



Министерство науки и высшего образования Российской Федерации  
федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа – Инженерная школа информационных технологий и робототехники  
Направление подготовки – 09.04.01 Информатика и вычислительная техника  
ООП/ОПОП – Data Science in Software Engineering / Анализ данных в разработке программного обеспечения  
Отделение школы (НОЦ) – Отделение информационных технологий

### ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРАНТА

Тема работы
Программа автоматического построения онтологической модели предметной области на основе анализа текстовых корпусов

УДК 004.032.26:81

Обучающийся

Группа	ФИО	Подпись	Дата
8ВМ1И	Михайлов Никита Игоревич		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Губин Е.И	к.ф.-м.н		

Научный консультант

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель ОИТ	Кайда А.Ю			

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент (ОСГН, ШБИП)	Спицына Л.Ю	к.э.н		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент (ООД, ШБИП)	Антоневич О.А	к.б.н		

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП, должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент (ОИТ, ИШИТР)	Зарубин А. Г	к.х.н		

## ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ООП

по направлению 09.04.01 Информатика и вычислительная техника

Код компетенции	Наименование компетенции
Универсальные компетенции	
УК(У)-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий
УК(У)-2	Способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели
УК(У)-4	Способен применять современные коммуникативные технологии, в том числе на иностранном (-ых) языке (-ах), для академического и профессионального взаимодействия
УК(У)-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия
УК(У)-6	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки
Общепрофессиональные компетенции	
ОПК(У)-1	Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте
ОПК(У)-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач
ОПК(У)-3	Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

<b>Код компетенции</b>	<b>Наименование компетенции</b>
ОПК(У)-4	Способен применять на практике новые научные принципы и методы исследований
ОПК(У)-5	Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
ОПК(У)-6	Способен разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования
ОПК(У)-7	Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий
ОПК(У)-8	Способен осуществлять эффективное управление разработкой программных средств и проектов
<b>Профессиональные компетенции</b>	
ПК(У)-1	Способен разрабатывать и администрировать системы управления базами данных
ПК(У)-2	Способен к созданию программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов
ПК(У)-3	Способен управлять процессами и проектами по созданию (модификации) информационных ресурсов
ПК(У)-4	Способен осуществлять руководство разработкой комплексных проектов на всех стадиях и этапах выполнения работ
ПК(У)-5	Способен проектировать и организовывать учебный процесс по образовательным программам с использованием современных образовательных технологий



Министерство науки и высшего образования Российской Федерации  
федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский Томский политехнический университет» (ТПУ)

Инженерная школа информационных технологий и робототехники  
Направление подготовки – 09.04.01 Информатика и вычислительная техника  
ООП/ОПОП – Data Science in Software Engineering / Анализ данных в разработке  
программного обеспечения  
Отделение информационных технологий

УТВЕРЖДАЮ:  
Руководитель ООП

\_\_\_\_\_ А.Г. Зарубин  
(Подпись) (Дата) (Ф.И.О.)

**ЗАДАНИЕ  
на выполнение выпускной квалификационной работы**

В форме:

магистерской диссертации

Студенту:

Группа	ФИО
8ВМ1И	Михайлову Никите Игоревичу

Тема работы:

Программа автоматического построения онтологической модели предметной области  
на основе анализа текстовых корпусов

Утверждена приказом директора (дата, номер)

№ 138-44/с от 18.05.2023

Срок сдачи студентом выполненной работы:

14.09.2023г.

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ:**

**Исходные данные к работе**

*(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду; энергозатратам; экономический анализ и т. д.).*

1. Текстовые корпуса из источника <https://sports.ru>
2. Использование библиотеки синтаксического анализа natasha
3. Анализ текстовых корпусов на наличие онтологических связей

<p><b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b></p> <p><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ol style="list-style-type: none"> <li>1. Анализ научных источников</li> <li>2. Обзор методов исследования</li> <li>3. Проектирование и разработка программы автоматического построения онтологических моделей</li> <li>4. Финансовый менеджмент</li> <li>5. Социальная ответственность</li> <li>6. Раздел на английском языке</li> </ol>
--	---

<p><b>Перечень графического материала</b> <i>(с точным указанием обязательных чертежей)</i></p>	<ol style="list-style-type: none"> <li>1. Диаграмма Ганта</li> </ol>
---	--

**Консультанты по разделам выпускной квалификационной работы**  
*(с указанием разделов)*

Раздел	Консультант
«Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»	Спицына Л. Ю., Доцент (ОСГН, ШБИП), к.э.н
«Социальная ответственность»	Антоневич О. А., Доцент (ООД, ШБИП), к.б.н.
Английский язык	Ануфриева Т. Н., Старший преподаватель (ОИЯ, ШБИП)

**Названия разделов, которые должны быть написаны на русском и иностранном языках:**


Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	05.09.2022 г.
--	---------------

**Задание выдал руководитель:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Губин Е.И	к.ф.-м.н		05.09.2022

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ВМ1И	Михайлов Никита Игоревич		05.09.2022



Министерство науки и высшего образования Российской Федерации  
федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа инженерная школа информационных технологий и робототехники  
Направление подготовки (ООП/ОПОП) 09.04.01 Информатика и вычислительная техника  
Отделение школы (НОЦ) Отделение информационных технологий  
Период выполнения \_\_\_\_\_ (осенний / весенний семестр 2021/2022 учебного года)

**КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН  
выполнения выпускной квалификационной работы**

Обучающийся:

Группа	ФИО
8ВМ1И	Михайлов Никита Игоревич

Тема работы:

Программа автоматического построения онтологической модели предметной области на основе анализа текстовых корпусов
--

Срок сдачи обучающимся выполненной работы:

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
14.06.2023	Основная часть ВКР, включая раздел на английском языке	70
10.06.2023	Раздел «Социальная ответственность»	15
12.06.2023	Раздел «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»	15

**СОСТАВИЛ:**

**Руководитель ВКР**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Губин Е.И	к.ф.-м.н		

**СОГЛАСОВАНО:**

**Руководитель ООП**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент (ОИТ, ИШИТР)	Зарубин А. Г.	к.х.н		

**ЗАДАНИЕ ДЛЯ  
РАЗДЕЛА  
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ,  
РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа	ФИО
8ВМ1И	Михайлову Никите Игоревичу

Институт	ИШИТР	Кафедра	
Уровень образования	магистр	Направление/специальность	09.03.01 Информатика и вычислительная техника

**Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:**

<i>1. Стоимость ресурсов научного исследования (НИ): материал но-технических, энергетических, финансовых, информационных и человеческих</i>	Стоимость ресурсов определялась по средней рыночной стоимости. Оклады в соответствии с окладами сотрудников организации.
<i>2. Нормы и нормативы расходования ресурсов</i>	районный коэффициент - 30%
<i>3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	Коэффициент отчислений на уплату во внебюджетные фонды 30%.

**Перечень вопросов, подлежащих исследованию, проектированию и разработке:**

<i>1. Оценка количественного и инновационного потенциала НТИ</i>	Проведение предпроектного анализа
<i>2. Разработка устава научно-технического проекта</i>	Разработка устава научно-технического проекта
<i>3. Планирование процесса управления НТИ. структура и график проведения, бюджет, риски и организация закупок</i>	Разработка плана управления НТИ
<i>4. Определение ресурсной, финансовой, экономической эффективности</i>	1. Определить интегральный финансовый показатель разработки; 2. Определить интегральный показатель ресурсоэффективности разработки; 3. Определить интегральный показатель эффективности;

**Перечень графического материала (с точным указанием обязательных чертежей):**

<ol style="list-style-type: none"> <li><i>1. Сравнение конкурентных технических решений</i></li> <li><i>2. Матрица SWOT</i></li> <li><i>3. Перечень продолжительности работ</i></li> <li><i>4. Оценка экономической эффективности НИ</i></li> <li><i>5. Бюджет НТИ</i></li> <li><i>6. Оценка ресурсной эффективности НТИ</i></li> </ol>	
---	--

Дата выдачи задания для раздела по линейному графику	
--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень,	Подпись	Дата
Доцент ОСГН, ШБИП	Спицына.Л.Ю	Канд.экон.наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ1И	Михайлов Н.И		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа		ФИО	
8ВМ1И		Михайлов Никита Игоревич	
Школа	ИШИТР	Отделение (НОЦ)	
Уровень образования	магистратура	Направление/специальность	09.03.01 Информатика и вычислительная техника

Тема ВКР:

Программа автоматического построения онтологической модели предметной области на основе анализа текстовых корпусов	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
<p><b>Введение</b></p> <ul style="list-style-type: none"> <li>– Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика) и области его применения.</li> <li>– Описание рабочей зоны (рабочего места) при разработке проектного решения/при эксплуатации</li> </ul>	<p><u>Объект исследования процесс разработки программы автоматического построения</u></p> <p><i>Область применения:</i> поисковые системы, педагогика, бизнес</p> <p><i>Рабочая зона:</i> <u>офис</u></p> <p><i>Размеры 3м x 4м</i></p> <p><i>помещения</i> <span style="margin-left: 150px;"><i>климатическая</i></span> <span style="margin-left: 100px;"><i>зона*)</i></span></p> <hr/> <p><i>Количество и наименование оборудования рабочей зоны</i> <u>ПЭВМ</u></p> <p><i>Рабочие процессы, связанные с объектом исследования, осуществляющиеся в рабочей зоне:</i> <u>формирование дорожных карт, оптимизация поисковых запросов</u></p>
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
<p><b>1. Правовые и организационные вопросы обеспечения безопасности при разработке проектного решения</b></p> <ul style="list-style-type: none"> <li>– специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства;</li> </ul>	<p>1) Трудовой Кодекс Российской Федерации</p> <p>2) Федеральный закон от 28 декабря 2013 г. N 426-ФЗ "О специальной оценке условий труда.</p> <p>3) ГОСТ Р 50948-2001. Средства отображения информации индивидуального пользования. Общие эргономические требования и требования</p>

