

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет

вычислительной техники

Кафедра

«Вычислительная техника»

Направление подготовки: 09.04.01 «Информатика и вычислительная техника»

Магистерская программа: «Вычислительные машины, комплексы, системы и сети»

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

на тему

**АВТОМАТИЗАЦИЯ АНАЛИЗА ТЕКСТОВ
ДОКУМЕНТОВ ОРГАНИЗАЦИИ С ЦЕЛЬЮ
ОПТИМИЗАЦИИ ПОТОКОВ ДОКУМЕНТОВ**

Студент

(подпись, дата)

Баканова Наталия Борисовна

(ФИО полностью)

Руководитель

(подпись, дата)

7.06.2023

Зинкин С.А.

(фамилия, инициалы)

Нормоконтролёр

(подпись, дата)

7.06.2023

Бычков А.С.

(фамилия, инициалы)

Рецензент

нат. орг. дела
ИТДП, КФ «Информатика»
(должность, место работы)

08.06.23
(подпись, дата)

Арекин А.Ю.
(фамилия, инициалы)

Работа допущена к защите (протокол зас. кафедры от 02.06.2023 № 14)

Заведующий кафедрой

(подпись)

Митрохин М.А.

(фамилия, инициалы)

Работа защищена с отметкой отлично (протокол заседания ГЭК от 9.06.23 № 5)

Секретарь ГЭК

(подпись)

Гурин Е.И.

(фамилия, инициалы)

Пенза 2023

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет
вычислительной техники

Кафедра
«Вычислительная техника»

«Утверждаю»
Заведующий кафедрой
М.А. Митрохин
« 29 » ноября 2021 г.

**ЗАДАНИЕ
НА МАГИСТЕРСКУЮ ДИССЕРТАЦИЮ**

1. Студент Баканова Наталия Борисовна гр. 21ВВВм1 факультета вычислительной техники, направления 09.04.01

2. Тема работы «Автоматизация анализа текстов документов организации с целью оптимизации потоков документов».

Тема утверждена приказом ПГУ № 924/о от "29" ноября 2021 г.

3. Руководитель магистерской диссертации д.т.н., профессор Зинкин С.А.

4. Задание на работу (назначение разработки, исходные данные и т.п.)

Назначение разработки: осуществление анализа текстов документов для принятия решения о повышении эффективности функционирования системы документационного обеспечения управления организации.

Реализация аналитического решения исследования наборов ключевых слов документов для нахождения внутренних резервов оптимизации документопотоков.

5. Перечень подлежащих разработке вопросов

1. Исследование существующих методов анализа состава и движения документов.

2. Составление методики анализа документопотоков организации на основе сравнения множеств ключевых слов, описывающих документы и функции должностных лиц.

3. Разработка методики определения ключевых слов документов и сотрудников.

4. Разработка автоматизации аналитического решения попарного сравнения наборов ключевых слов.

5. Тестирование автоматизированного аналитического решения анализа множества ключевых слов документов.

6. Календарный график выполнения работы

№ п/п	Наименование этапов работы	Объем работы	Срок выполнения	Подпись руководителя
1.	Исследование существующих методов анализа состава и движения документов	10%	29.11.2021- 20.12.2021	
2.	Составление методики анализа документопотоков организации на основе сравнения множеств ключевых слов, описывающих документы и функции должностных лиц	13%	21.12.2021- 28.02.2022	
3.	Разработка методики определения ключевых слов документов и сотрудников	10%	01.03.2022- 01.05.2022	
4.	Разработка автоматизации аналитического решения попарного сравнения наборов ключевых слов	39%	02.05.2022- 27.12.2022	
5.	Тестирование автоматизированного аналитического решения анализа множества ключевых слов документов	15%	28.12.2022- 29.03.2023	
6.	Оформление выпускной работы	10%	30.03.2023- 31.05.2023	

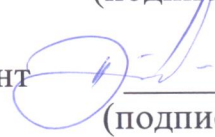
Дата выдачи задания " 29 " ноября 2021 г.

Руководитель магистерской диссертации


(подпись, дата)

Зинкин С.А.

Задание к исполнению принял студент


(подпись, дата)

Баканова Н.Б.

Работу к защите допустить
Декан факультета

 08.06.2023
(подпись, дата)

Фионова Л.Р.

РЕФЕРАТ

Магистерская работа содержит 60 л., 20 рис., 3 табл., 20 источн., 1 прил.

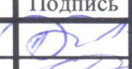
АВТОМАТИЗАЦИЯ, ДОКУМЕНТОПОТОКИ, СИСТЕМА
ДОКУМЕНТАЦИОННОГО ОБЕСПЕЧЕНИЯ УПРАВЛЕНИЯ,
ОПТИМИЗАЦИЯ, ДОКУМЕНТ, КЛЮЧЕВЫЕ СЛОВА, LOGINOM

Цель работы – исследование текстов документов, сопровождающих бизнес процесс, с целью повышения эффективности документопотоков.

В данной магистерской диссертации объектом исследования выступает бизнес-процесс. Предметом исследования являются документационные процессы организаций, исследование которых с помощью ключевых слов может помочь найти внутренние ресурсы для повышения качества документопотоков.

В результате проведенной работы разработаны методика анализа документопотоков организации на основе анализа ключевых слов документов, методика определения ключевых слов документов и сотрудников, автоматизация аналитического решения исследования ключевых слов документов с целью оптимизации документопотоков.

Дальнейшее развитие разработанной системы может включать в себя раздел автоматизации поддержки принятия решения на основе полученных в ходе исследования данных.

					ПГУ1.090401.01.001 ПЗ		
Изм.	Лист	№ докум.	Подпись	Дата			
Разраб.		Баканова Н.Б.			<i>Автоматизация анализа текстов документов организации с целью оптимизации потоков документов. Пояснительная записка</i>		
Провер.		Зинкин С.А.					
Н. Контр.		Бычков А.С.					
Утверд.		Митрохин М.А.					
					Лит.	Лист	Листов
						2	60
					ФВТ, каф. ВТ гр. 21ВВВм1		

СОДЕРЖАНИЕ

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ.....	5
ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ	6
ВВЕДЕНИЕ	7
1 Анализ существующих методов анализа состава и движения документов.....	9
1.1 Графический метод исследования информационных документопотоков.....	10
1.2 Описание потоков информации с использованием теории графов	10
1.3 Описание потоков информации с помощью семантической сети.....	11
1.4 Метод анализа оперограмм и технологических карт документооборота.....	12
1.5 Выводы.....	12
2 Методика анализа документопотоков организации	15
2.1 Составление списка документов процесса.....	17
2.2 Выделение ключевых словосочетаний нормативных документов, регламентирующих выполнение бизнес-процесса.....	20
2.3 Выделение ключевых словосочетаний документов, сопровождающих бизнес-процесс.....	24
2.4 Сравнительный анализ устойчивых словосочетаний и ключевых слов текстов документов, сопровождающих бизнес-процесс.....	24
2.5 Выводы.....	27
3 Автоматизация сравнительного анализа устойчивых словосочетаний и ключевых слов текстов документов.....	29
3.1 Выбор платформы для решения задачи автоматизации	29
3.2 Разработка требований к решению задачи автоматизации	33
3.3 Подготовка исходных данных	34
3.4 Автоматизация алгоритма попарного сравнения наборов ключевых слов.....	37
3.5 Выводы.....	38

4	Тестирование автоматизированного решения сравнительного анализа устойчивых словосочетаний и ключевых слов текстов документов.....	39
4.1	Описание тестовых данных	39
4.2	Составление списка документов процесса итоговой аттестации	39
4.3	Выделение ключевых словосочетаний из нормативных документов, участвующих в реализации процесса	43
4.4	Сравнительный анализ устойчивых словосочетаний и ключевых слов текстов распорядительных документов.....	48
4.5	Выводы.....	54
	ЗАКЛЮЧЕНИЕ	56
	СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	58
	ПРИЛОЖЕНИЕ А. Презентация	61

ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящей работе применяют следующие термины с соответствующими определениями

Семантическая сеть – информационная модель предметной области, имеющая вид ориентированного графа, вершины которого соответствуют объектам предметной области, а дуги (рёбра) задают отношения между ними

Система управления базами данных – комплекс программ, позволяющих создать базу данных и манипулировать данными

Система документационного обеспечения управления – комплексная организация работы с документами в организации

Словарь устойчивых словосочетаний понимается устойчивая конструкция из нескольких последовательно расположенных слов, имеющая определенный смысл только в рамках этой конструкции

Loginom – low-code платформа реализации аналитических решений

Low-code – метод проектирования и разработки приложений с помощью интуитивно понятных графических инструментов и встроенных функций, которые снижают традиционные (профессиональные) требования к написанию кода

Парсер – программа для сбора и систематизации информации

Дашборд – это информационная панель, которая получает данные из других систем и отображает их в понятном виде

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

В настоящей работе применяют следующие сокращения и обозначения

ДОУ – документационное обеспечение управления

СЭД – система электронного документооборота

ИноУП – кафедра «Информационное обеспечение управления и производства»

ПГУ – Пензенский государственный университет

ПО – программное обеспечение

СУБД – система управления базами данных

БД – база данных

ГАК – государственная аттестационная комиссия

ГЭК – государственная экзаменационная комиссия

УР – учебная работа

УМУ – учебно-методическое управление

ПУ – правовое управление

SQL (Structured Query Language) – структурированный язык запросов

API (Application Programming Interface) – описание способов взаимодействия одной компьютерной программы с другими

ВВЕДЕНИЕ

Реалии функционирования современных организаций требуют обязательного документного подтверждения свершившихся фактов и событий. Поэтому перед каждой организацией неизменно встает проблема эффективного управления, рациональной организации и оптимизации документопотоков предприятия. Эффективное решение обозначенных задач позволяет вывести на качественно новый уровень не только реализацию протекающих в рамках любого предприятия бизнес-процессов, но и функционирование организации в целом. Оптимизация и постоянное совершенствование документооборота повышает управляемость отдельных бизнес-процессов, улучшает качество их реализации, снижает текущие расходы, улучшает управляемость всей организации за счет своевременного получения необходимой и актуальной информации, что в конечном итоге приводит к повышению конкурентоспособности в жестких условиях современной рыночной экономики.

Правильно организованный документооборот должен основываться на определенных принципах, обеспечивающих оперативное прохождение документов по наиболее короткому пути с минимальными затратами времени.

Основные принципы организации документооборота:

- целевой принцип. Все документы должны иметь определенную цель и направлены на достижение определенных задач,
- принцип единства методов регистрации. Регистрация документации должна происходить единообразно, в соответствии с установленными правилами и процедурами,
- принцип краткости и ясности документов. Документы должны быть составлены кратко и ясно, без лишних подробностей,
- принцип правильной адресации. Документы должны быть адресованы непосредственно тому, кто должен знать содержимое документа,

- принцип согласованности и утверждения. Любые изменения в документации должны проходить через процедуру согласования и утверждения,
- принцип контроля. Документы должны проходить контроль на каждом этапе их прохождения, чтобы убедиться в их правильности и соответствии установленным требованиям,
- принцип минимизации временных затрат. Необходимо максимально сократить время, затрачиваемое на прохождение документов через систему делопроизводства,
- принципы организации документооборота позволяют создать эффективную систему делопроизводства, где документы проходят быстро и без потерь времени и ресурсов [1].

Недостатки в оформлении, движении, обработке, хранении документов приводят к серьезным трудностям в работе как рядовых сотрудников, так и руководителей, что в конечном итоге сказывается на функционировании организации в целом. Устранение проблем документооборота требует применения комплексного подхода и подразумевает анализ состояния документооборота и бизнес-процессов, как источника создания документов, а также реализацию системы мер по совершенствованию состояния документопотоков.

Задача анализа состава создаваемых документов и особенностей их движения становится особенно важной, если в организации принимается решение о переходе к электронному документообороту. Исследование и оценка состояния документопотоков – приоритетная функция службы документационного обеспечения управления (службы ДОУ). Однако практика показывает, что организации, иногда даже крупные, не всегда имеют компетентного сотрудника службы ДОУ, знакомого с тонкостями ведения документации и способного провести всестороннее исследование документооборота. Одним из возможных вариантов решения данной проблемы является автоматизация процесса анализа документооборота.

1 Анализ существующих методов анализа состава и движения документов

Документооборот предприятия или учреждения, как известно, не должен быть стихийным или произвольным. Для его правильной организации требуется тщательный анализ как документальной, так и управленческой деятельности предприятия, т. к. рационализация ведет к сокращению инстанций, через которые проходит документ при согласовании и подписании, к прямоотчности движения документов, исключению повторных инстанций и возвратности. В конечном итоге происходит значительная экономия рабочего времени и материальных ресурсов. Движение документов на предприятии предопределено его организационной структурой и особенностями реализуемых бизнес-процессов и управления.

Анализ документооборота предполагает тщательное изучение внутренних нормативных документов: устава (положения) организации и положений о структурных подразделениях, штатного расписания, должностных инструкций работников. Это позволит определить тип делопроизводства, проанализировать распределение обязанностей между руководителями, четко очертить их права и компетенцию, изучить взаимосвязь разных уровней управления и производства.

Структура организации во многом определяет тип ее делопроизводства: централизованное, децентрализованное или смешанное. Как правило, небольшая фирма использует централизованную форму ДООУ, когда за создание, хранение, регистрацию, контроль документов отвечает конкретное должностное лицо или отдел (канцелярия). При децентрализованной системе все виды работ с документами производятся непосредственно в структурных подразделениях организации, то есть делопроизводство каждое функциональное подразделение ведет самостоятельно. Предприятия со сложной структурой управления прибегают к смешанному типу делопроизводства, когда часть документов создается, регистрируется и хранится в подразделениях, а другая часть подлежит

централизованной обработке. Структурная часть исследования для наглядности и удобства восприятия обычно оформляется в виде схем, рисунков, диаграмм [2].

Следующий этап исследования – изучение качественного состава документопотоков. При этом учитываются все виды документов, существующие в организации. В качестве объекта исследования можно рассматривать как отдельные виды документов, так и группы документов. Но основополагающим объектом исследования и совершенствования в процессе управления документацией считается документопоток, управляющими параметрами которого являются: содержание или функциональная принадлежность; структура потока; направление потока и его объем.

Цель анализа документооборота состоит в том, чтобы исключить возвратность и задержку движения документов, а значит, повысить скорость и качество исполнения конкретного задания.

1.1 Графический метод исследования информационных документопотоков

Графический метод используется для описания документопотоков небольшой размерности на макроуровне, для выявления общей структуры и функций системы управления, а также для совершенствования существующих потоков информации.

Основные элементы потока – документы. Отношение между ними изображается в виде графической схемы. Процедуры преобразования элементов потока (обработки документов) записываются в виде кратких пояснений на схеме потока. Система координат графика двумерная [3].

1.2 Описание потоков информации с использованием теории графов

Описание потоков информации с помощью графов является одним из наиболее разработанных методов. С его помощью достигается наглядность функционирования системы управления и движения потоков информации; применение математического аппарата теории графов позволяет

оптимизировать работу управления и каналов связи; имеется возможность также представить динамику управления и движения информации, которая ускользает при использовании другими методами. В настоящее время имеется много примеров использования теории графов в описании данных процессов. Они различаются по характеру описываемых объектов, по видам графов [3].

1.3 Описание потоков информации с помощью семантической сети

Семантическая сеть — информационная модель предметной области, имеющая вид ориентированного графа, вершины которого соответствуют объектам предметной области, а дуги (рёбра) задают отношения между ними. Объектами могут быть понятия, события, свойства, процессы. Семантическая сеть является одним из способов представления знаний. Знания — это закономерности предметной области (принципы, связи, законы), полученные в результате практической деятельности и профессионального опыта, позволяющие специалистам ставить и решать задачи в этой области [3].

Использование онтологий позволяет сделать шаг к решению задачи анализа содержания текстов документов и сопоставлению текстов документов с описанием должностных обязанностей работников.

Онтология представляет собой описание понятий (классов) в рассматриваемой предметной области, свойств (слотов), описывающих каждое понятие, и ограничений, наложенных на слоты. Онтология вместе с набором индивидуальных экземпляров классов входит в базу знаний интеллектуальной системы. Характерной особенностью онтологий является однородность смысловых связей между входящими в нее понятиями. Однако современные интеллектуальные системы и усложнение структуры предметных областей требуют обеспечения возможности использования разнородных понятий и различающихся по смыслу связей между ними. В такой ситуации более предпочтительно применение аппарата семантических сетей, позволяющего реализовать указанное требование.

Достоинством семантических сетей является их универсальность, достигаемая за счет выбора соответствующего предметной области набора отношений. В принципе с помощью семантической сети можно описать сколь угодно сложную ситуацию, факт или предметную область. Поэтому использование аппарата семантических сетей позволяет как нельзя лучше описать процедуру принятия решений по рационализации документооборота, которая, как известно, имеет неформальный характер и должна учитывать особенности предметной области.

1.4 Метод анализа оперограмм и технологических карт документооборота

Основным методом совершенствования технологии документооборота является метод моделирования.

В маршрутной карте технологического процесса документооборота отражается маршрут движения документа с момента его составления до сдачи в архив. По вертикали технологической карты располагаются операции в их технологической последовательности, по горизонтали – продолжительность процесса. Технологическая карта документооборота дает возможность регулировать последовательность операций, производимых с основными видами документов.

Особенностью анализа документопотоков является включение в исследование не только самих процессов обработки документов, но и предшествующих им этапов формирования содержания проектов документов и придание этим проектам юридической силы документов [4].

1.5 Выводы

Сегодня не существует единого и универсального метода для решения этой задачи. Существующие методы не позволяют провести комплексный анализ документооборота организации. Ни один из методов не может претендовать на универсальность для решения задачи анализа документооборота.

Сравним эти методы и определим обязательные этапы, необходимые для проведения анализа документопотоков (таблица 1).

Таблица 1 - Сравнение существующих методов анализа документопотоков

Метод	Преимущества	Недостатки	Возможность использования для анализа документооборота
Графический	Является простым, наглядным, универсальным и экономичным методом описания потоков информации	При увеличении размерности потока схема может стать настолько велика, что потеряет свою ценность как средство анализа, или будет настолько поверхностна в деталях, что не окажет помощи при анализе потоков информации	Целесообразно использовать в совокупности с формализованными методами. Позволяет проследить пути документов, выявить моменты их образования, операции, которые с ними осуществляются, порядок, в котором документы объединяются или расчленяются
Теория графов	Наглядность, интерпретируемость, компактность представления, структурность	Для проведения структурного анализа исследуемого объекта необходимо иметь в распоряжении довольно подробную информацию о нем, о взаимодействии всех элементов его структуры	Применение совместно с математическим аппаратом, позволяющим производить расчет различных числовых характеристик, например, загруженности должностных лиц на рабочих местах.

Метод	Преимущества	Недостатки	Возможность использования для анализа документооборота
Семантическая сеть понятий	Универсальность, достигаемая за счет выбора соответствующего предметной области набора отношений	Практическая необозримость при описании модели мира реального уровня сложности. При этом появляется проблема размещения семантической сети в памяти ЭВМ	Вербальное описание процесса, выявление «узких мест»
Анализ оперограмм и технологических карт документо-потоков	Возможность регулирования последовательности операций, производимых с основными видами документов.	Невозможность оценки содержания выполняемых действий и совмещения сведений большого количества карт	Можно систематизировать содержание, рационализировать маршруты движения документов, определить виды работ по документированию управленческих процессов

Можно сделать вывод, что наиболее эффективно с точки зрения анализа документооборота является использование нескольких методов в комплексе.

2 Методика анализа документопотоков организации

Для эффективного анализа документооборота необходимо начать с изучения внутренних нормативных документов, которые регулируют делопроизводство в организации. Это включает в себя положения о порядке регистрации, хранения и передачи документов, правила оформления документов, сроки их рассмотрения, утверждения и выполнения [5].

После анализа внутренних нормативных документов можно перейти к анализу фактического документооборота организации. Это позволит выявить отличия между установленными процедурами и реальными практиками в организации. Например, может быть выявлено, что сроки рассмотрения документов не соблюдаются, процедуры согласования не выполняются в полном объеме, а контроль за исполнением документов отсутствует.

Таким образом, анализ внутренних нормативных документов позволяет определить, как происходит документооборот в организации на законодательном уровне, а анализ фактического документооборота поможет выявить проблемы и пропуски в текущей практике. Эти данные позволят разработать рекомендации для улучшения системы делопроизводства и повышения эффективности работы организации.

При анализе внутренних нормативных документов основное внимание должно быть уделено содержанию документов. На кафедре ИнОУП ПГУ разработана методика анализа ключевых слов документов [1]. Полученная информация позволяет сформировать исходные данные для проведения собственно анализа документопотоков организации. Анализ включает следующие этапы:

- структурирование бизнес-процесса и составление полного списка документов, сопровождающих его,
- выделение ключевых словосочетаний из числа нормативных документов, участвующих в реализации процесса,
- оценка рациональности маршрутов прохождения документов.

Эффективность управления документами во многом определяет успешность бизнеса. В свою очередь, для повышения эффективности управления документами в современных условиях динамично изменяющегося рынка необходимо проводить мониторинг бизнес-процессов и своевременно оптимизировать документооборот предприятия [7].

Процедуры проектирования и оптимизации документооборота организации должны основываться на моделировании маршрутов движения документов. Последнюю задачу предлагается решить с помощью исследования ключевых слов. Оно заключается в сопоставлении наборов ключевых слов, содержащихся в документах и/или описывающих фактически исполняемые должностные обязанности конкретного сотрудника, представленные в положениях о подразделениях, должностных регламентах или инструкциях сотрудников и др. В случае отсутствия нормативных документов необходимо детальное обследование организации с целью выявления существующих бизнес-процессов и порождаемых ими документопотоков.

Исследовательский этап методики состоит из следующих шагов [2]:

- подготовка текста к анализу. При этом из текста необходимо исключить маркировку списков, проверить орфографию и пунктуацию и др.,
- построение последовательного списка слов текста документа, с сохранением пунктуации предложений,
- определение грамматических и морфологических признаков слов (часть речи, число, время, род и др.). В данной работе используется кодировка, принятая в морфологическом анализаторе текстов на русском языке *mystem* (компания Яндекс). Например, существительные — S, прилагательные — A, глаголы — V, предлоги — PR, союзы — CONJ и т.д.,
- приведение слов к словарной форме. Выявление устойчивых словосочетаний с помощью словаря устойчивых словосочетаний.

Под термином словарь устойчивых словосочетаний понимается устойчивая конструкция из нескольких последовательно расположенных

слов, имеющая определенный смысл только в рамках этой конструкции. В словаре для каждого термина указывается часть речи основного слова.

2.1 Составление списка документов процесса

Для составления полного списка документов, сопровождающих бизнес-процесс, необходимо произвести структурирование списка. Этот процесс включает в себя несколько шагов:

- определение бизнес-процесса и его этапов. Для этого нужно раздробить процесс на отдельные этапы, определить последовательность действий и уточнить, какие документы необходимы для каждого этапа,
- построение структуры документов. Проанализировав каждый этап бизнес-процесса, необходимо составить список документов, которые потребуются на каждом этапе. Это может быть список документов, которые нужно заполнить, передать, проверить, подписать и т.д.,
- группировка документов. Составив список документов, необходимо определить их взаимосвязь и группировать их по типу, содержанию или функциональному назначению. Например, все документы для одного процесса закупок можно определить в отдельную группу,
- проверка и уточнение списка документов. После завершения первых трех этапов необходимо провести анализ структуры документов и уточнить ее. В процессе анализа могут выявиться новые типы документов, которые нужно добавить в список, а также проясниться некоторые вопросы по тому, какие документы нужны на каждом этапе,
- внесение изменений в список. После проведения анализа и уточнения структуры документов, необходимо внести изменения в список. В конечном итоге каждый этап бизнес-процесса будет иметь свой список документов, который необходимо использовать для эффективной работы с документами в рамках данного процесса.

Входными и выходными данными для каждого этапа и будут являться документы, сопровождающие выполнение процесса [8].

Структурирование процесса можно осуществлять, в том числе, на основе стандартов организации. К примеру, если в организации используется определенный стандарт документооборота или методология управления бизнес-процессами, то структурирование процесса может осуществляться с учетом этих стандартов и методологий.

Стандарты и методологии могут содержать рекомендации по использованию определенных типов документов, процедур обработки документов и требований к документообороту. Поэтому, при структурировании процесса на основе стандартов, можно использовать эту информацию, чтобы составить список документов, которые должны быть привязаны к каждому этапу процесса.