	<p>безопасности.</p> <p>4) ГОСТ 12.2.033-78 «Рабочее место при выполнении работ сидя»</p>
<p><b>2. Производственная безопасность при разработке проектного решения</b></p> <ul style="list-style-type: none"> <li>– Анализ выявленных вредных и опасных производственных факторов</li> <li>– Расчет уровня опасного или вредного производственного фактора</li> </ul>	<p><b>Вредные факторы</b></p> <ul style="list-style-type: none"> <li>- Опасные и вредные производственные факторы, связанные с аномальными микроклиматическими параметрами воздушной среды на месте работающего</li> <li>- Производственные факторы, обладающие свойствами психофизиологического воздействия на организм человека, а именно нервно-психические перегрузки, связанные с напряженностью трудового процесса</li> <li>- Опасные и вредные производственные факторы, связанные с отсутствием и недостатком естественного освещения</li> </ul> <p><b>Опасные факторы:</b></p> <ul style="list-style-type: none"> <li>- Поражение электрическим током</li> </ul> <p><b>Расчёт</b></p> <p>Расчёт системы искусственного освещения</p>
<p><b>3. Экологическая безопасность при разработке проектного решения</b></p>	<p><b>Воздействие на селитебную зону:</b> нет</p> <p><b>Воздействие на литосферу:</b> твёрдые отходы оборудования ПЭВМ (утилизация ПЭВМ и периферийных устройств (принтеры, веб-камеры, наушники)</p> <p><b>Воздействие на гидросферу:</b> нет</p> <p><b>Воздействие на атмосферу:</b> нет</p>
<p><b>4. Безопасность в чрезвычайных ситуациях при разработке проектного решения</b></p>	<p><b>Возможные ЧС:</b> Пожар</p> <p><b>Наиболее типичная ЧС:</b> Пожар</p>

Дата выдачи задания для раздела по линейному графику	
--	--

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Антоневич Ольга Алексеевна	Кандидат биологических наук		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ВМ1И	Михайлов Никита Игоревич		

## РЕФЕРАТ

Выпускная квалификационная работа содержит: 122 с., 46 рис., 21 табл., 71 источник., 1 прил.,

Объектом исследования является процесс разработки программы автоматического построения онтологической модели предметной области на основе анализа текстовых корпусов.

Цель работы – разработка программы автоматического построения онтологической модели предметной области на основе анализа текстовых корпусов.

В результате исследования были найдены эффективные методы для построения онтологических моделей и оценки их эффективности, а также спроектировано и разработано программное обеспечение для автоматического построения онтологических моделей предметной области на основе анализа текстовых корпусов. Источником анализа текстовых корпусов выступал сайт <https://sports.ru>. Во время разработки программного обеспечения были созданы следующие динамически подключаемые модули:

1. Модуль анализа (парсинга) текстовых корпусов с применением библиотеки BeautifulSoup.
2. Модуль распознавания именованных сущностей.
3. Модуль синтаксического анализа текстовых корпусов и построения онтологических моделей с применением библиотеки natasha.
4. Модуль статического анализа текстовых корпусов.
5. Модуль отображения результата работы программного обеспечения с применением библиотек matplotlib и numpy.

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ .....	15
ОБЪЕКТ И МЕТОДЫ ИССЛЕДОВАНИЯ.....	19
ГЛАВА 1. ОБЗОР ТЕОРЕТИЧЕСКИХ АСПЕКТОВ ПОНЯТИЯ ОНТОЛГИИ И ОНТОЛОГИЧЕСКОЙ МОДЕЛИ.....	23
1.1 Анализ термина онтологии.....	23
1.2 Методы онтологического моделирования.....	29
1.3 Методы оценки качества онтологической модели.....	35
ГЛАВА 2. ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ПОСТРОЕНИЯ ОНТОЛОГИЧЕСКОЙ МОДЕЛИ.....	37
2.1 Рассмотрение общей архитектуры программного обеспечения.....	37
2.2 Проектирование и разработка модуля анализа содержимого сайта.....	43
2.3 Проектирование и разработка модуля распознавания именованных сущностей.....	47
2.4 Проектирование и разработка модуля синтаксического анализатора текстовых корпусов.....	50
2.5 Проектирования и разработка модуля статистического анализатора текстовых корпусов.....	59
2.6 Проектирование и разработка модуля представления результата.....	65
ГЛАВА 3. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ.....	66
3.1 Предпроектный анализ.....	66
3.2 Анализ конкурентных технических решений.....	67
3.3 SWOT анализ.....	69
3.4 Оценка готовности проекта к коммерциализации.....	71
3.5 Методы коммерциализации результатов научно-технического исследования.....	73

3.6	Инициация проекта .....	73
3.6.1	Цели и результаты проекта.....	73
3.6.2	Организационная структура проекта.....	74
3.6.3	Ограничения и допущения проекта.....	74
3.7	Планирование управления научно-техническим проектом .....	75
3.8	Бюджет научно-технической разработки.....	80
3.9	Оценка сравнительной эффективности исследования.....	84
ГЛАВА 4. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ.....		89
4.1	Правовые и организационные вопросы обеспечения безопасности.....	88
4.2	Производственная безопасность.....	91
4.2.1	Опасные и вредные производственные факторы, связанные с аномальными микроклиматическими параметрами воздушной среды.....	92
4.2.2	Производственные факторы, обладающие свойствами психофизиологического воздействия на организм человека.....	94
4.2.3	Опасные и вредные производственные факторы, связанные повышенным уровнем шума на рабочем месте.....	96
4.2.4	Опасные и вредные производственные факторы, связанные с отсутствием и недостатком естественного освещения.....	98
4.3	Экологическая безопасность.....	101
4.4	Безопасность в чрезвычайных ситуациях.....	102
4.5	Выводы по разделу.....	103
ЗАКЛЮЧЕНИЕ .....		105
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ .....		106
ПРИЛОЖЕНИЕ А The program of automatic construction of an ontological model of the subject area based on the analysis of text corpora.....		114

## **ВВЕДЕНИЕ**

Современные информационные технологии сегодня содержат множество способов обработки и интерпретации данных. Бизнес-задачи требуют всё большей автоматизации производственных систем. Формализация знаний о производственном процессе является одним из наиболее важных аспектов повышения эффективности работы компании как с точки зрения адаптации к изменениям, так и с точки зрения экономического аспекта.

Одним из методов эффективного построения семантической модели предметной области являются онтологические модели. Они широко распространены в таких процессах как построение поисковых систем, систем представления знаний, а также при семантической интеграции информационных ресурсов. Под онтологией в целом понимается процесс формирования концептов в некоторой предметной области. Процесс формирования концептов является процедурой описания множества понятий и связей между ними в контексте определённой предметной области.

Онтология выступает в качестве предмета исследования уже на протяжении 20 лет. В исследовании существует несколько направлений:

- применение методов искусственного интеллекта для создания баз знаний
- использование онтологических техник для семантического структурирования разметки интернет ресурсов
- изменение сетевых информационных технологий с целью создания различных интернет ресурсов

В процессе исследования сущности онтологии существует два направления исследования, до некоторых пор развивающихся отдельно:

1. направление исследования, связанное с пониманием онтологии как формальной системы, базирующейся на математических утверждения.
2. Направление, развивающееся в контексте компьютерной лингвистики и семантических наук.

Данные направления обуславливают существование двух подходов к созданию и воспроизведению онтологических моделей: формального, основанного на представлении онтологии как сущностей математической моделей и рассматривающего их именно с точки зрения формальной логики и лингвистического, то есть в контексте лингвистических наук.

Данные подходы взаимно влияют друг на друга. Процесс поиска семантической связи обуславливает необходимость их чередования и комбинирования.

Онтологическая модель представляет собой совокупность отношений типа «Субъект + предикат + объект».

В качестве субъектов выступают экземпляры классов, то есть предметы, которые выступают активными сущностями. Обычно это существительные, либо местоимения.

Предикатами являются связи между объектом и субъектом. Они указывают на зависимость объекта от субъекта.

Класс является сущностью абстрактного типа, описывающая однотипные явления, имеющие общее строение и однородное поведение. Классы характеризуют большое количество объектов и явлений, а также устанавливают разные связи с другими классами [62].

Существует несколько стандартизированных типов связей между классами [26]:

1. Зависимый класс является частью активного класса (part of).
2. Зависимый класс является разновидностью активного класса (kind of).
3. Зависимый класс является синонимом активного класса (known as).

Сущности класса – это воплощение определённого класса, его свойств и характера. Сущности могут иметь только одну реализацию класса и описывают конкретный объект окружающего мира. Стоит отметить, что сущности могут принадлежать сразу нескольким классам.

Между сущностями существует несколько типов связей: наследственные (унаследованные от классов-родителей) и индивидуальные (определены в самой сущности).

Предикатами являются связи, установленные между сущностями.