Как правило, стандарты и методологии управления процессами и документооборотом организации должны поддерживать общие принципы и стандарты, выработанные в соответствии с международными стандартами и рекомендациями. Поэтому, использование стандартов и методологий в структурировании процесса позволяет не только упростить и ускорить выполнение работ, но и обеспечить соответствие документооборота организации национальным и международным стандартам и регуляторным требованиям.

Термины, выделенные в процессе анализа текстов документов, неоднократно встречаются в текстах стандарта и употребляются в различных формах. Часто стандарты используют специальные термины для описания процессов, инструментов, методологий и других элементов, входящих в их сферу действия. Использование этих терминов обеспечивает унификацию понимания и общение между специалистами, работающими в данной области [9].

Кроме того, в стандартах могут присутствовать различные формы слов, такие как множественное число, формы глаголов или существительных в различных временах и т.д. Это может вызывать затруднения в понимании текста стандарта, особенно для тех, кто не имеет достаточного опыта в

данной области или не является носителем языка, на котором написан стандарт.

Поэтому, для более эффективного понимания и использования стандартов, рекомендуется ознакомиться с определениями и инструкциями по применению терминов, используемых в стандарте, а также изучать соответствующие грамматические правила для правильного использования терминов в различных формах.

Создание словаря синонимов может значительно помочь повысить эффективность анализа текстов документов. Такой словарь подойдет для случаев, когда в документах используются различные термины, но они имеют одинаковый или похожий смысл. В этом случае, словарь синонимов поможет определить схожесть значений терминов и избежать ошибок в их интерпретации.

Создание такого словаря может быть осуществлено путем обработки большого количества текстовых документов с помощью специальных алгоритмов и методов машинного обучения. В результате, можно получить список слов, которые могут использоваться в качестве синонимов для каждого термина, встречающегося в текстах документов.

Кроме того, создание словаря синонимов может быть осуществлено вручную при помощи специального программного обеспечения или просто составлением списков синонимов вручную. Этот метод может быть чрезвычайно полезен для создания специализированных словарей, которые содержат термины и синонимы, используемые в конкретной области знаний или отрасли.

Однако, стоит учитывать, что создание словаря синонимов - это сложный процесс, требующий достаточно больших усилий. Более того, такой словарь должен регулярно обновляться и модернизироваться, чтобы учитывать изменения в употреблении терминов в текстах документов. Данный словарь поможет провести более подробный сравнительный анализ устойчивых словосочетаний и ключевых слов.

2.2 Выделение ключевых словосочетаний нормативных документов, регламентирующих выполнение бизнес-процесса

Анализ бизнес-процессов можно начать с внутренних стандартов предприятия. Внутренние стандарты – это документы, содержащие правила и процедуры для выполнения различных операций внутри предприятия. Они определяют, как должны быть выполнены определенные задачи, а также устанавливают нормы и стандарты для качества выполнения этих задач.

Анализ внутренних стандартов может помочь определить, насколько эффективно выполняются бизнес-процессы в предприятии. Использование внутренних стандартов может позволить предприятию добиться более высокой эффективности и точности выполнения задач, а также обеспечить единый подход к выполнению бизнес-процессов. Анализ внутренних стандартов может также позволить выявить проблемы в выполнении бизнес-процессов и необходимость их улучшения.

Однако, стоит также учитывать, что внутренние стандарты могут быть несовершенными или устаревшими, поэтому необходимо проводить их регулярный анализ и обновление. Кроме того, при анализе бизнес-процессов необходимо учитывать не только внутренние стандарты, но и внешние факторы, такие как требования законодательства или стандартов отрасли. Поэтому, успешный анализ бизнес-процессов должен основываться на комплексном подходе и учитывать множество различных факторов.

Таким образом, проанализировав стандарт, можно определить информативность его текста. Далее необходимо выделить из текста документа устойчивые словосочетания, содержащиеся в разделе Порядок проведения.

Выделение устойчивых словосочетаний очень полезно при анализе текстов документов. Устойчивые словосочетания – группы слов, которые часто встречаются вместе и образуют определенное значение, которое не всегда можно легко определить по отдельности каждого слова.

Выделение устойчивых словосочетаний может помочь улучшить понимание текста, особенно если он содержит специфические термины и технические термины. Это позволяет увидеть связи между словами и определить, как они связаны друг с другом в рамках текста. Кроме того, устойчивые словосочетания могут использоваться для создания словарей терминов и сокращений.

Существует несколько способов выделения устойчивых словосочетаний, таких как использование частотного анализа, анализа соотнесения, статистики биграмм и триграмм, а также использование специальных программ, которые могут автоматически выделять устойчивые словосочетания на основе алгоритмов машинного обучения.

Начать можно с обнаружения частых и повторяющихся сочетаний слов в тексте документа, возможно, даже при помощи поисковых запросов. Затем следует проводить анализ связей между этими словами и определять наиболее часто используемые устойчивые словосочетания

Именно эти термины отражают содержание действий, выполняемых в рамках данного процесса, и тем или иным образом должны присутствовать в должностных инструкций, участвующих в процесс сотрудников, а также в тексте создаваемых документов.

Выделение устойчивых словосочетаний позволяет оценить информационный объем документа как отношение количества ключевых слов к общему количеству слов в документе [10].

Информативность документа – это степень его смысловой и содержательной новизны. информативность документа оценивается по степени его содержательной новизны и смысловой ценности. Информативность документа зависит от того, насколько он содержит новую и полезную информацию.

Новизна документа может заключаться в наличии новых данных, фактов, исследований или методологий, которые ранее не были известны или не были представлены в других документах. Смысловая ценность документа

определяется его способностью привнести новый взгляд на известные факты или представить эту информацию в ином контексте.

Чтобы оценить информативность документа, следует рассмотреть его содержание, структуру, доступность и точность. Документ должен быть информативным, если содержит весомые данные и факты, обоснованные исследования, приложения, карты, графики и т.д.

Содержательная новизна и смысловая ценность документа могут быть оценены и в сравнении с аналогичными документами на ту же тему. Если документ имеет преимущества по сравнению с другими документами, то он считается информативным.

Информативный документ должен помочь пользователю получить полное и точное представление о теме. Он должен быть понятным, доступным и содержать корректную и проверенную информацию. Это поможет воспринимать и использовать документ более эффективно и принимать информированные решения.

В связи с возможностью разного способа представления информации в тексте документа – экономного и избыточного – существенной оказывается проблема авторского намерения и читательского восприятия сообщения, т.е. различной оказывается «глубина прочтения» текста. Для разных читателей критерий «экономности-избыточности» окажется разным [3]. Избыточной информации не должно быть слишком много, иначе текст будет труден для прочтения, но все-таки она должна быть, иначе текст будет труден для восприятия.

Проблема авторского намерения и читательского восприятия сообщения заключается в том, что автор может иметь определенное намерение, создавая текст, но читатели могут трактовать его по-разному, в зависимости от своего опыта, культурных ценностей и восприятия мира.

Авторское намерение относится к тому, что автор хотел бы донести до своих читателей. Например, автор может писать статью, которая призывает к

действию на определенной теме, или может написать книгу с целью повысить осведомленность людей о конкретной теме.

Однако, при чтении текстов, читатели могут истолковывать информацию и воспринимать ее по-разному. Это может зависеть от того, насколько точно автор передал свою мысль, насколько однозначен язык, насколько сходны автор и читатели в их культурных установках и понимания мира.

Когда сообщение неоднозначно, читатели могут прочесть его иначе, чем автор предполагал. Это может привести к формированию различных точек зрения, мнений и даже споров. Также это может привести к неправильному пониманию текста и созданию образа автора, который не соответствует истинному намерению автора.

Решить проблему авторского намерения и читательского восприятия может помочь тщательная редакция и перепроверка текстов, чтобы авторское намерение было еще более ясно и понятно для читателей. Также может использоваться контекстуальное прочтение, дополнительный двусторонний обмен мнениями и понимание культурных и языковых особенностей читателей.

Считается, что соотношение ключевых слов к общему количеству слов, должно быть 1:2. В этом случае текст документа будет понятным и легким для восприятия, что необходимо для исключения ошибок в исполнении документа [11].

Далее должны быть рассмотрены должностные инструкции участников процесса. Для этого необходимо выделить из текста документа устойчивые словосочетания, определяющие основные функции в деятельности сотрудников.

Для того, чтобы проверить, соответствуют ли должностные обязанности сотрудника, формально закрепленные в его должностной инструкции, содержанию конкретного бизнес-процесса, необходимо сопоставить ключевые слова должностных инструкций и стандарта.

Очевидным препятствием к осуществлению такой проверки является общий характер используемых в нормативных документах словесных формулировок, а также то обстоятельство, что нормативные документы создаются в разные периоды существования организации и разными людьми.

2.3 Выделение ключевых словосочетаний документов, сопровождающих бизнес-процесс

Как и в предыдущем этапе, здесь необходимо составить таблицы ключевых слов, определяющих тексты документов, сопровождающих бизнес-процесс. Каждый конкретный бизнес-процесс может иметь свои уникальные документы и формы для заполнения, но указанные ключевые слова представляют общую базу для классификации документов бизнес-процесса. Такие таблицы должны быть сформированы для всего состава документов процесса.

Можно также произвести анализ текстов документов на избыточность. Избыточность в текстах документов может привести к ненужному усложнению, затруднениям в понимании и увеличению времени на чтение/анализ.

2.4 Сравнительный анализ устойчивых словосочетаний и ключевых слов текстов документов, сопровождающих бизнес-процесс

Этот вид анализа проводится для определения рациональности существующих маршрутов прохождения документов.

Полученные результаты являются исходными данными для второго (аналитического) этапа методики анализа текстов документов, целью которой является формулирование рекомендаций по оптимизации бизнес-процессов и электронных документопотоков.

Для анализа применяются операции над множествами (объединение, пересечение, вычитание). В качестве элементов множеств рассматриваются ключевые слова различных документов и/или ключевые слова, полученные

по результатам обследования предметной области и отнесённые к категориям «субъект», «объект», «роль», «вид деятельности».

Анализу могут подвергаться: документы системы менеджмента качества организации; инструкции (в том числе должностные); положения о структурных подразделениях; документы, образовавшиеся в оперативной деятельности организации. Сравнение ключевых слов документов позволяет найти:

- общие вопросы, рассматриваемые несколькими субъектами (должностными лицами),
- вопросы, рассматриваемые одним субъектом, но не рассматриваемые другим,
- вопросы, не рассматриваемые ни одним из субъектов.

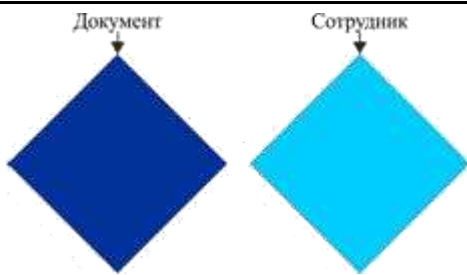
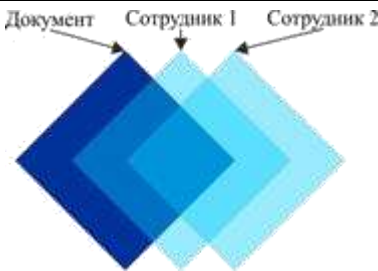
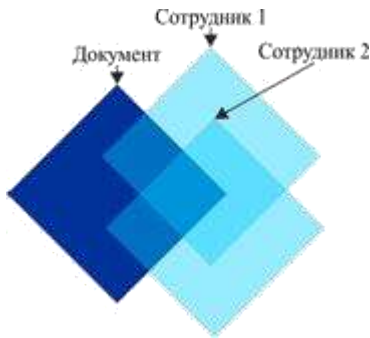
Отдельную проблему составляет интерпретация полученных результатов и выработка рекомендаций по улучшению сложившейся ситуации. Для снижения зависимости от уровня квалификации пользователя предлагается разработка комплекса стандартных автоматизированных процедур анализа полученных результатов, основными из которых являются:

- анализ полноты состава документов, сопровождающих выполнение бизнес-процесса,
- проверка наличия дублирования работ, выполняемых должностными лицами,
- выявление недокументируемых участков деятельности структурных подразделений,
- проверка актуальности содержания нормативных документов, регламентирующих основные бизнес-процессы и делопроизводство.

Перечисленные процедуры предполагают попарное сравнение интересующих пользователя документов и дополняются наборами рекомендаций по устранению выявляемых проблем.

В таблице 2 представлены разновидности ситуаций, возникающих при сопоставлении ключевых слов распорядительного документа, например приказа, и должностных инструкций сотрудников, участвующих в согласовании документа.

Таблица 2 – Примеры ситуаций, возникающих при сопоставлении ключевых слов распорядительного документа и должностных инструкций сотрудников, участвующих в его согласовании

№ п/п	Описание ситуации	Графическая иллюстрация
1.	$T_d \cap T_c = \emptyset$	
2.	$(T_d \cap T_{c1}) \cap (T_d \cap T_{c2}) = \emptyset$	
3.	$(T_d \cap T_{c1}) \cap (T_d \cap T_{c2}) = \emptyset$, причём $(T_d \cap T_{c1}) \gg (T_d \cap T_{c2})$	
4.	$(T_d \cap T_{c1}) \cap (T_d \cap T_{c2}) = \emptyset$, причём $ T_d \cap T_{c1} \approx T_d \cap T_{c2} $	

№ п/п	Описание ситуации	Графическая иллюстрация
5.	$(T_d \cap T_{c1}) \cap (T_d \cap T_{c2}) = \emptyset$, причём $T_{c2} \in T_{c1}$	

В таблице использованы следующие обозначения:

- T_d, T_c – множества ключевых слов распорядительного документа и должностной инструкции сотрудника соответственно;
- T_{c1}, T_{c2} – множества ключевых слов должностных инструкций первого и второго сотрудников соответственно.

Аналогичным образом проанализируем все документы, участвующие в данном бизнес-процессе, в жизненном цикле которых участвует конкретное должностное лицо.

2.5 Выводы

Результаты проведенного анализа дают возможность оценить эффективность документооборота организации. На первом этапе исследования создаются словари, содержащие ключевые словосочетания, отображающие основные функции должностных лиц. На основе данных словосочетаний документы оценивались на их информативную полноту. Это производилось потому, как есть вероятность, что тексты локальных документов являются информационно неточными и могут быть восприняты не верно. Таким образом, возникает риск снижения эффективности управленческой деятельности.

Таким образом, применение рассматриваемой методики анализа документопотоков делает возможным регулирование объема документооборота за счет исключения из него документов, не относящихся к рассматриваемому процессу.

Сравнительный анализ ключевых словосочетаний и ключевых слов распорядительных документов позволяет определить рациональность

существующих маршрутов прохождения документов. Исследования могут показать, что локальные документы отчасти носят формальный характер и не отображают полной картины деятельности должностного лица.

Рассмотренная методика является эффективной, так как позволяет достигнуть основных целей исследования документопотоков организации, а именно:

- повышение эффективности управленческой деятельности,
- ускорение движения документов в организации,
- уменьшение трудоемкости обработки документов.

Применение методики позволяет выявить недостатки комплекта нормативных документов и сформулировать предложения по совершенствованию текстов нормативов.

3 Автоматизация сравнительного анализа устойчивых словосочетаний и ключевых слов текстов документов

3.1 Выбор платформы для решения задачи автоматизации

Аналитические информационные системы обладают различными функциональными возможностями. Все эти возможности позволяют аналитическим информационным системам ускорить процессы принятия решений, улучшить качество анализа данных и повысить эффективность работы организации в целом.

Информация является важнейшей основой для принятия эффективных управленческих решений. Однако, только наличие большого количества данных не гарантирует успешного принятия решения. Для этого необходимо:

- правильно и точно собрать данные. Данные должны быть достоверными, а их сбор и хранение должны происходить в соответствии с требованиями безопасности и конфиденциальности,
- систематизировать и преобразовать данные, чтобы они стали информацией, которую можно использовать для принятия решений. Это может включать в себя очистку данных, переклассификацию, группирование, агрегирование и другие методы обработки,
- организовать данные для удобства использования, чтобы пользователи могли быстро найти нужную информацию. Это включает выбор правильных форматов отображения данных, создание отчетов, дашбордов, и других инструментов отображения информации,
- анализировать данные, чтобы получить из них новые знания и инсайты, применять методы анализа данных, проводить моделирование и прогнозирование,
- использовать полученную информацию в процессе принятия решений. При этом необходимо учитывать интересы всех заинтересованных сторон, учитывать ограничения и риски, связанные с решением.

Одним из ключевых элементов эффективного использования данных является их правильная организация. Данные могут быть представлены в

различных форматах и находиться в разных местах. Для того, чтобы данные были полезны и доступны для анализа, необходимо их собрать и организовать в единую базу данных.

После сбора и организации данных, следующий этап – их анализ. Он включает использование различных методов, таких как статистический анализ, моделирование и визуализация данных. Этот этап позволяет получить лучше понимание данных, их взаимосвязей и закономерностей.

Когда данные проанализированы, можно перейти к их использованию в процессах принятия решений. Для этого можно использовать специальные инструменты для построения отчетов и дашбордов. Они позволяют пользователям легко получать доступ к данным и использовать их для принятия решений.

Наконец, эффективное использование данных требует постоянного мониторинга и оценки результатов. Оценка результатов позволяет убедиться в том, что данные используются правильно и принятые решения помогают достигать поставленных целей. Если результаты оказываются неудовлетворительными, необходимо пересмотреть стратегию и принять новые решения на основе данных, чтобы улучшить результаты.

Бизнес-аналитика относится к использованию методов и инструментов для анализа данных и извлечения полезной информации, которая может использоваться в процессе принятия управленческих решений. Она включает в себя использование различных технологий и методов анализа данных, таких как статистический анализ, моделирование, предсказывающий анализ и машинное обучение [13].

Таким образом, бизнес-аналитика является важным инструментом для улучшения эффективности и эффективности бизнес-процессов, а также для принятия управленческих решений.

В настоящее время на рынке представлено множество разнообразных средств автоматизации анализа данных. У каждого из них есть свои плюсы и минусы. Набор инструментов, предлагаемый такими системами, и их

функциональность в чём-то схожи. Выбор средства автоматизации исходил из постановки задачи данной диссертационной работы.

Всем этим условиям наилучшим образом отвечает аналитическая low-code платформа Loginom, которая позволяет проводить анализ данных любого уровня сложности без программирования.

Платформа Loginom включает в себя набор инструментов и методов, направленных на сбор, организацию, анализ и использование данных для принятия управленческих решений.

Loginom предоставляет возможность проводить глубокий анализ данных с использованием различных методов, от простого отображения данных до сложной моделирования и машинного обучения. Это позволяет пользователям получать точные и достоверные данные для принятия управленческих решений.

В целом, Loginom – это мощный инструмент для анализа данных и принятия управленческих решений, который может быть использован в различных отраслях бизнеса. Она помогает лучше понимать данные и использовать их для обеспечения более эффективного и производительного бизнеса [14].

Данная система – продукт компании Loginom Company (ранее BaseGroup Labs ООО «Аналитические технологии»).

К основным функциям платформы Loginom относятся следующие:

- проведение сложных расчетов – это процесс математического анализа данных и выполнения сложных вычислений, используемых для решения задач в различных областях, таких как финансы, наука, инженерия и т.д.,
- консолидация данных – это процесс объединения данных из различных источников в единую структуру данных для анализа и использования в бизнес-процессах,

- очистка данных - это процесс проверки, обновления, удаления и коррекции неправильных, неактуальных или поврежденных значений в наборе данных,

- интеграция с различными хранилищами данных - это процесс связывания нескольких различных источников данных вместе для получения полной информации и ее дальнейшего анализа. Интеграция данных позволяет более эффективно использовать данные, которые хранятся в разных системах и форматах,

- структурный и объектно-ориентированный подходы к моделированию - это два различных подхода к описанию и проектированию сложных систем.

Loginot полностью построен на базе веб-технологий, а веб-интерфейс этой платформы обладает целым рядом значительных преимуществ.

Первое и, пожалуй, самое важное преимущество веб-интерфейса - это доступность. Доступность веб-интерфейса означает, что вы можете получить доступ к системе из любого устройства, имеющего доступ к Интернету и имеющего браузер в своей оснастке. Это обеспечивает беспрепятственную работу не только на рабочих машинах, но и на мобильных устройствах и планшетах.

Второе преимущество - наличие всей информации на одной платформе и свобода в настройке. Веб-интерфейс позволяет объединить в себе информацию из различных источников, таких как базы данных, учетные системы, социальные сети и т.д. Это дает возможность получить детальный анализ всех аспектов бизнес-процессов на одной платформе. Кроме того, платформа Loginot предлагает мощные средства настройки и создания отчетов, что позволяет настроить систему под свои нужды [15].

Третье преимущество веб-интерфейса – это многофункциональность. Браузер может выполнить ряд функций, которые были бы неподъемными для более традиционного программного обеспечения. Веб-интерфейс Loginot, например, позволяет создавать диаграммы, графики, отчеты и дашборды. Все

эти функции позволяют пользователю получить всю необходимую информацию в удобном для анализа и понимания виде.

Четвертое преимущество - это гибкость. Веб-интерфейс Loginom позволяет пользователям работать из любой точки мира, а также дают возможность гибко настраивать и перенастраивать систему в соответствии с потребностями и изменениями в компании. Веб-платформа Loginom имеет открытые API и может быть интегрирована с другими системами, с чем не всегда могут справиться традиционные ПО [16].

В целом, веб-интерфейс платформы Loginom обладает множеством преимуществ, которые делают его удобным, доступным, гибким и многофункциональным инструментом для управления бизнес-процессами любой сложности.

В том числе и то, что эта платформа обладает пользовательским интерфейсом, не требующим для работы специальной подготовки. Loginom имеет широкие возможности для анализа: от простой логики до машинного обучения [17].

3.2 Разработка требований к решению задачи автоматизации

Разрабатываемая система должна обеспечивать возможность попарного сравнения ключевых слов документов, сопровождающих бизнес процессы, и функциональных обязанностей должностных лиц. Полученные в результате данные являются основой для предложений по оптимизации документопотоков и бизнес-процессов организации. Для решения этой задачи применяются методы операций над множествами, а также SQL [18].

Входными данными для работы системы являются ключевые слова документов, сопровождающих бизнес-процессы организации.

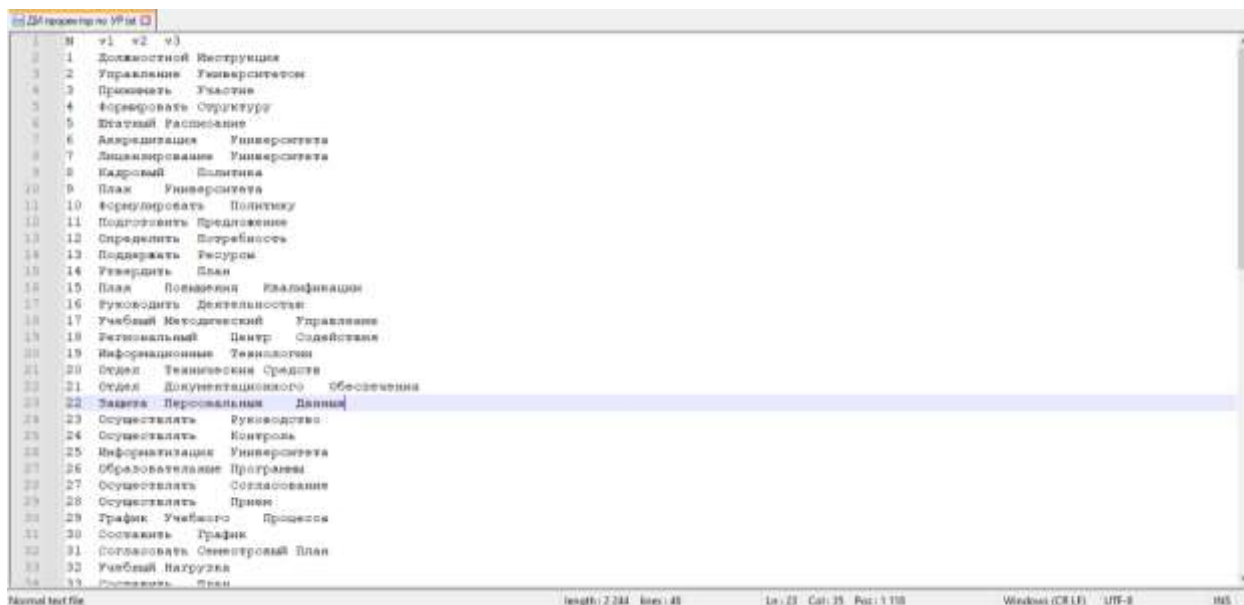
Автоматизация предполагает попарное сравнение наборов ключевых слов.