Для более информативного отображения онтологических моделей применяют языки разметки онтологий. Существует несколько групп языков разметки онтологий:

1. Языки разметки, чья идея заключается в аннотации семантических единиц и их разметки (OWL, OIL).

2. Языки разметки, которые базируются на логике первого порядка и традиционном программировании (Common Logic, Cycl) [26].

К тому же стоит учесть, что онтологические модели также способны повысить эффективность восприятия и интерпретации потребностей стейкхолдеров.

Цель работы – разработка программы автоматического построения онтологической модели предметной области на основе анализа тестовых корпусов.

Актуальность данной работы обусловлена развитием различных технологий представления информации. Технология представления информации на основе онтологии является костяком во многих информационных системах, основанных на познании. Со временем, область применения онтологических моделей расширилась от использования их в разработке искусственного интеллекта до процесса информационной интеграции, а также в семантических веб приложениях.

Исследования по проектированию и построению онтологий и онтологических моделей ведутся по всему миру уже на протяжении долгого времени. Все усилия сконцентрированы на изучении методологии онтологического инжиниринга и построении условий для его дальнейшего использования. Однако практическое использование пока ограничено его использованием в исключительно исследовательских и научных целях.

Таким образом, не смотря на активное развитие направления исследования онтологии, на данный момент пока не существует эффективных инструментов построения онтологических моделей. Данный факт также обуславливает актуальность данной работы

Объектом исследования является процесс разработки программы автоматического построения онтологической модели предметной области на основе анализа текстовых корпусов.

Предметом исследования является онтология и онтологическая модель.

Для достижения поставленной цели необходимо выполнить следующие задачи:

1. Изучение научной литературы на предмет исследования сущности онтологии и онтологической модели, а также методов построения онтологий
2. Проектирование и написание программного обеспечения автоматического построения онтологических моделей предметной области на основе анализа текстовых корпусов.

В первой главе отражена общая теоретическая база разрабатываемого программного обеспечения, указаны методы исследования и архитектура программы, а также обусловлен выбор источника анализируемых текстовых корпусов.

Во второй главе рассматриваются теоретические аспекты онтологии и онтологической модели, классификация онтологий, онтологических связей и языки разметки, а также методология построения онтологических моделей.

В третьей главе рассматриваются архитектура и процесс разработки программного обеспечения.

В четвёртой главе отражены вопросы, связанные с ресурсоэффективностью, социальной ответственностью и ресурсосбережением.

В ходе работы были задействованы следующие методы исследования: поиск литературы, анализ информационных источников и их сравнение, методы построения онтологий и их визуализации.

## ОБЪЕКТ И МЕТОДЫ ИССЛЕДОВАНИЯ

В данной работе объектом исследования является процесс разработки программы автоматического построения онтологической модели предметной области на основе анализа текстовых корпусов.

В качестве источника текстовых корпусов используется ресурс <http://sports.ru>. Выбор данного источника обусловлен эффективностью синтаксического анализа именно данной области знаний.

Как уже было упомянуто ранее, существует множество языков разметок онтологий. В данной работе будет использоваться их графовое обозначение. В процессе проектирования и разработки программного обеспечения возникла необходимость в предварительной загрузке и обработке анализируемых текстовых корпусов. Данное мероприятие сопровождалось следующими шагами:

1. Получение HTML-страницы с ссылками на свежие статьи посредством GET запроса.
2. Анализ и отбор содержимого страницы на предмет ссылок на новостные блоги (парсинг).
3. Последовательное получение страниц новостных блогов по полученным ссылкам.
4. Анализ страниц на предмет содержимого текстовых корпусов.
5. Предварительная обработка полученных текстовых корпусов.

Следующим шагом является проведение синтаксического анализа текстовых корпусов на наличие именованных сущностей и их классификация. Именованными сущностями являются имена собственные, которые могут быть разделены на три класса: имена людей, наименования организаций, наименования локаций (городов, деревень). Также анализируются все существительные и местоимения.

Далее происходит выделение главных концептов путём их ранжирования.

Процесс ранжирования происходит при помощи оценки важности слов или словосочетаний с применением методики TF-IDF. TF-IDF является

статистическим способом оценки важности слова в контексте содержащего документа.

TF (Term Frequency) – отражает частоту появления слова в документе. Представляет собой отношение числа вхождений слова  $D$  к общему числу слов документа  $K$ .

$$TF = \frac{D}{\sum k}$$

IDF (Inverse document frequency) – представляет обратную частоту появления слова в документа. IDF призван корректировать вес широкоупотребляемых слов.

$$IDF = \frac{D}{(d_i \in D | t \in d_i)},$$

где

$D$  – Число документов в корпусе.

$(d_i \in D | t \in d_i)$  - число документов коллекции, где встречается слово  $t$

Величина TF-IDF является произведением двух множителей  $TF * IDF$  и отображается как мера важности использования определённого слова или словосочетания [67].

Следующим шагом является процесс выделения концептов высшего уровня, для чего необходимо выявить значимость слова  $w_i$

$$K(w_i) = \frac{1}{|R_1 - R_2|} * \frac{Q(w_i)}{N_{max} - N_{min}},$$

$R_1$  и  $R_2$  являются частными рангами анализируемого слова  $w_i$ , то есть значения, обратные позициям этих слов в списках для TF и TF-IDF,  $Q(w_i)$  представляет собой общее количество анализируемых текстовых корпусов, содержащих слово  $w_i$ ,  $N_{max}$  и  $N_{min}$  – наибольшее и наименьшее вхождение слова  $w_i$  в корпус текстов [67].

Для анализа смысла словосочетания производится его повторный синтаксический анализ, где выделяется главное слово.

Из ранжированных списков выделяются главные концепты, Дочерние концепты выделяются путем нахождения наибольших совпадений с главными концептами.

Установление связей между сущностями онтологии может происходить тремя путями:

1. Поиск подходящего слова на связь и дальнейший подбор связываемых концептов.
2. Определение для концепта списка вероятно подходящих слов-связок.
3. Нахождение двух концептов и последующие поиск слова-связки для них.

Недостатки первого подхода:

- в общем множестве найденных связей между концептами присутствуют несущественные или несуществующие связи;
- некоторые важные концепты предметной области не имеют связей сформированного множества с другими концептами проектируемой онтологии [67].

В данном программном обеспечении будет применён третий вариант нахождения слов-связок по причине его эффективности в контексте данных текстовых корпусов.

Далее необходимо выделить словосочетания, в которые входят именованные сущности. Для этого необходимо проанализировать предложения, содержащие именованные сущности на наличие зависимых или главных по отношению к концепту именованной сущности членов предложения в зависимости от его положения. Если слово, содержащее концепт именованной сущности, является главным (именем существительным в именительном падеже или глаголом), то необходимо проанализировать предложение на наличие второстепенных слов (имён существительных не в именительном падеже или прилагательных). В случае если слово второстепенное, нужно найти главное по отношению к этому слово (производится поиск имени существительного в именительном падеже или глагола).

После нахождения словосочетания, содержащее концепт именованной сущности нужно найти слово-связку с другой сущностью. Обычно в качестве него выступает глагол или оно опускается.

Для извлечения отношений типа «is-a», то есть отношения иерархии концептов необходимо воспользоваться методом, основанным на лексико-синтаксических шаблонах. Они задают тип отношений как «NP-это NP» и «NP является NP», где NP является словосочетанием. Также необходимо применить лексико-синтаксический анализ, и, получив синтаксическую структуру извлекать отношения.

Извлечение семантических отношений другого типа базируются на том же принципе. Шаблон типа «Adj NP», где Adj прилагательное, а NP существительное либо словосочетание позволяет получить свойства объекта. В качестве слова-связки также может выступать глагол, тогда шаблон будет представлять собой тип «NP V NV», где NP является подлежащим (главным словом), V подлежащим (глаголом), а NV зависимым словом или словосочетанием [67].

Поиск данных связей осуществляется путём синтаксического анализа отдельных предложений текстовых корпусов и последовательным выделением частей речи в тексте.

Статистические методы включают в себя процесс ранжирования слов и словосочетаний и нахождение степени их важности с применением метода TF-IDF.

# ГЛАВА 1. ОБЗОР ТЕОРЕТИЧЕСКИХ АСПЕКТОВ ПОНЯТИЯ ОНТОЛГИИ И ОНТОЛОГИЧЕСКОЙ МОДЕЛИ

## 1.1 Анализ термина онтологии

Понятие онтология происходит от греческого "онтос" - сущее, "логос" - учение, понятие, т.е. это раздел философии, изучающий бытие. Онтология определена как попытка всеобщей и детальной формализации определённой области знания с применением концептуальной схемы. Концептуальная схема подразумевает совокупность понятий и терминов, а также информацию о них. Наиболее устоявшееся определение онтологии определяется как «спецификации концептуализации».

Также по определению Тома Грубера, онтология является спецификацией концептуализации, являющейся описанием множества предметов(концептов), предметных областей и связей между ними.

В контексте семантики онтология является абстракцией, отображающей информацию в определённой предметной области.

Более широкие понятия онтологии определяются следующим образом:

1. Онтология - это эксплицитная спецификация концептуализации, где в качестве концептуализации выступает описание множества объектов предметной области и связей между ними.
2. Онтология - это знания, формально представленные на базе концептуализации. Формально онтология состоит из терминов, организованных в таксономию, их определений и атрибутов, а также связанных с ними аксиом и правил вывода.
3. Онтология - формальная спецификация разделяемой концептуализации, которая имеет место в некотором контексте предметной области.
4. Онтология - база знаний, описывающая факты, которые предполагаются всегда истинными в рамках определённого общества на основе общепринятого смысла используемого словаря.

Отличительными чертами моделей онтологии являются:

1. Использование универсальных инструментов знаний.
2. Ориентация на структурирование знаний и их хранение.
3. Невозможность применения для планирования будущих состояний предметной области.

В данной главе будут рассмотрены концепция и сущность понятия онтологической базы знаний и произведена её категоризация.

База знаний онтологии — это машинно-читаемая база знаний с формальной семантической моделью. Она может быть рассмотрена как специальная база данных, содержащая в себе информацию об определённой предметной области.

Она классифицируется следующим образом:

1. Онтология нижнего уровня (Upper level ontology) — онтология, состоящая из основного знания концептов более высокого уровня, которые не являются специфичными для определённой предметной области.
2. Онтология среднего уровня — онтология, состоящая из концептов основного знания.
3. Базовая онтология (core ontology) - онтология, состоящая из минималистичных или базовых концептов, необходимые для понимания понятий другой предметной области.
4. Онтология предметной области — онтология, содержащая понятия определённой предметной области.
5. Гибридная онтология — комбинация между онтологиями верхнего и нижнего уровней.
6. Онтология задачи - онтология, содержащая понятия, специфичные для одной задачи.

Для избежание фундаментальных противоречий процесс накопления онтологий должен:

1. Быть интегрирован в рабочие процессы и не отвлекать пользователей от их работы.

2. Процесс должен быть простым в исполнении и использовании.
3. Предоставлять исчерпывающую информацию пользователю, избавляя его от необходимости прибегать к дополнительной информации.
4. Допускать некоторые пропуски в формализации информации [47].

В данной работе преимущественно будет рассмотрена онтологическая база знаний, применимая для определённой предметной области.

Базы знаний на основе правил являются одними из баз знаний, где каждое правило может быть выражено утверждениями «If» и «Then». Она характеризуется наличием нескольких элементов условия.

Разработка и создание баз знаний на основе правил выполняется при помощи техник накопления знаний, которые могут варьироваться от простых до сложных алгоритмов. Накопление знаний является долгим процессом превращения данных низкого уровня в данные высокого уровня.

Итак, процесс проектирования системы базы знаний онтологии или так называемого «Менеджмента знаний» является одним из самых эффективных подходов к управлению сложными информационными системами.

В процессе разработки и проектирования «менеджмента знаний» выделяются следующие этапы:

1. Накопление — стихийное накопление информации.
2. Извлечение — процесс добычи информации из различных источников знаний. Является одним из самых трудных этапов создания онтологической системы базы знаний.
3. Структурирование — выработка наглядной структуры представления информации.
4. Формализация — представление структурированной информации в виде машинной обработки, то есть языках описания данных по запросу
5. Обслуживание — исправление формализованных данных, удаление или обновление устаревшей информации.

База знаний включает в себя терминологию — формально определённые термины и правила, и логические отношения между ними. Сама база знаний

описывается с помощью языков программирование, что позволяет применять различные возможности в запросе и обработке данных.

Еще Аристотель рассматривал сущность термина онтологии с точки зрения его деления на такие нематериальные объекты как предмет и понятие. Данные сущности создают фактически полный набор вопросов, ответы на которые способны воспроизвести предсказание новой сущности.

В целом, компьютерная онтология предметной области может быть формально представлена упорядоченной тройкой в виде концепт, отношение, концепт

$$O = \langle X, R, F \rangle$$

Анализ граничных случаев множеств включает в себя проверку четырех комбинаций значений  $R$  и  $F$ , а именно:  $R = \emptyset$  ;  $R \neq \emptyset$  ;  $F = \emptyset$  ;  $F \neq \emptyset$ . Это позволяет создать различные онтологические конструкции, начиная от простого словаря и таксономии, и заканчивая формальной структурой концептуальной базы знаний для высокоинтеллектуальных знание-ориентированных систем [67].

По степени полноты функциональности различают три типа онтологий:

1. Простая онтология – состоит только из одного элемента, главного объекта  $X$ , зависимый объект  $F$  и предикат  $R$  отсутствуют. Служит для однозначного восприятия.
2. Полная (строгая) онтология – характеризуется наличием всех объектов и предиката их отношения. В данной онтологии отношения между концептами максимально полны. Все компоненты описываются при помощи формального языка разметки.

Полная онтология представляет собой следующую структуру

$$O \langle X, R, F, A(D, R) \rangle$$

, где  $X$  – множество концептов;  $R$  – множество концептуальных отношений между ними,  $F : X \times R \rightarrow \text{bool}$  – конечное множество функций интерпретации, заданных на концептах или отношениях;  $A$  – конечное множество аксиом, которые используются для записи всегда истинных

высказываний (определений и ограничений);  $D$  – множество дополнительных определений понятий;  $R$  – множество ограничений, определяющих область действия понятийных структур.

3. Промежуточная (неполная) онтология представляет собой либо полное отсутствие концептов и предикатов, либо наличие лишь только зависимого объекта [67].

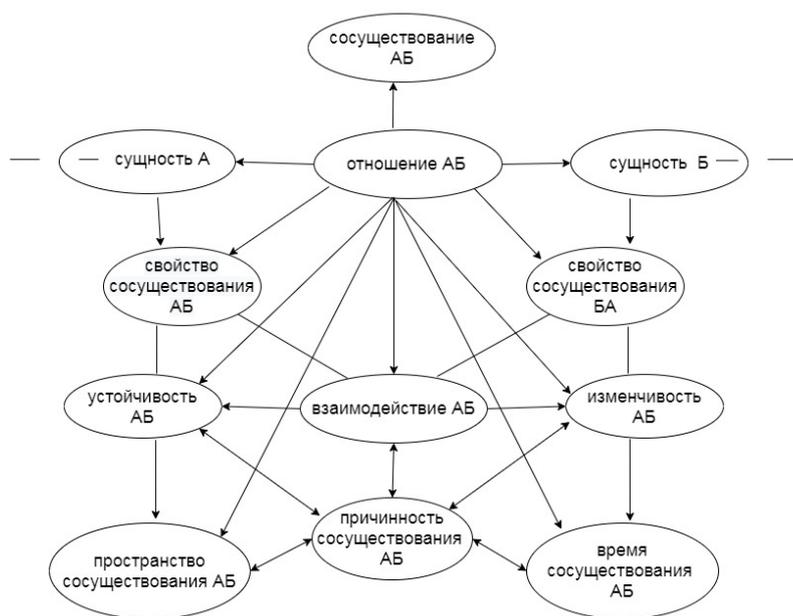


Рисунок 1 – Пример схемы онтологии

На Рисунке 1 видно, что помимо отношений, сущности также могут иметь свойства, характеризующие их. На практике свойства обычно выражаются прилагательными, наречиями или словосочетаниями, содержащими их.

Для более эффективного представления онтологической модели используется так называемая многослойная онтология. Она представляет собой разделение связи сущностей на слои, то есть части онтологии, связанные определёнными отношениями. Одним из ключевых элементов многослойной онтологии является ядро слоя.

В онтологии имеются разные вариации слоёв, обусловленные типом отношений между основой (ядром) и составными компонентами каждого слоя. Слои включают в себя:

1. Функциональный слой – получение одних слоёв при помощи других.
2. Казуальный слой – характеризуется причинно-следственной природой связей между слоями.
3. Структурный слой – составные части элементов представлены в иерархическом виде.
4. Семантический слой – элементы имеют семантическую структуру и связи [62].

При создании онтологического слоя происходит рекурсивный процесс, где начальный слой состоит из элементов, напрямую связанных с ядром. Далее новые элементы добавляются на каждом шаге, расширяя слой. Глубина рекурсии определяется пользователем и должна быть ограничена, чтобы избежать проблем с зацикленными связями и бесконечной рекурсией. В новой методике предлагается устанавливать нижний предел семантической близости элементов к ядру, при достижении которого рекурсия останавливается, вместо явного ограничения числа шагов. Это позволит создавать более качественные и эффективные онтологические слои [62].

Таким образом, онтология представляет собой формализованное описание и отображение информации определённой предметной области, содержащее концепты и тип связи между ними.

## 1.2 Методы онтологического моделирования

Как уже было отмечено ранее, онтологическая модель является концептуальным представлением какой-либо области знаний.

Для более эффективного понимания принципов онтологического моделирования необходимо ознакомиться с процессом построения модели мышления с точки зрения человека в соответствии с методикой онтологического моделирования.

В процессе построения мыслительной модели человек сначала выделяет основные объекты и даёт им наименование. Данный процесс является концептуализацией, то есть для каждого объекта был определён его понятие (концепт). В результате была установлена связь между концептом и его мысленным образом, возникающем при восприятии изображения конкретным человеком.

При этом было задействовано внутреннее определение термина, его идею, называемую интенционалом.

Процесс построения модели мышления включает в себя:

1. Предмет или объект материального мира.
2. Его понятие в сознании человека.
3. Понятие (знак), обозначающее объект.
4. Идею или интенционал.

Далее, определив объекты, человек оценивает отношения между ними. Связи обычно выражены глаголами – отношениями между субъектом и зависимым объектом. Машина едет по дороге. В данной связке выражена обычная связь, существительное машина является активным субъектом по отношению к объекту дороги. Однако если человек увидит полицейскую машину, он сразу поймёт, что на картинке изображены полицейские и они едут по делам (гонятся за преступником и т.д). В данном случае человек приходит к более высокому уровню концептуализации, сопоставляя предмет (машину) и его понятие в сознании человека [62].