Выходными данными работы системы являются наборы ключевых слов, полученных в результате выполнения операций над входным набором данных.

Для проверки корректности работы системы предполагается проведение тестирования на опытных данных.

3.3 Подготовка исходных данных

Исходные данные должны быть оформлены в виде текстовых файлов (рисунок 1).



N	v1	v2	v3
1	Должностной Инструкции		
2	Управления Университетов		
3	Приказывать Участие		
4	Формировать Структуру		
5	Штатный Расписание		
6	Аккредитация Университета		
7	Лицензирование Университета		
8	Кадровый Политика		
9	План Университета		
10	Формулировать Политику		
11	Подготавливать Предложение		
12	Определять Потребности		
13	Поддерживать Ресурсы		
14	Утверждать План		
15	План Повышения Квалификации		
16	Руководить Деятельностью		
17	Учебный Методический Управление		
18	Региональный Центр Содействия		
19	Информационные Технологии		
20	Судей Технические Средства		
21	Судей Документационного Обеспечения		
22	Задача: Персональные Данные		
23	Осуществлять Руководство		
24	Осуществлять Контроль		
25	Информатизация Университета		
26	Образовательные Программы		
27	Осуществлять Согласование		
28	Осуществлять Прием		
29	Трафик Учебного Процесса		
30	Составить График		
31	Согласовать Сменотурный План		
32	Учебный Нагрузка		
33	Составить План		

Рисунок 1 – Исходные данные

Проблема поиска ключевых слов суть задача общей лингвистики. Существует ряд программных средств, которые направлены на анализ текстов. Некоторые из них представлены, например, на сайте Русской виртуальной библиотеки.

Многие программы анализа и лингвистической обработки текстов нацелены на другие задачи, такие как построение базы ключевых слов для индексации сайтов и повышения эффективности поиска информации в интернет, выполнения смысловых запросов, с учётом наиболее часто встречающихся терминов (например, TextAnalyst) или средства анализа текстов для копирайтеров (например, TextusPro) [19].

Многие из них в результате выполнения анализа формируют перечень абсолютно всех одиночных слов текста или не приводят слова к общей грамматической форме. Однако при анализе должностных инструкций (регламентов), положений и других нормативных документов придётся иметь

дело и с одиночными словами, и со словосочетаниями, так как необходимо описывать направления деятельности организации, которые зачастую невозможно свести к одному слову. Нужно также отметить исключительную важность формализации функциональных обязанностей должностных лиц, которые позволят определить степень участия сотрудников в каждом бизнес-процессе предприятия, а, следовательно, и степень участия в подготовке документа, что и должно быть выявлено посредством анализа ключевых слов.

В практике лингвистических исследований, особенно выполняемых документоведами, не имеющими специальной подготовки в области информационных технологий, «рутинные» операции лингвистического анализа автоматизированы очень слабо. К таким операциям могут быть отнесены:

- создание картотек/конкордансов словоупотреблений с указанием частотности,
- морфологический разбор словоупотреблений и построение частотного списка лексем (а не словоформ) для исследуемого текста,
- выделение грамматически связанных сочетаний слов, расположенных как контактно, так и дистантно.

Каждая из этих задач, так или иначе, решается с помощью применения корпусных методов без создания представительного корпуса текстов языка (или подъязыка).

Старшим редактором ООО «Технотранс» (г. Северодвинск) Барановым А.Г. в рамках проекта MINICORPUS разработаны специальные средства, облегчающие выполнение некоторых задач частотного, морфологического, семантического и дистрибутивного лингвистического анализа художественных и специальных текстов на русском языке, к которым относятся и тексты документов. Все эти материалы выложены на персональном сайте Баранова А.Г. [5] для свободного использования. В основе этих материалов положена методика, основанная на использовании

утилиты xMarkup, парсера mystem (компания Яндекс) и СУБД Microsoft Access.

Утилита xMarkup, предназначена для выполнения процедур строковых преобразований во множестве текстовых файлов. Эти процедуры могут быть сколь угодно сложными в отличие от обычных процедур поиска и замены текста. Фактически, в утилите реализован процедурный язык, на котором можно описать любые алгоритмы преобразования данных. Однако, в большинстве случаев, использование процедурных расширений является излишним. Пользователю достаточно определить маркеры начала и конца искомых элементов текста и задать шаблоны их преобразования. Утилита поддерживает кодировки ANSI и UTF-8 [6].

Данная утилита может успешно использоваться для генерации, редактирования или удаления любых текстовых элементов в исходных файлах. Утилита xMarkup первоначально разрабатывалась как любительская программа, однако сейчас продолжает развиваться и служит для решения серьёзных прикладных задач. Основные цели использования утилиты:

- генерация или редактирование навигационных гиперссылок в совокупности HTML-документов;
- анализ и преобразование структуры и/или содержимого SGML/XML/HTML-документов;
- редактирование текстовых файлов в пакетном режиме;
- разрезка и склейка текстовых файлов по заданным правилам;
- анализ и обработка текста;
- генерация скриптов.

Утилита xMarkup реализована в виде консольного 32-бит приложения для Windows и написана на языке Icon. Этот язык доступен также для ряда других систем. Для удобства пользования программа имеет удобный графический интерфейс.

Программа mystem производит морфологический анализ текста на русском языке. Для слов, отсутствующих в словаре, порождаются гипотезы.

Существует версия программы, предназначенная для некоммерческого использования.

В результате выполнения анализа предложенного текста можно получить полный список входящих в него лемм, из которого можно выделить, например, существительные, прилагательные и глаголы, как носителей основной семантики.

3.4 Автоматизация алгоритма попарного сравнения наборов ключевых слов

Осуществляем импорт файлов по документам и должностным лицам. Далее необходимо провести аналог операции внутреннего соединения. Сценарий представлен на рисунке 2.

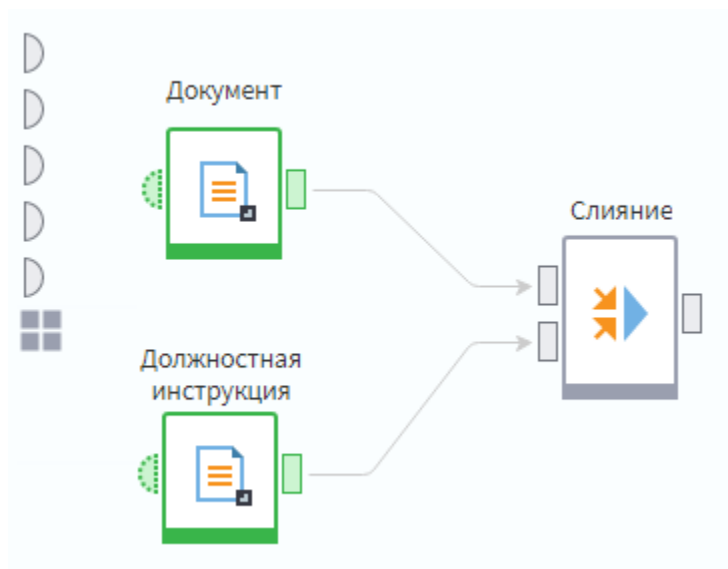


Рисунок 2 – Сценарий попарного сравнения ключевых слов

Настраиваем узел слияния для вычисления пересечений множеств ключевых слов (рисунок 3).

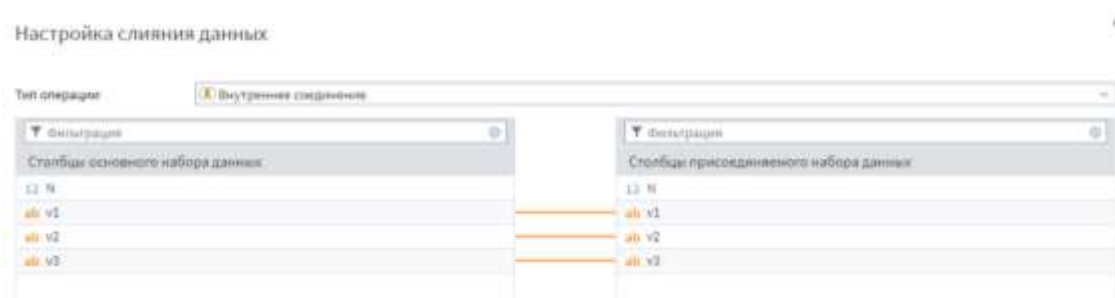


Рисунок 3 – Настройка узла слияния данных

3.5 Выводы

В рамках этого раздела были достигнуты следующие результаты: проведено глубокое исследование предметной области систем интеллектуального анализа данных, лингвистического анализа текстов, принято решение о выборе средства автоматизации алгоритма попарного сравнения множеств ключевых слов, разработан сценарий на основе модели анализа.

4 Тестирование автоматизированного решения сравнительного анализа устойчивых словосочетаний и ключевых слов текстов документов

4.1 Описание тестовых данных

Для проведения тестирования был выбран процесс «Итоговая аттестация» Пензенского государственного университета «Итоговая государственная аттестация». Диаграмма процесса представлена на рисунке 4. Связующими элементами являются документы.

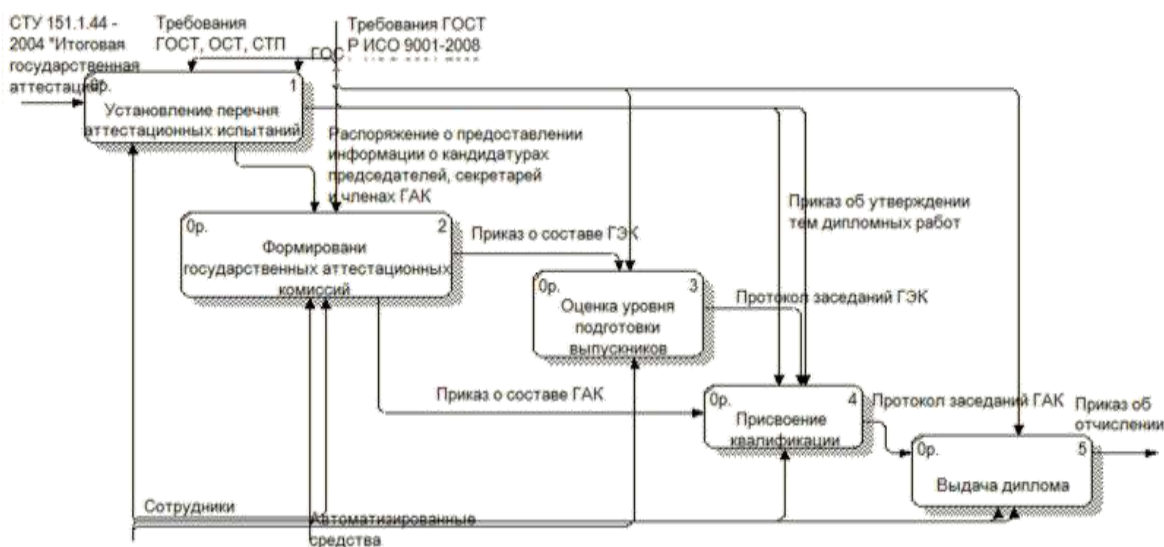


Рисунок 4 - Детализация процесса «Итоговая государственная аттестация»

4.2 Составление списка документов процесса итоговой аттестации

Произведем структурирование процесса «Итоговая государственная аттестация». Для этого проанализируем стандарт университета

Выделим основные этапы (действия), выполняемые при реализации процесса. Очевидно, что данные этапы должны быть связаны с процессом напрямую, значит, к ним относятся следующие термины: оценка уровня подготовки выпускника, аттестационные испытания, государственная аттестационная комиссия, государственная экзаменационная комиссия, присвоение квалификации, выдача диплома.

Сформулируем названия этапов и расставим их в хронологической последовательности:

- установление перечня аттестационных испытаний,
- формирование государственных аттестационных комиссий,
- формирование государственных экзаменационных комиссий,
- оценка уровня подготовки выпускника,
- присвоение квалификации,
- выдача диплома.