Человеческое мышление характеризуется заданием ярлыков на воспринимаемые объекты и явления. Это обусловлено необходимостью адаптации к восприятию окружающего мира.

Вся информация также проходит через эмоции человека, что, несомненно, влияет на конечную картину мира.

В конечном итоге была построена мысленная модель, которая характеризуется упрощением воспринимаемой информации, процессом абстракции понятий, то есть их концептуализированностью, взаимосвязанностью элементов, событийным прогнозированием путём применения логических правил.

Далее производится декомпозиция, состоящая в разделении объекта реальности на отдельные фрагменты. Состав объекта может быть нам до конца не известен: а вдруг у полицейского за пазухой спрятан пистолет? Тогда результаты нашего мысленного эксперимента существенно изменятся, события могут пойти по совсем другому сценарию. Это вопрос, в том числе, целостности полученного нами информационного сообщения, достаточности информации для моделирования.

Глубина разложения элементов модели, определение границ объектов и принятие решения о создании или удалении объектов зависят от практической цели моделирования. Если мы используем модель для тренировки действий правоохранительных органов при задержании, то подозреваемый останется подозреваемым, независимо от того, какие повреждения ему будут нанесены. Если одежда или другие предметы могут повлиять на исход событий, то мы должны включить их в модель как отдельные объекты [30].

Оптимальное моделирование является процессом нахождения эффективного соотношения между степенью детализации модели и требованием ресурсов.

Далее необходимо произвести классификацию – соотношение каждого объекта определённый класс.

Чтобы приступить к классификации, необходимо сначала определить сам набор доступных классов. Сам класс представляет собой идею (интенционал), а

его название отражает символ, обозначающий совокупность индивидуальных объектов или других сущностей, которые определяют идею. Простейшее определение класса не включает ни признаки, относящие объекты к определённому классу, ни его отношения с другими классами, ни характеристики, которыми обладают все члены класса. Хотя вся эта информация может быть отражена в модели, например – в виде правил отнесения объектов к классам.

Процесс концептуализации, или классификации объектов нужен для построения логических выводов и формулирования знаний. Для этого нужно включить классификацию в построение логики первого порядка (все кошки животные)

Концептуализация (классификация) позволяет построить иерархию классов от общих к специфичным.

Оперируя объектами, человек придаёт им свойства. Следующим шагом в процессе построения мыслительной модели является выделение свойств объектов. Свойства объектов отражают его характеристики, основные из них которые:

1. Название свойства.
2. Ограничения на тип и диапазон значений.
3. Набор объектов, которые могут и/или должны являться носителем этого свойства.

Связи между объектами также могут выступать в качестве свойств. Таким образом, свойства разделяются на два типа:

1. Свойства, описывающие характеристики предметов, имеющие конкретные значения.
2. Свойства, отражающие связи между объектами или особенностями их взаимодействия.

На данный момент существует три пути разработки онтологии:

1. Формализация опыта действующего лица. В данном случае построение онтологии происходит на языке метамоделей.

2. Разработка онтологии происходит при наличии инфраструктуры базы знаний уже имеющихся апробированных онтологий. Построение онтологий происходит путём человеко-машинных процедур.

3. Вывод отображение онтологии из уже имеющихся данных.

Разработка и проектирование онтологических моделей с целью решения задач предусматривает создание трансформационных объектных моделей. - денотатами, то есть объектами, определённые целями решения задач.

Проектирование трансформационных моделей представляет собой спецификацию сценария решения задач, описание действий для достижения цели, в результате которых она должны приобрести необходимые свойства.

На этапе выделения и классификации концептов высшего уровня используются инструменты семантического анализа текста: распознавание именованных сущностей, сущностей характеристик и их роли словосочетаниях (зависимые или главные), а также определение их роли в тексте [30].

Одним из методов поиска концептов высшего уровня является метод, основанный на оценке уровня важности концептов при помощи значений TF/IDF. Сначала происходит ранжирование слов по убыванию их значимости в соответствии с показателями TF/IDF.

Значимость определённого слова  $w_i$  в тексте  $i$  рассчитывается как:

$$K(w_i) = \frac{1}{|R_1 - R_2|} * \frac{Q(w_i)}{N_{max} - N_{min}}$$

,где

$Q(w_i)$  – количество текстов, в которых содержится слово  $w_i$ .

$\max \min N$ ,  $N$  – наибольшее и наименьшее число вхождений слов  $w_i$  в корпус текстов соответственно [67].

Для создания связей между концептами необходимы ключевые термины, которые будут присутствовать в большинстве текстов корпуса и распределены равномерно в каждом из них.

Также не менее важным является правильный выбор длины словосочетаний. Её длина прямо пропорциональна вероятности присутствия словосочетаний, имеющих какой-либо смысл.

Процесс анализа словосочетаний позволил оценить влияние длины словосочетаний на вероятность осмысленности фразы.

Для полной оценки степени важности слова или словосочетания  $w_i$  необходимо найти коэффициент  $M(w_i)$

$$M(w_i) = K(w_i) * (\prod R) * \min(|R_{1j} - R_{2j}|) * n * \max(TF/IDF_j)$$

,где  $n$  – количество слов в словосочетании  $iW$  (не считая стоп-слов);  $TF/IDF$  – коэффициенты оценки важности слова;  $R_i, R_{1j}, R_{2j}$   $R$  – общий и частные ранги слова  $w_j$  из словосочетания  $W_i$  соответственно.

Данный метод оценивает степень важности концептов по коэффициенту  $M(w_i)$  и слов по коэффициенту  $K(w_i)$ .

Общий алгоритм данного метода представлен следующими шагами:

1. Ранжирование по степени важности списков слов (сущностей) и определение степени их значимости.
2. Изменение ранжирования список сущностей.
3. Формирование словосочетаний из списка ранжированных сущностей.
4. Определение степени важности словосочетаний  $M(w_i)$ .

Также стоит отметить, что для анализа смысла концептов полученные словосочетания нужно привести к одному слову для максимальной абстракции.

Словосочетание анализируется с позиции наличия главного и зависимых слов. Выделяются части речи как существительные, глаголы и т.д.

Для выделения дочерних концептов определяются слова или словосочетания с наибольшим совпадением слов с главными концептами высшего уровня. Далее выделяются связи между ними.

Существует четыре варианта представления концептов и связей между ними:

$$W_1 \uparrow L_1 \uparrow W_2$$

$$W_1 \downarrow L_1 \uparrow W_2$$

$W_1 \downarrow L_1 \downarrow W_2$

$W_1 \uparrow L_1 \downarrow W_2$

,где  $\uparrow$  - определяет слово-связку  $L_1$  как специфичную для концепта  $L_i$ , а  $\downarrow$  - слово-связка  $L_1$  не специфична для концепта  $L_i$ .

Статистический анализ текстов определяет вероятности принадлежности слов-связок [67].

Для определения актуальности рассматриваемой онтологической тройки  $\langle X, R, F \rangle$  необходимо воспользоваться формулой:

$$P_{place} = \frac{k * \left( \left( \frac{|m - n|}{\min(m, n)} + 2 \right) + 1 \right)}{n + m + 1}$$

Где  $n$  – расстояние от слова-связки  $L_1$  до первого концепта  $W_1$ ,  $m$  – расстояние слова-связки  $L_1$  до второго концепта  $W_2$ .

Таким образом, чем больше расстояние между концептом и словом-связкой, тем меньше актуальность онтологической тройки [67].

Принцип построения связей между онтологическими концептами представляет собой:

1. Определение концепта и, при необходимости его нормализация до одного слова.
2. Определение для текущего концепта потенциальных слов-связок.
3. Определение наиболее вероятной тройки онтологий путем выявления ее актуальности.
4. Включение наиболее вероятной тройки онтологий в онтологическую модель.

### 1.3 Методы оценки качества онтологической модели

Рассматривая методы оценки качества, стоит определить понятие качества как пригодность к использованию.

Согласно теории деятельности, контекст можно рассматривать как взаимодействие между общими культурными и общественными структурами (языком, нормами, конвенциями, социальными сетями и отношениями) и структурой конкретной деятельности.

Также качество онтологической модели может определяться как её характеристика, оценивающая степень пригодности для моделирования.

Следовательно, чтобы измерить качество онтологии или любого информационного объекта, необходимо понимать общий культурный контекст онтологии и сообщества, понимать контекст конкретных действий по созданию, использованию и сопровождению онтологии и определять механизмы, фиксирующие различия свойств онтологий, лежащих в их основе сущностях и соответствующих контекстуальных (культурных, деятельностных) особенностях и связях и преобразовывать их в показатели качества.

Основными критериями оценки онтологий являются:

1. Точность и полнота словаря предметной области.
2. Адекватность.
3. Восприимчивость.
4. Производительность.

Требованиями, предъявляемыми к качеству модели, являются:

1. Адекватность (соответствие) моделируемому объекту, включающая в себя её полноту и отсутствие противоречивости.
2. Обоснованность.
3. Сложность, представляющая собой как размер занимаемой памяти, так и сложность вычисления.
4. Результативность, включающая в себя достоверность модели и её точность.

5. Адаптивность к изменениям.

6. Удобство восприятия модели, как человеком, так и техникой [27].

Также важными параметрами являются ширина и глубина онтологии.