Определим входные и выходные данные для каждого этапа.

В состав аттестационных испытаний входят выпускная квалификационная работа и итоговый государственный экзамен. В стандарте указано, что данные испытания выбираются на основе государственного стандарта об образовании, следовательно, данный стандарт будет относиться к входным данным. На выходе должен быть создан приказ об утверждении тем дипломных работ (рисунок 5).

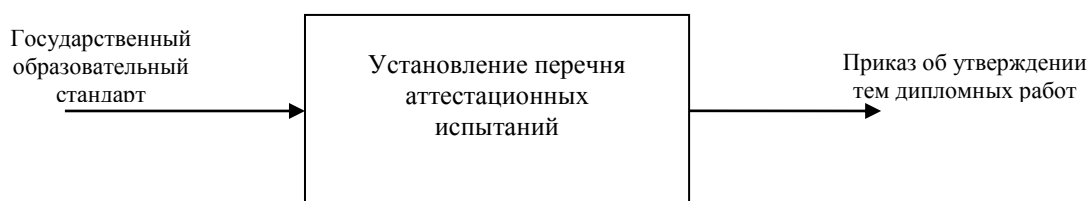


Рисунок 5 - Детализация этапа Установление перечня аттестационных испытаний

Следующим рассмотрим этап формирования государственных аттестационных комиссий. На этом этапе формируются аттестационные комиссии, определяется их состав. Соответственно, входными данными для реализации данного этапа является распоряжение о предоставлении информации о кандидатурах председателей, секретарей и членах ГАК. На выходе должны быть созданы такие документы, как приказ о составе ГАК, приказ о рецензентах (в данном случае рецензентами являются сотрудники, приглашенные из других организаций) (рисунок 6).

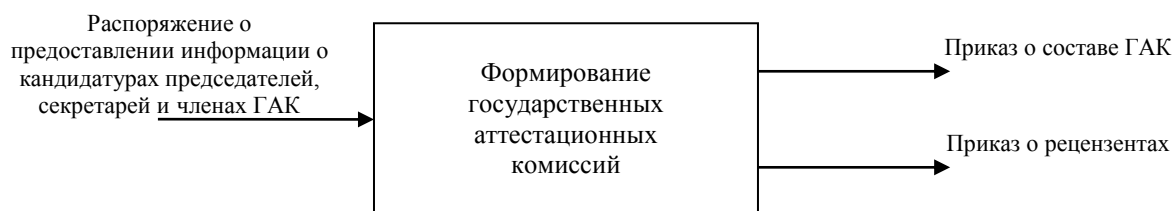


Рисунок 6 - Детализация этапа Формирование государственных аттестационных комиссий

Одновременно с этапом формирование государственных аттестационных комиссий проводится этап формирование государственных экзаменационных комиссий. Входными данными, необходимыми для реализации этапа, также является распоряжение о предоставлении информации о кандидатурах председателей, секретарей и членах ГАК. На выходе необходимо определить конкретный состав комиссии – приказ о составе ГЭК (рисунок 7).

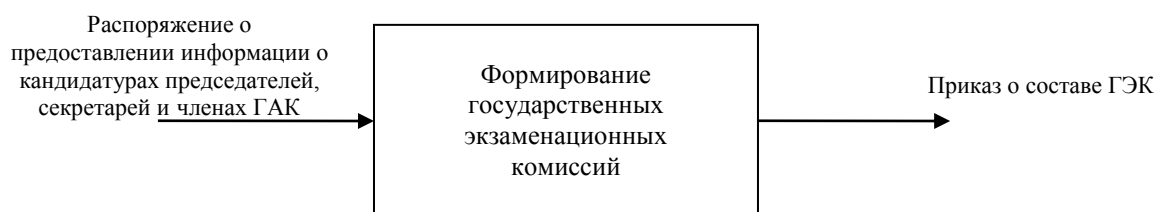


Рисунок 7 - Детализация этапа Формирование государственных экзаменационных комиссий

Исходя из данных, представленных в семантической сети понятий, видно, что сформированные экзаменационные комиссии осуществляют проведение следующего этапа – оценка уровня подготовки выпускников. Выпускников оценивают по результатам сдачи аттестационных испытаний, решение принимается на заседаниях комиссий и фиксируется в протоколах заседаний, следовательно, выходными данными данного этапа являются протоколы заседаний государственных экзаменационных комиссий (рисунок 8).

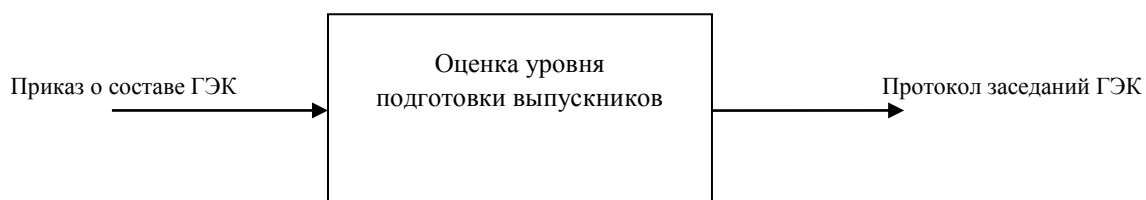


Рисунок 8 - Детализация этапа Оценка уровня подготовки выпускников

Проанализируем следующий этап - присвоение квалификации. На основании предоставленной квалификационной работы государственной аттестационной комиссией принимается решение о присвоении квалификации выпускнику по соответствующему направлению. Очевидно, что входными данными будет являться приказ об утверждении тем дипломных работ. Решение о присвоении квалификации также фиксируется в протоколах заседаний аттестационных комиссий. Следовательно, данные документы будут являться выходными данными (рисунок 9).

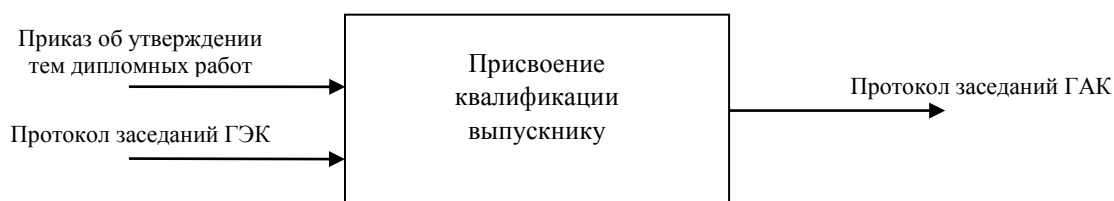


Рисунок 9 - Детализация этапа Присвоение квалификации выпускнику

Заключительным этапом в реализации процесса «Итоговая государственная аттестация» является выдача диплом. Решение о получении выпускником диплома о высшем профессиональном образовании принимается на основании протокола заседаний ГАК. На основании полученных данных делается вывод, какой из видов диплома должен получить выпускник. Данные предоставляются в официальном виде приказом об отчислении – данный документ играет роль выходных данных (Рисунок 10).

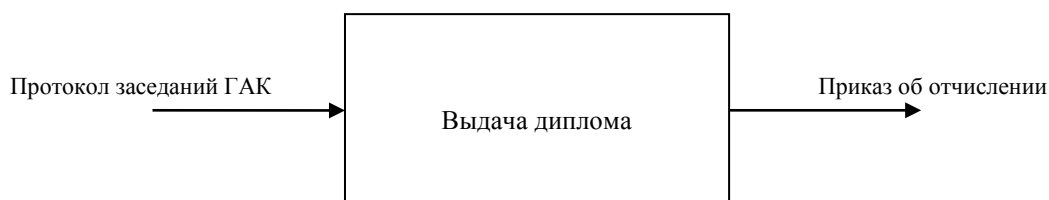


Рисунок 10 - Детализация этапа Выдача диплома

Таким образом, на основании семантической сети понятий выделены основные этапы реализации процесса итоговой аттестации.

Термины, выделенные в процессе анализа текстов документов и включенные в семантическую сеть, неоднократно встречаются в тексте стандарта и употребляются в различных формах. Следовательно, для повышения эффективности анализа текстов документов целесообразно создание словаря синонимов (таблица 3). Данный словарь поможет провести более подробный сравнительный анализ устойчивых словосочетаний и ключевых слов.

Таблица 3 – Словарь синонимов

Индекс	Синоним	Дескриптор
1	Аттестационные испытания	Итоговые аттестационные испытания
2	Итоговая аттестация	Итоговая государственная аттестация
3	Государственная аттестация	Итоговая государственная аттестация
4	Государственный экзамен	Итоговый государственный экзамен
5	Экзамен	Итоговый государственный экзамен
6	Итоговый экзамен	Итоговый государственный экзамен
7	Дипломная работа	Выпускная квалификационная работа
8	Дипломный проект	Выпускная квалификационная работа
9	Диплом	Выпускная квалификационная работа
10	Стандарт	Государственный образовательный стандарт
11	Образовательный стандарт	Государственный образовательный стандарт

4.3 Выделение ключевых словосочетаний из нормативных документов, участвующих в реализации процесса

Проанализируем стандарт университета «Итоговая государственная аттестация», для определения информативности его текста. Выделим из текста документа устойчивые словосочетания, содержащиеся в разделе Порядок проведения итоговой аттестации (таблица 4).

Таблица 4 - Словарь устойчивых словосочетаний процесса Итоговая аттестация выпускников

Индекс	Часть речи	Часть1	Часть2	Часть3
1	A	Итоговый	Аттестационный	Испытание
2	A	Государственный	Экзамен	
3	A	Выпускной	Квалификационный	Работа
4	A	Итоговый	Государственный	Аттестация
5	V	Подготовить	Условия	
6	V	Проводить	Консультация	
7	A	Основной	Образовательный	Программа
8	S	Направление	Подготовка	
9	A	Высший	Профессиональный	Образование
10	A	Государственный	Образовательный	Стандарт
11	A	Государственный	Экзаменационный	Комиссия
12	A	Государственный	Аттестационный	Комиссия
13	S	Протокол	Заседание	
14	V	Присвоить	Выпускник	Квалификация
15	S	Выдача	Диплома	
16	A	Закрытый	Заседание	
17	A	Открытый	Заседание	
18	S	Председатель	Комиссия	
19	S	Заместитель	Председатель	Комиссия
20	S	Приложение	Диплом	
21	A	Незаконченный	Профессиональный	Образование
22	A	Теоретический	Курс	
23	A	Медицинский	Показания	
24	S	Подача	Заявления	
25	A	Ученый	Совет	Факультет
26	A	Профессиональный	Подготовка	Выпускников
27	S	Совершенствование	Качества	
28	A	Учебный	Методический	Управление
29	A	Сопроводительный	Письмо	
30	A	Федеральный	Агентство	
31	A	Выпускающий	Кафедра	
32	S	Архив	Университет	
33	A	Качественный	Состав	
34	V	Поступить	Аспирантура	
35	A	Учебный	Нагрузка	

Именно эти термины отражают содержание действий, выполняемых в рамках данного процесса, и тем или иным образом должны присутствовать в должностных инструкций участвующих в процесс сотрудников, а также в тексте создаваемых документов.

Произведем анализ информативности текста стандарта университета «Итоговая аттестация выпускников». Выделение устойчивых словосочетаний

позволяет оценить информационный объем документа как отношение количества ключевых слов к общему количеству слов в документе.

Рассмотрим данные, приведенные в таблице 3 и разделе «Порядок проведения итоговой аттестации» стандарта. Исключим повторно встречающиеся словосочетания в тексте документа.

Анализ таблицы 3 показывает, что общее количество слов в словосочетаниях равно 87, а количество слов в разделе стандарта Итоговая аттестация выпускников 628, из них повторно встречающихся слов в ключевых словосочетаниях – 192. В результате расчёта процентного соотношения получим 19,9.

Вывод: полученные данные показывают, что текст раздела «Порядок проведения итоговой аттестации» стандарта содержит большое количество слов, не относящихся к процессу. Следовательно, текст данного раздела является информативно неточным и может быть воспринят не верно. Таким образом, должностные лица, отвечающие за проведение итоговой аттестации, могут допустить ряд серьезных ошибок при реализации данного процесса.

Рассмотрим должностную инструкцию проректора по УР как одного из важнейших участников процесса итоговой аттестации. Для этого выделим из текста документа устойчивые словосочетания, определяющие основные функции в деятельности проректора по УР (рисунок 11).