Существует три метода для расчёта глубины онтологии:

1. Абсолютная глубина – является суммой длин онтологического графа.
2. Средняя глубина – отношение абсолютной глубины к количеству рёбер онтологического графа.
3. Максимальная глубина – отражает значение максимальной длины онтологического графа [27].

Метрика ширины онтологии содержит следующие параметры:

1. Абсолютная ширина – является суммой вершин каждого графа.
2. Средняя ширина – представляет пропорцию значения ширины графа к количеству графов.
3. Максимальная ширина – является количеством вершин на наибольшем уровне [27].

Также существуют субъективные метрики оценки качества онтологии, среди которых:

1. Метрики Ингве-Миллера – для качественной онтологии её количество связей не должно превышать значения  $7 \pm 2$ .
2. Метрика циклов – более глубокое восприятие онтологической модели требует минимального количества циклов.
3. Метрика глубины – производит оценку степени сбалансированности онтологических моделей. Значения длин путей в графе не должны различаться, либо их количество должно быть минимальным [27].

## ГЛАВА 2. ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ПОСТРОЕНИЯ ОНТОЛОГИЧЕСКОЙ МОДЕЛИ

### 2.1 Рассмотрение общей архитектуры программного обеспечения

В качестве источника онтологической модели текстовых корпусов был выбран новостной сайт <https://sports.ru>.

Итак, для выполнения поставленной задачи будет использован язык программирования Python в среде разработки Jupyter-notebook. Сначала необходимо реализовать функцию, которая бы получала методом GET запроса содержимое HTML страницы. Воспользуемся библиотекой requests, которая входит в основную среду языка Python.

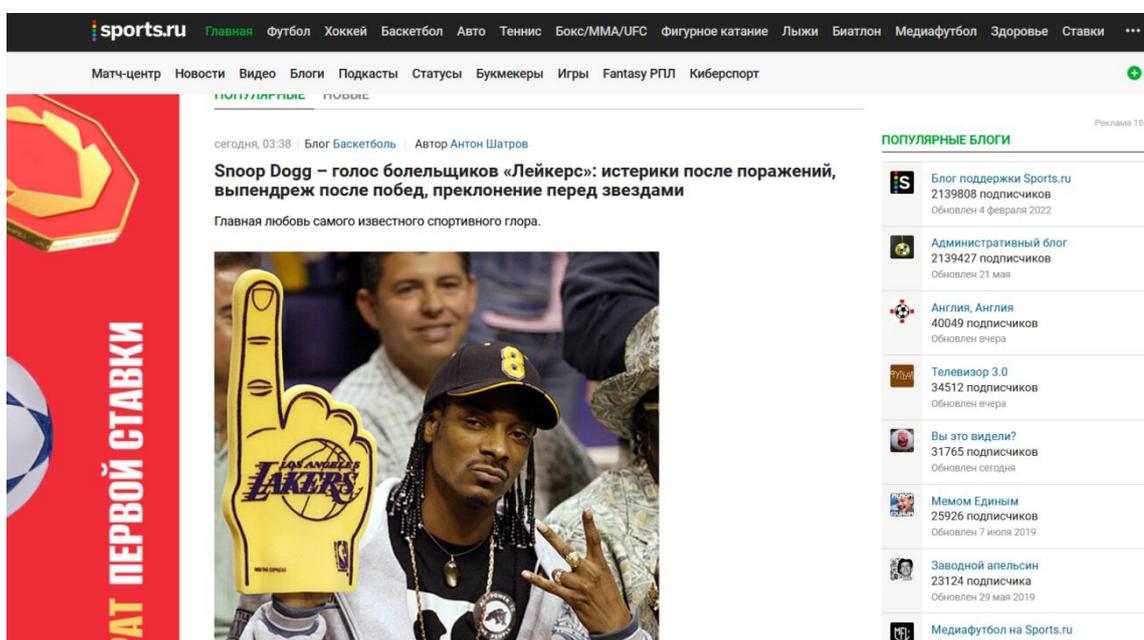


Рисунок 2 – Главная страница анализируемого сайта

Каждый модуль, то есть динамически подключаемая часть программного обеспечения разделён на функции, которые отвечают за выполнение определённой части модуля. В данном программного обеспечении было решено не использовать объектно-ориентированное программирование в виду отсутствия необходимости построения взаимозависимых частей. Каждый модуль не является в какой-либо мере зависимым от других модулей, то есть

нет надобности использовать парадигмы ООП как наследование, инкапсуляция и т.д.

При проектировании данного программного обеспечения упор делался на возможность подключения модулей, то есть частей программного обеспечения к другому программному обеспечению естественно в контексте уместности их использования.

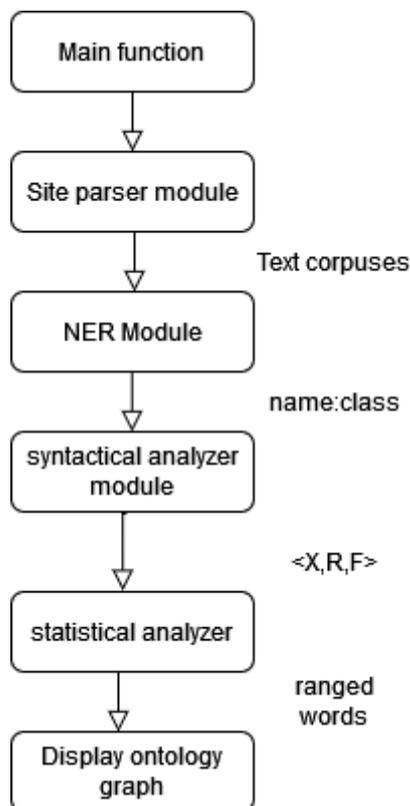


Рисунок 3 – Схема архитектуры программного обеспечения

Общая архитектура представляет собой схему использования трёх модулей главной точкой входа в программу (Рисунок 3). Модуль парсинга сайта выдаёт на выходе список обработанных текстовых корпусов. Модуль NER отвечает процесс распознавания именованных сущностей текстовых корпусов, принимая на входе текстовых корпус в строковом формате и выдавая на выходе список сущностей и их типа в виде сущность: тип.

Синтаксический анализатор проводит морфологический анализ слова на наличие главных и зависимых концептов и связей между ними, в результате получается тройка вида  $\langle X,R,F \rangle$ .

В модуле статистического анализатора происходит оценка важности и значимости слов, используя метрики TF-IDF.

Модуль отображения результата принимает на вход список в виде сущности: тип. На выходе данный модуль формально ничего не возвращает, однако он отображает схему распознанных онтологий.

В функции точки входа в программу происходит поочерёдное использование модулей, сначала используется модуль парсинга сайта, далее модуль распознавания именованных сущностей и в конечном итоге их отображение в виде таблицы. Рассмотрим более подробный план каждого модуля.

Модуль парсинга сайтов состоит из четырёх функций:

1. Функция, отвечающая за получение кода общей страницы с ссылками.
2. Функция, обрабатывающая результат предыдущей функции и получающая ссылки на отдельные новости.
3. Функция, занимающаяся обработкой текста и отделением его от HTML тегов.
4. Функция, получающая HTML код каждой страницы и обрабатывающая результат.

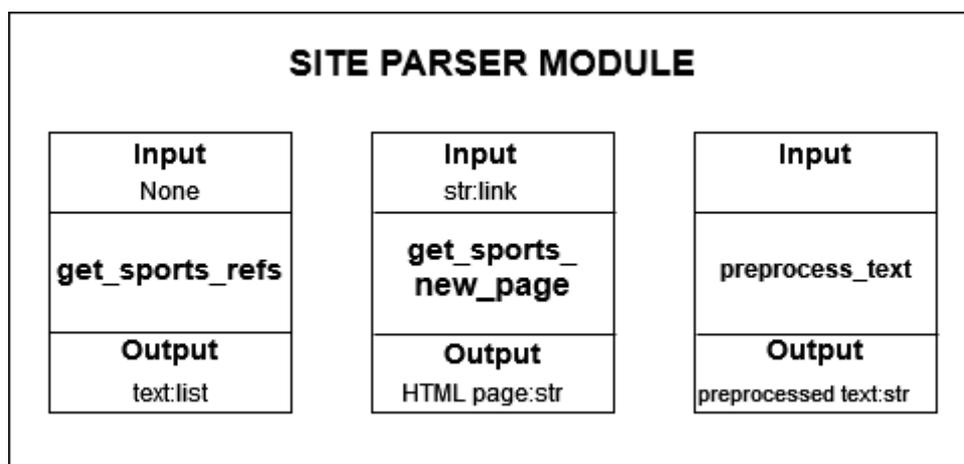


Рисунок 4 – Структура модуля парсера

Как видно из структуры, первая функция `get_sports_refs()` не принимает на входе никаких аргументов (Рисунок 4). На выходе функция возвращает ссылки в виде списка строковых переменных.

Функция `get_sports_new_page` отвечает за получение страниц новостей. На входе функция получает один аргумент – строковый формат, содержащий ссылку. На выходе функция выдаёт HTML код в строковом формате `preprocess_text` обрабатывает HTML код страницы, оставляя только читаемые человеком текст. На вход подаётся текст в виде строкового параметра, на выходе также строковый параметр текста.

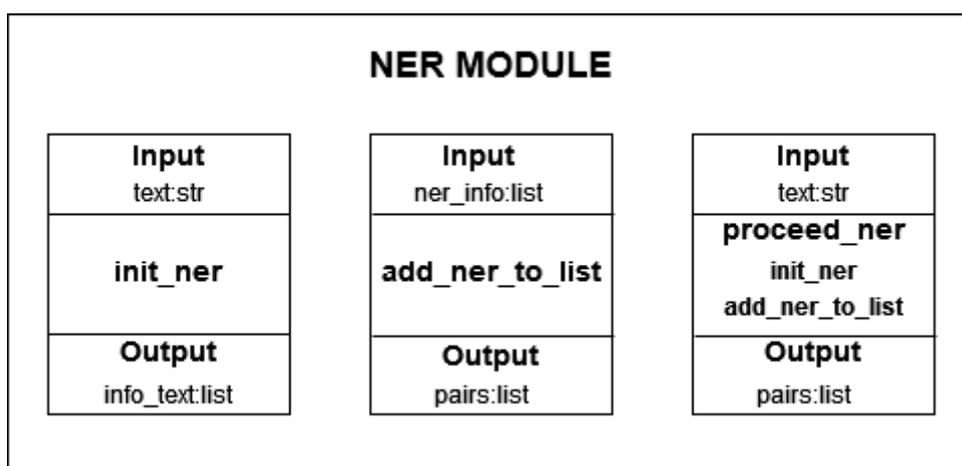


Рисунок 5 – Структура модуля обработки именованных сущностей

Модуль анализа именованных сущностей состоит из трёх функций (Рисунок 5):

1. Функция инициации анализатора.
2. Функция анализа именованных сущностей и добавление результата в список.
3. Функция, объединяющая работу предыдущих двух функций.

Модуль отображения содержимого состоит из одной функции, которая выводит поступающую информацию в виде графов (Рисунок 6).

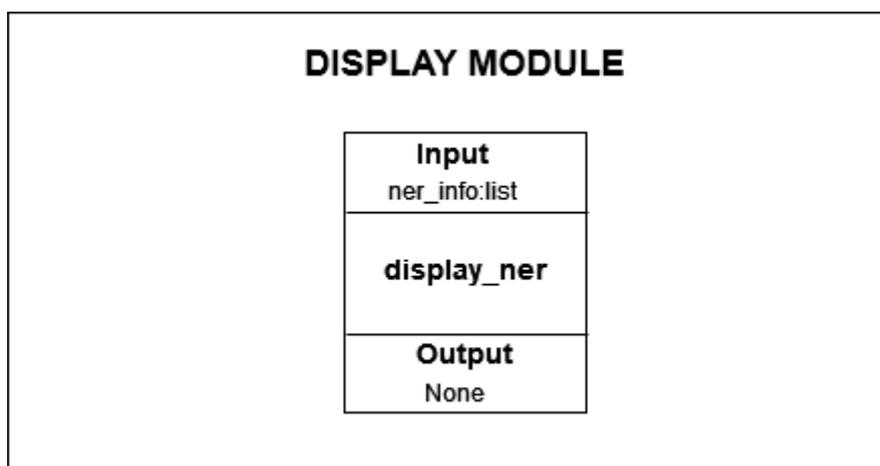


Рисунок 6 – Структура модуля отображения результатов

Модуль синтаксического анализатора отвечает за процесс анализа синтаксических связей и построение словосочетаний и онтологий. Он состоит из трёх функций:

1. Функция получения главного концепта.
2. Функция получения слова связки.
3. Функция зависимого концепта.

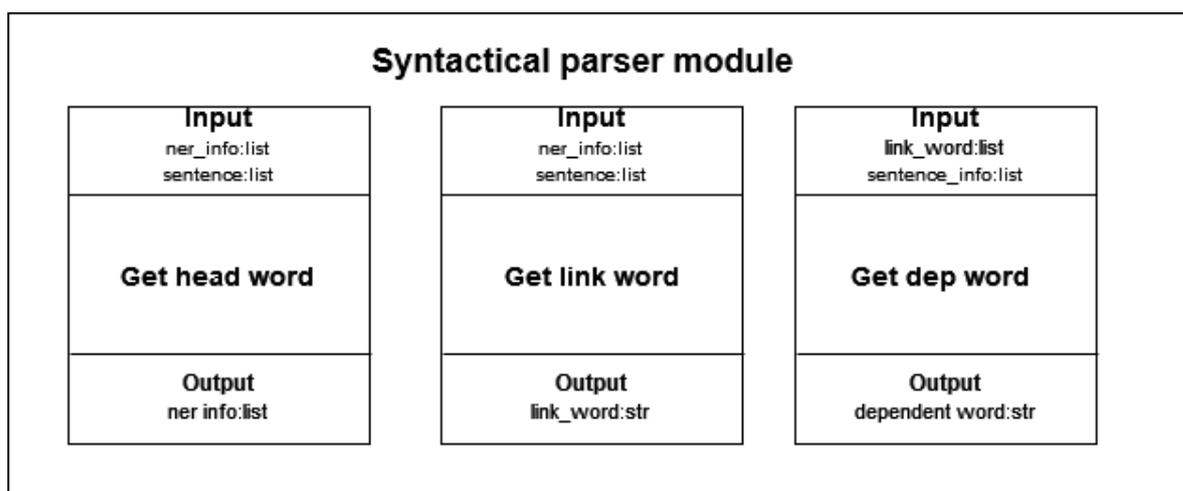


Рисунок 7 – Структура модуля синтаксического анализатора текстовых корпусов

Модуль статистического анализа онтологий проводит оценку слов всего текстового корпуса по метрикам TF-IDF, затем оценивает метрику важности каждого слова и получает метрику актуальности онтологической тройки.

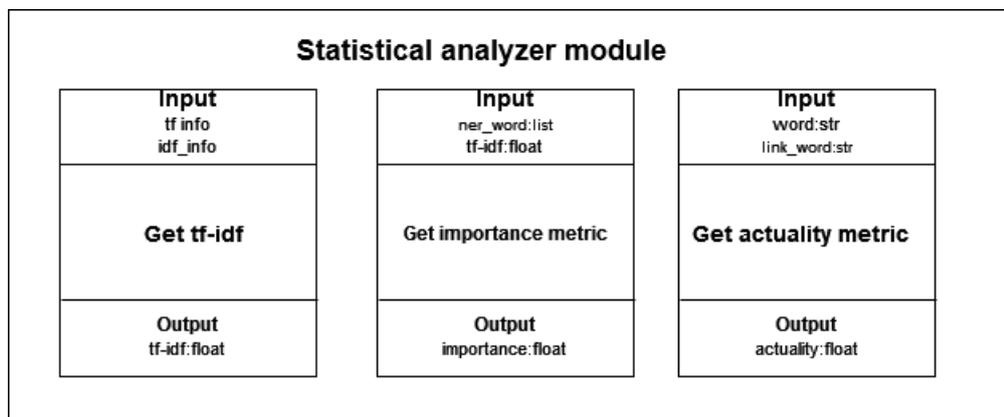


Рисунок 8 – Структура модуля статистического анализатора текстовых корпусов

Таким образом, основная структура программного обеспечения характеризуется наличием модулей, которые ограничены функционально, каждый модуль способен взаимодействовать с другим модулем для делегации необходимых функций. Функционал модулей может быть динамически дополнен необходимым функционалом.

Дальнейший функционал будет рассмотрен в следующих главах более подробно.

## 2.2 Проектирование и разработка модуля анализа содержимого сайта

Модуль анализа содержимого сайта отвечает за получение корпусов текстов для их дальнейшего анализа на предмет онтологий. В целом он состоит из трёх функций: функции получения ссылок на свежие новости, функции получения содержимого каждой страницы и предварительной обработки текста.

В функции получения ссылок на свежие новости сначала происходит получение GET запроса при помощи библиотеки requests, далее, с использованием библиотеки BeautifulSoup путём парсинга тегов short-text выделяются ссылки, которые итеративно добавляются в список. В итоге функция возвращает список ссылок.

```
def get_sport_refs():  
  
    lnk='https://www.sports.ru'  
    result=[]  
  
    text_html=requests.get(lnk+'/topnews')  
    text_html.encoding='utf-8'  
    bs=BeautifulSoup(text_html.text)  
  
    a_tags=bs.find_all('a',attrs={'class':'short-text'})  
    i=0  
    for a in a_tags:#which contain links  
        if(i==3):  
            break  
        if(len(a['title'])>0):  
            result.append(lnk+a['href'])  
            i=i+1  
    return result
```

Рисунок 9 – Код функции получения ссылок

Далее по ссылке происходит получение страницы содержимого путём GET запроса. При помощи BeautifulSoup происходит выделение тегов текста и возвращается строковый параметр содержимого страницы.

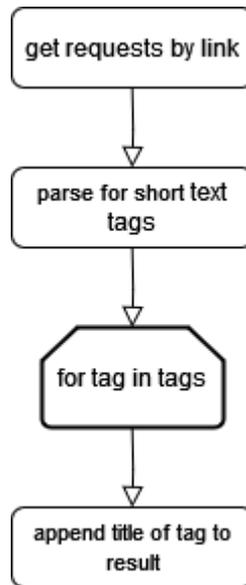


Рисунок 10 – Блок-схема функции получения ссылок

```

def get_sport_news_page(pg_url:str):
    th=requests.get(pg_url)
    th.encoding='utf-8'
    bs=BeautifulSoup(th.text)
    result=bs.find_all('p')
    return result
  
```

Рисунок 11 – Код функции получения ссылок

Данная функция принимает аргумент в виде строкового параметра, содержащего ссылку на новостной ресурс. Возвращает строковый параметр содержимого страницы.