Если принимать каждое слово устойчивого словосочетания (рисунок 11) за ключевое слово, то можно рассмотреть соотношение количества слов из словосочетаний к количеству слов в разделе должностные обязанности должностной инструкции проректора по УР. Выбор для анализа раздела должностные обязанности объясняется следующим соображением. Любые документы являются результатом реализации некоторых действий в рамках существующих в организации бизнес-процессов. Эти действия, в свою очередь, выполняются должностными лицами и рассматриваются организацией как должностные обязанности или их компоненты и закреплены в разделе Должностные обязанности.

Набор данных				
#	12 N	ab v1	ab v2	ab v3
1	1	Должностной	Инструкция	
2	2	Управление	Университетом	
3	3	Принимать	Участие	
4	4	Формировать	Структуру	
5	5	Штатный	Расписание	
6	6	Аккредитация	Университета	
7	7	Лицензирование	Университета	
8	8	Кадровый	Политика	
9	9	План	Университета	
10	10	Формулировать	Политику	
11	11	Подготовить	Предложение	
12	12	Определить	Потребность	
13	13	Поддержать	Ресурсы	
14	14	Утвердить	План	
15	15	План	Повышения	Квалификации
16	16	Руководить	Деятельностью	
17	17	Учебный	Методический	Управление
18	18	Региональный	Центр	Содействия
19	19	Информационные	Технологии	
20	20	Отдел	Технических	Средств
21	21	Отдел	Документационного	Обеспечения
22	22	Защита	Персональных	Данных
45	23	Осуществлять	Руководство	

Рисунок 11 - Словарь устойчивых словосочетаний
должностных обязанностей проректора по УР

Подсчет слов произведем с помощью БД minicorpus. Анализ таблицы на рисунке 11 показал, что общее количество слов в словосочетаниях равно 107, а количество слов в разделе должностной инструкции равно 337. В результате расчёта процентного соотношения получим 31,75.

Вывод: полученные данные показывают, что текст раздела должностные обязанности должностной инструкции проректора по УР содержит большое количество слов, не относящихся к основным функциям проректора по УР. Следовательно, текст данного пункта необходимо

конкретизировать, для того чтобы минимизировать риски по неточному восприятию текста должностным лицом.

Далее проанализируем должностную инструкцию первого проректора. Для этого необходимо выделить из текста документа устойчивые словосочетания, определяющие основные функции в деятельности первого проректора (рисунок 12).

Набор данных				
#	12 N	ab v1	ab v2	ab v3
1	1	Управление	Университетом	
2	2	Общий	Структура	Университет
3	3	Штатный	Расписание	
4	4	Кадровый	Политика	
5	5	Административный	Управленческий	Персонал
6	6	Профессорский	Преподавательский	Состав
7	7	Учебный	Вспомогательный	Персонал
8	8	План	Университета	
9	9	Формулировать	Политика	Университет
10	10	Область	Качества	
11	11	Готовить	Предложения	
12	12	Установление	Надбавок	
13	13	Оплата	Труда	Сотрудникам
14	14	Поддерживать	Ресурсы	
15	15	Утверждать	План	Работ
16	16	План	Повышение	Квалификация
17	17	Годовой	План	
18	18	Самооценка	Подразделений	
19	19	Координировать	Работа	
20	20	Учебный	Работа	
21	21	Научный	Исследовательский	Деятельность
22	22	Воспитательный	Работа	
53	23	Административный	Хозяйственный	Работа

Рисунок 12 - Словарь устойчивых словосочетаний должностных обязанностей первого проректора

Произведем анализ текста должностной инструкции первого проректора на его информативность. Рассмотрим данные приведенные на рисунке 12 и разделе должностные обязанности должностной инструкции.

Произведем подсчет слов (подсчет слов производился с помощью БД minicorpus). Анализ данных показал, что общее количество слов в словосочетаниях равно 128, а количество слов в разделе должностной инструкции равно 342. Расчёт процентного соотношения показал 37,43.

Вывод: полученные данные показывают, что текст раздела должностные обязанности должностной инструкции первого проректора содержит большое количество слов, не относящихся непосредственно к основным функциям первого проректора. Следовательно, текст данного раздела необходимо конкретизировать для уменьшения риска неточного восприятия текста должностным лицом.

4.4 Сравнительный анализ устойчивых словосочетаний и ключевых слов текстов распорядительных документов

Рассмотрим приказ о составе экзаменационной комиссии. Приведем слова из текста приказа к словарной форме и проанализируем их (рисунок 13).

1	Приказ		
2	Состав		
3	Государственный	Аттестационный	Комиссия
4	Проведение		
5	Государственный	Итоговый	Аттестация
6	Год		

Рисунок 13 - Список слов из приказа о составе экзаменационной комиссии

В таблице содержится список слов, составляющих текст приказа. Слова выстроены в том же порядке, что и в тексте приказа, и приведены к словарной форме. Данные слова могут быть использованы для выявления устойчивых словосочетаний, которые будут отражать содержание данного документа. Следовательно, можно произвести анализ текста на информационную полноту.

Данный приказ согласовывается с проректором по УР, начальником УМУ, главным бухгалтером, начальник ПУ. Необходимо проверить целесообразность проставления данных виз согласования. Для этого

необходимо сравнить слова из текста приказа с устойчивыми словосочетаниями из должностных инструкций соответствующих должностных лиц.

В результате проведения сравнительного анализа текста приказа и устойчивых словосочетаний должностной инструкции проректора по УР было выявлено совпадение: государственная аттестация (рисунок 14).

Выходной набор данных					
#	12 N	ab v1	ab v2	ab v3	12 N
1	5	Государственный	Аттестация		35

Рисунок 14 – Результат сравнения ключевых слов приказа и должностной инструкции

Вывод: проректор по УР в своей деятельности осуществляет координирование составления плана по проведению государственной аттестации выпускников (в данном случае словосочетания государственная аттестация и государственная аттестация выпускников являются синонимами), следовательно, проставление визы согласования на данном приказе целесообразно.

Аналогичным образом проанализируем все документы, участвующие в данном бизнес-процессе (итоговая аттестация выпускников), которые содержат визу согласования проректора по УР.

Выполним анализ текста приказа о рецензентах. Данные представлены на рисунке 15.

По результатам анализа документов приказа и должностной инструкции проректора по УР совпавших или синонимичных словосочетаний не выявлено (рисунок 16).

Набор данных					
#	12 N	ab v1	ab v2	ab v3	
1	1	Приказ			
2	2	Утверждение			
3	3	Рецензент			
4	4	Дипломный	Работа		
5	5	Специальность			
6	6	Документоведение и ДОУ			
7	7	Учебный	План		

Рисунок 15 – Список слов из приказа о рецензентах

Выходной набор данных					
#	12 N	ab v1	ab v2	ab v3	12 N

Рисунок 16 - Результат сравнения ключевых слов приказа и должностной инструкции

Вывод: данные анализа указывают на то, что проставление визы согласования проректором по УР на данном приказе не целесообразно, но в тексте должностной инструкции указано, что проректор координирует процесс «присвоения выпускникам квалификации и выдаче диплома о высшем образовании». Следовательно, данные полученные из контекста распорядительного документа, указывают на необходимость проставления визы согласования данным должностным лицом.

Аналогичным образом необходимо проанализировать текст приказа о направлении на преддипломную практику, так как он содержит визу согласования проректора по УР (рисунок 17).

Набор данных					
#	12 N	ab v1	ab v2	ab v3	
1	1	Приказ			
2	2	Преддипломный	Практика		
3	3	Основание			
4	4	Учебный	План		
5	5	Специальность			
6	6	Договор			
7	7	Практика			
8	8	Предприятие			
9	9	Организация			
10	10	Студент			
11	11	Курс			
12	12	Факультет			
13	13	Руководитель			
14	14	Заведующий	Кафедра		

Рисунок 17 - Список слов из приказа о направлении на преддипломную практику

В результате проведения сравнительного анализа текста приказа и устойчивых словосочетаний должностной инструкции проректора по УР совпадения не выявлены (рисунок 18).

Выходной набор данных					
#	12 N	ab v1	ab v2	ab v3	12 N

Рисунок 18 - Результат сравнения ключевых слов приказа и должностной инструкции

Вывод: проставление визы согласования проректором по УР на данном приказе при текущем состоянии организационных документов не является необходимым. Следовательно, необходимо внести соответствующие коррективы в должностную инструкцию проректора по УР или исключить

этап согласования приказа с проректором по УР из маршрута прохождения данного документа. Решение по этому вопросу должно приниматься экспертами, в роли которых выступают руководители высшего звена.

Кроме того, необходимо отметить, что в соответствии с должностной инструкцией проректор по УР должен согласовывать документы, относящиеся к производственной практике. Если рассматривать словосочетания преддипломная практика и производственная практика, то в случае, если распорядительный документ содержит информацию по бакалаврам и магистрам, эти термины будут рассматриваться как синонимы и тогда визу проректора по УР можно рассматривать как необходимую.

Аналогичным образом необходимо проанализировать текст приказа об утверждении тем дипломных работ, так как он содержит визу согласования проректора по УР (рисунок 19).

Набор данных				
#	12 N	ab v1	ab v2	ab v3
1	1	Приказ		
2	2	Утверждение		
3	3	Тема		
4	4	Дипломный	Работа	
5	5	Руководитель		
6	6	Дипломный	Проектирование	
7	7	Студент		
8	8	Очный		
9	9	Отделение		
10	10	Факультет		
11	11	Специальность		

Рисунок 19 - Список слов из приказа об утверждении тем дипломных работ

Произведем сравнительный анализ слов из текста Приказа и словосочетаний, отображающих основные функции проректора по УР (рисунок 11).

В результате проведения сравнительного анализа текста приказа и устойчивых словосочетаний должностной инструкции проректора по УР совпадения не выявлены.

Вывод: данные анализа указывают на то, что проставление визы согласования проректором по УР на данном приказе не целесообразно. Однако в тексте должностной инструкции указано, что проректор координирует процесс «присвоения выпускникам квалификации и выдачи диплома о высшем образовании». Это означает, что проректор координирует все действия, связанные с выдачей диплома. Следовательно, данные полученные из контекста распорядительного документа, указывают на необходимость проставления визы согласования данным должностным лицом.

Необходимо произвести анализ слов теста приказа об отчислении, также содержащего визу согласования проректора по УР (рисунок 20).

Набор данных					
#	12 N	ab v1	ab v2	ab v3	
1	1	Приказ			
2	2	Отчисление			
3	3	Основание			
4	4	Решение			
5	5	Государственный	Аттестационный	Комиссия	
6	6	Университет			
7	7	Студент			
8	8	Очный	Форма	Обучение	
9	9	Факультет			
10	10	Учебный	План		
11	11	Специальность			
12	12	Защитивший			
13	13	Дипломный	Работа		
14	14	Квалификация			
15	15	Выдать			
16	16	Диплом			

Рисунок 20 - Список слов из приказа об отчислении

Сравнительный анализ документов показал, что совпадений по словосочетаниям не выявлено.

Вывод: по данным анализа, проставление визы согласования проректором по УР на данном приказе не требуется. Обратимся к контексту распорядительного документа. В нем сказано, что проректор осуществляет «согласование предложений по приёму, выпуску, отчислению и восстановлению на учёбу студентов и слушателей университета», следовательно, проставление визы согласования на данном приказе целесообразно.

4.5 Выводы

Данные, полученные в ходе тестирования сферы применения методики анализа текстов документов, помогли проанализировать полноту и избыточность комплекта документов, сопровождающих один из производственных процессов Пензенского государственного университета «Итоговая государственная аттестация». Использование данных, позволило ранее сформулировать основные этапы процесса. Затем была произведена детализация каждого этапа, что позволило определить входные данные, необходимые для реализации каждого из этих этапов, а также состав документов, необходимых на каждом этапе работы [20].

Анализ текстов нормативных документов показал, что они содержат неполное описание процесса итоговой аттестации. Так, в них отсутствует описание некоторых действий (этапов), фактически реализуемых в ходе процесса. Комплект документов, фактически создаваемых в процессе итоговой аттестации, полностью отражает его содержание.

Таким образом, применение рассматриваемой методики анализа документопотоков делает возможным регулирование объема документооборота за счет исключения из него документов, не относящихся к рассматриваемому процессу.

Сравнительный анализ ключевых словосочетаний и ключевых слов из распорядительных документов позволяет определять рациональность

существующих маршрутов прохождения документов. Исследования показали, что ряд документов проходят процедуру согласования с должностными лицами, в обязанности которых формально не входят функции, отображенные в документах. Таким образом, можно сделать вывод о том, что локальные документы отчасти носят формальный характер и не отображают полной картины деятельности должностного лица.

Вывод: рассмотренная методика является эффективной, так как позволяет достигнуть основных целей исследования документопотоков организации, а именно:

- повышение эффективности управленческой деятельности,
- ускорение движения документов в организации,
- уменьшение трудоемкости обработки документов.

Применение методики позволяет выявить недостатки комплекта нормативных документов и сформулировать предложения по совершенствованию текстов нормативов.

ЗАКЛЮЧЕНИЕ

Одной из главных целей предприятий и организаций является эффективное функционирование и максимизация прибыли. Для этого они применяют различные методы и стратегии, например, оптимизируют производственные процессы, управляют расходами и обеспечивают высокий уровень качества продукции или услуг. Но это невозможно без оптимальной организации документооборота.

С увеличением объемов и сложности документов, переход к новым формам и способам их обработки, необходимо проводить периодический анализ и оптимизацию документопотоков, чтобы устранить узкие места, уменьшить время и сократить затраты на обработку документов. Также, оптимизация документооборота позволяет повысить качество обслуживания клиентов, ускорить принятие решений и повысить эффективность работы сотрудников. Поэтому, анализ и оптимизация документопотоков являются важными этапами в повышении конкурентоспособности организации на рынке.

В дипломной работе рассмотрена методика анализа документопотоков организации, основанная на применении методики анализа ключевых слов документов.

В рамках данной дипломной работы произведен анализ существующих методов анализа информационных потоков и предложена уникальная методика анализа ключевых слов документов.

В работе приведены результаты тестирования методики, проведенного в ходе дипломного проектирования и позволившего доказать ее эффективность для решения задачи анализа документопотоков организации. Кроме того, предложены новые инструменты, позволяющие расширить сферу применения методики анализа текстов документов.

В ходе дипломного проектирования были выполнены следующие исследования:

– выделены ключевые слова внутренних нормативных документов, регламентирующих выполнение процесса итоговой аттестации в Пензенском государственном университете. Для анализа были использованы стандарт университета, регламентирующий процесс прохождения итоговой аттестации выпускников и должностные инструкции сотрудников заинтересованных в реализации рассматриваемого процесса,

– выделены ключевые слова управленческих документов, сопровождающих процедуру итоговой аттестации,

– реализована автоматизация исследования ключевых слов документов, включённых в бизнес-процесс,

– выполнено поэтапное сравнение этапов бизнес-процесса, отражающих содержание отдельных действий процесса итоговой аттестации, с ключевыми словами управленческих документов, выполняющих роль входных и выходных данных для каждого действия процесса, позволило определить полноту и избыточность комплекта создаваемых документов,

– сравнительный анализ ключевых слов управленческих документов и ключевых словосочетаний должностных инструкций сотрудников, участвующих в подготовке управленческих документов, позволил оценить реальную необходимость проставляемых виз согласования и тем самым проверить оптимальность маршрутов прохождения документов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Акулич, И.Л. Математическое программирование в примерах и задачах [Электронный ресурс] : учеб. пособие — Электрон. дан. — Санкт-Петербург : Лань, 2011. — 352 с. — Режим доступа: <https://e.lanbook.com/book/2027> — Загл. с экрана..
2. Кузнецов, А.В. Высшая математика. Математическое программирование [Электронный ресурс] : учеб. / А.В. Кузнецов, В.А. Сакович, Н.И. Холод. — Электрон. дан. — Санкт-Петербург : Лань, 2013. — 352 с. — Режим доступа: <https://e.lanbook.com/book/4550>.
3. Громницкий В.С. Методы оптимизации. Курс лекций: Учебное пособие – Нижний Новгород: Изд-во Нижегородского госуниверситета, 2010. – 104 с.
4. Барсегян, А.А. Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненков, И.И. Холод. – Санкт-Петербург : БХВ-Петербург, 2012. – 335 с.
5. Паклин Н. Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям /Н.Б.Паклин, В.И. Орешков – СПб.: Питер, 2013. – 704 с.
6. Практикум по анализу данных на компьютере / И.А. Кацко, Н.Б. Паклин /под ред. Г.В. Гореловой – М.: КолосС, 2009. – 278 с.
7. Рафалович, В. Data Mining, или интеллектуальный анализ данных для занятых. Практический курс / В. Рафалович. – М.: SmartBook, 2018. – 352 с.
8. Официальный сайт компании Loginom Company [Электронный ресурс]. – Режим доступа: <https://loginom.ru/>.
9. Энциклопедия по Бизнес-анализу [Электронный ресурс] режим <https://wiki.loginom.ru>
10. Loginom – Руководство пользователя [Электронный ресурс]. – Режим доступа: <https://help.loginom.ru/userguide/>.
11. Баранов, Александр. Моделирование применения корпусных методов для локальных лингвистических исследований [Электронный

ресурс] / Доклады, принятые к публикации на сайте международной конференция по компьютерной лингвистике «Диалог». – М., 2010. – Электрон. дан. – Режим доступа: <http://www.dialog-21.ru/digests/dialog2010/materials/html/Baranov.htm>. – Загл. с экрана.

12. Методика лингвистического анализа текста с помощью xMarkup, mystem и minicorpus.mdb [Электронный ресурс]/ Баранов А.Г.; ООО «Технотранс» – Северодвинск, 2010 . – Электрон. дан. – Режим доступа: <http://minicorpus.narod.ru/files/guide.htm>. – Загл. с экрана.

13. О программе mystem [Электронный ресурс] : Морфологический анализатор текста на русском языке / ООО «ЯНДЕКС» – [2011] . – Электрон. дан. – Режим доступа: <http://company.yandex.ru/technologies/mystem/>. – Загл. с экрана.

14. Русская виртуальная библиотека (РВБ) [Электронный ресурс]. – [2005–] . – Электрон. дан. – Режим доступа: <http://www.rvb.ru/>. – Загл. с экрана.

15. Усманова И.В., Баканова Н.Б. Метод ключевых слов как инструмент решения задачи реинжиниринга документопотоков организации // Известия ПГПУ им. В.Г. Белинского. 2012. № 30. С. 302-309.

16. Усманова И.В., Баканова Н.Б. Оптимизация маршрутов прохождения документов на основе анализа ключевых слов их текстов // Materiály VIII mezinárodní vědecko - praktická konference «Dny vědy - 2012». - Díl 86. Moderní informační technologie: Praha. Publishing House «Education and Science» s.r.o - 80 stran – str. 47-52.

17. Утилита xMarkup Преобразования Текста [Электронный ресурс] : Утилита для выполнения процедур строковых преобразований во множестве текстовых файлов / Логичев С.В. – [2012] . – Электрон. дан. – Режим доступа: <http://xmarkup.sourceforge.net/ru/>. – Загл. с экрана.

18. Усманова И.В., Забродина Л.В. Взаимодействие интеллектуальных компонентов информационного обеспечения

автоматизированной системы анализа и оптимизации документооборота организации – Известия ПГПУ им. В.Г. Белинского. № 30 - 2012. с. 310-318.

19. Садовников В.И., Эпштейн П.Л. Потоки информации в системах управления. – М.: энергия, 1973. – 240с. Садовников В.И., Эпштейн П.Л. Потоки информации в системах управления. – М.: энергия, 1973. – 240с.

20. Усманова И.В. Сетевые модели представления знаний для решения задачи анализа документооборота организации – Известия ПГПУ им. В.Г. Белинского. № 26 – 2011- С. 412-417.

ПРИЛОЖЕНИЕ А

Презентация

ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ
КАФЕДРА «ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА»

АВТОМАТИЗАЦИЯ АНАЛИЗА ТЕКСТОВ ДОКУМЕНТОВ ОРГАНИЗАЦИИ С ЦЕЛЬЮ ОПТИМИЗАЦИИ ПОТОКОВ ДОКУМЕНТОВ

ВЫПОЛНИЛ: СТ. ГР. 21ВВВМ1
БАКАНОВА Н.Б.
РУКОВОДИТЕЛЬ: Д.Т.Н., ПРОФЕССОР КАФ. ВТ
ЗИНКИН С.А.

МЕТОДИКА АНАЛИЗА ДОКУМЕНТОПОТОКОВ ОРГАНИЗАЦИИ

- Составление списка документов процесса
- Выделение ключевых словосочетаний нормативных документов, регламентирующих выполнение бизнес-процесса
- Выделение ключевых словосочетаний документов, сопровождающих бизнес-процесс
- Сравнительный анализ устойчивых словосочетаний и ключевых слов текстов документов, сопровождающих бизнес-процесс

СОСТАВЛЕНИЕ СПИСКА ДОКУМЕНТОВ ПРОЦЕССА

- **Определение бизнес-процесса и его этапов.** Для этого нужно раздробить процесс на отдельные этапы, определить последовательность действий и уточнить, какие документы необходимы для каждого этапа
- **Построение структуры документов.** Проанализировав каждый этап бизнес-процесса, необходимо составить список документов, которые потребуются на каждом этапе. Это может быть список документов, которые нужно заполнить, передать, проверить, подписать и т.д.
- **Группировка документов.** Составив список документов, необходимо определить их взаимосвязь и группировать их по типу, содержанию или функциональному назначению.
- **Проверка и уточнение списка документов.** После завершения первых трех этапов необходимо провести анализ структуры документов и уточнить ее. В процессе анализа могут выявиться новые типы документов, которые нужно добавить в список, а также проясниться некоторые вопросы по тому, какие документы нужны на каждом этапе
- **Внесение изменений в список.** После проведения анализа и уточнения структуры документов, необходимо внести изменения в список. В конечном итоге каждый этап бизнес-процесса будет иметь свой список документов

ВЫДЕЛЕНИЕ КЛЮЧЕВЫХ СЛОВСОЧЕТАНИЙ НОРМАТИВНЫХ ДОКУМЕНТОВ, РЕГЛАМЕНТИРУЮЩИХ ВЫПОЛНЕНИЕ БИЗНЕС-ПРОЦЕССА

- Анализ бизнес-процессов можно начать с внутренних стандартов предприятия
- Внутренние стандарты – это документы, содержащие правила и процедуры для выполнения различных операций внутри предприятия. Они определяют, как должны быть выполнены определенные задачи, а также устанавливают нормы и стандарты для качества выполнения этих задач.

ВЫДЕЛЕНИЕ КЛЮЧЕВЫХ СЛОВСОЧЕТАНИЙ ДОКУМЕНТОВ, СОПРОВОЖДАЮЩИХ БИЗНЕС-ПРОЦЕСС

- Как и в предыдущем этапе, здесь необходимо составить таблицы ключевых слов, определяющих тексты документов, сопровождающих бизнес-процесс.
- Каждый конкретный бизнес-процесс может иметь свои уникальные документы и формы для заполнения, но указанные ключевые слова представляют общую базу для классификации документов бизнес-процесса.
- Такие таблицы должны быть сформированы для всего состава документов процесса.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ УСТОЙЧИВЫХ СЛОВСОЧЕТАНИЙ И КЛЮЧЕВЫХ СЛОВ ТЕКСТОВ ДОКУМЕНТОВ, СОПРОВОЖДАЮЩИХ БИЗНЕС-ПРОЦЕСС

- Для анализа применяются операции над множествами (объединение, пересечение, вычитание). В качестве элементов множеств рассматриваются ключевые слова различных документов
- Анализ могут подвергаться:
 - документы системы менеджмента качества организации;
 - инструкции (в том числе должностные);
 - положения о структурных подразделениях; документы, образовавшиеся в оперативной деятельности организации

ВЫБОР ПЛАТФОРМЫ ДЛЯ РЕШЕНИЯ ЗАДАЧИ АВТОМАТИЗАЦИИ

- Loginom предоставляет возможность проводить глубокий анализ данных с использованием различных методов, от простого отображения данных до сложной моделирования и машинного обучения. Это позволяет пользователям получать точные и достоверные данные для принятия управленческих решений.

ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ

- Проблема поиска ключевых слов суть задача общей лингвистики. Существует ряд программных средств, которые направлены на анализ текстов. Некоторые из них представлены, например, на сайте Русской виртуальной библиотеки.