Последняя функция данного модуля производит предварительную обработку текста, очищая его от ненужных символов. Данный процесс производится при помощи использования регулярных выражений из библиотеки re.

```

def preprocess_text(txt:str):
    #txt = re.sub('<[^>]*>', '', str(txt))
    txt = re.sub('<[^>]*>', '', str(txt))
    txt = re.sub('\xa0', '', str(txt))
    return txt
  
```

Рисунок 12 – Код функции обработки текста

Рассмотрим данный модуль с точки зрения проектирования. Функция `get_sports_refs()` производит GET запрос на удалённый сервер и в случае успешного выполнения (код состояния 201) производит поиск по тегам ссылок.

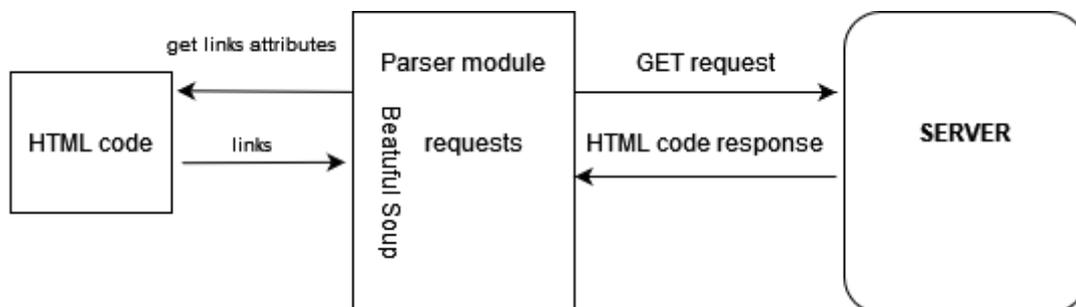


Рисунок 13 – Общая схема функции получения ссылок

Далее необходимо получить ссылки на новостные ресурсы страницы. Для этого нужно также использовать библиотеку `BeautifulSoup`, которая отвечает за анализ HTML кода.

Сначала также происходит GET запрос ссылки страницы, а затем при помощи метода `find_all` производится поиск всех текстовых тегов ссылок а. В конце используется функция обработки полученного текста от тегов `preprocess_text`. Здесь происходит поиск а элементов с атрибутами класса `short-text`.

Функция получения содержимого страницы получает на вход ссылку и аналогичным образом производит GET запрос содержимого страницы, однако после чего, в отличие от предыдущего модуля она ищет текстовые корпуса и в дальнейшем делегирует их функции предобработки текста.

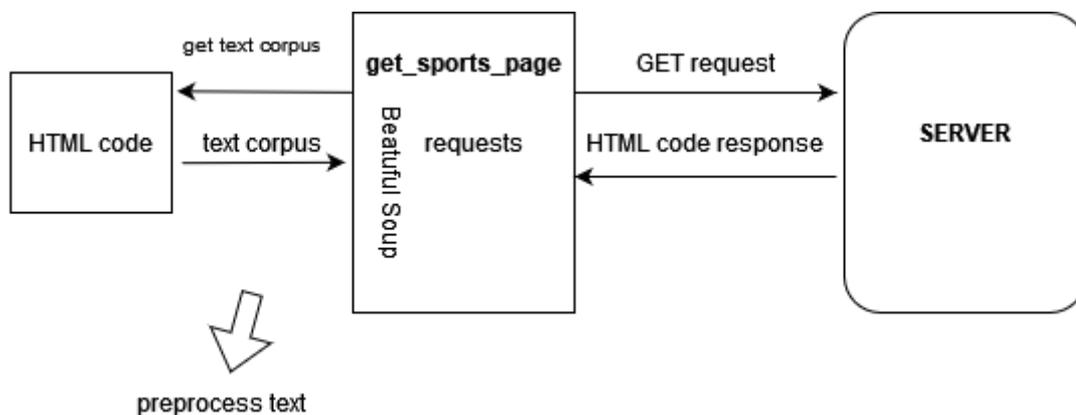


Рисунок 14 – Схема функции получения страницы

Данная функция также использует библиотеки BeautifulSoup и requests. Таким образом, данный модуль характеризуется наличием взаимопользующих функций, работа которых синхронизирована.

Лидер «Денвера» Никола Йокич помог команде завершить финал Западной конференции против «Лейкерс» в четвертом матче (113:111, 3-). Йокич в понедельник набрал 30 очков, 14 подборов и 13 передач. За 45 минут 2-кратный MVP реализовал 11 из 24 с игры, а также 5 из 6 с линии. В его активе 1 перехват и 3 блок-шота при 3 потерях. Это восьмой трипл-дабл в исполнении серба в плей-офф-023. Он обновил рекорд Уилта Чемберлена 1967 года по количеству трипл-даблов за один розыгрыш плей-офф (7). Всего для Николя это 14-й трипл-дабл в плей-офф за карьеру. Лидер – Мэджик Джонсон (30), на 2-м месте Леброн Джеймс (28)., Snoop Dogg – голос болельщиков «Лейкерс»: истерики после поражений, выпендрег после побед, преклонение перед звездами, Опубликовал Валерий Левкин  
Источник – ESPN, Выберите любимую команду]

Рисунок 15 – Результат работы модуля получения текста

## 2.3 Проектирование и разработка модуля распознавания именованных сущностей

Основная цель данной главы заключается в извлечении имён собственных и их классификации из полученных в предыдущей главе текстов. Для этого будет использован набор библиотек `natasha`, который позволит провести синтаксический и морфологический анализы русскоязычного текста, а также непосредственно провести анализ их именованных сущностей.

Для начала нужно инициировать сущности распознавания элементов текста.

```
def init_ner(text:str):
    segmenter=Segmenter()
    emb=NewsEmbedding()
    morph=MorphVocab()
    morph_tagger=NewsMorphTagger(emb)
    ner_tagger=NewsNERTagger(emb)
    names=NamesExtractor(morph)

    doc=Doc(text)
    doc.segment(segmenter)
    doc.tag_morph(morph_tagger)
    doc.tag_ner(ner_tagger)

    return doc
```

Рисунок 16 – Начальная инициация сущностей распознавания именованных текстов

На вышеизложенном рисунке функция иницирует объект `Doc`, который принимает один аргумент — анализируемый текст. Вызываются методы для начальной инициации объектов распознавания именованных сущностей. Метод `tag_ner` производит непосредственно анализ именованных сущностей текста `txt`. Возвращается объект `doc`, содержащий именованные сущности текста.

Для непосредственного отображения информации о полученных именованных сущностях воспользуемся вложенным списком, первый элемент которого будет отображать слово, а следующий — принадлежность к той или иной сущности.

```

def add_ner_to_list(ner_info:list):
    ret=[[], []]

    for sp in doc.spans:
        if(sp.text in ret[0]): #if already presented in list
            continue

        ret[0].append(sp.text)
        if(sp.type=='ORG'):
            ret[1].append('Organization')
        elif(sp.type=='LOC'):
            ret[1].append('Location')
        elif(sp.type=='PER'):
            ret[1].append('Name')

    return ret

```

Рисунок 17 – Распределение именованных сущностей по спискам

Список `ret` содержит в себе два вложенных списка, первый из которых принимает в себя само слово(свойство `sp.text`), а второй — сущность этого слова(свойство `sp.type`). Перед добавлением сущности во второй список происходит проверка для более информативного отображения. В данном случае `ORG` — наименование организации, `LOC` — наименование места(страны, населённого пункта), `PER` — собственное имя.

Свойство `spans` объекта `doc` является списком именованных сущностей. В результате работы данного кода будет создан двойной список, содержащий в себе два списка — объектов и их сущностей. Принимает функция один аргумент — информацию о именованных сущностей в виде списка.

Для объединения двух функций создания списка именованных сущностей создадим функцию `proceed_ner()`, которая также принимает один аргумент — текстовый корпус.

```

def proceed_ner(text:str):
    doc=init_ner(text)
    result=add_ner_to_list(doc.spans)

    return result

```

Рисунок 18 – Функция получения именованных списков

Возвращает данная функция также переменную в виде списка информации об именованных сущностях.

Рассмотрим работу модуля распознавания именованных сущностей с точки зрения проектирования архитектуры.

Результат работы функции представлен двойным списком list, содержащие имена собственные и их наименования.

	Объект	Наименование
0	Дмитрий Морозов	Name
1	DWH	Organization
2	ГКНПЦ им. М. В. Хруничева	Organization
3	Винды	Name
4	Standoff	Organization
..	...	...
75	Bimeister	Organization
76	Почта	Organization
77	Лимиты	Organization
78	UWB	Organization
79	Модуляция	Name

Рисунок 19 – Результат работы модуля распознавания именованных сущностей

## 2.4 Проектирование и разработка модуля синтаксического анализатора текстовых корпусов

Для выявления концептов и связей между ними необходимо провести синтаксический анализ текстовых корпусов.

Принципом его функционирования является постепенный анализ именованных сущностей на предмет зависимого или главного слова и последующем построении связей.

Как уже было сказано ранее, онтология представляет собой отношение  $\langle X, R, F \rangle$ , типа концепт – связь – концепт.

Главными словами выступают обычно существительные (в данном случае имена собственные) в именительном падеже. Зависимыми словами являются имена существительные или имена собственные, не стоящие в именительном падеже.

С точки зрения архитектуры данный процесс построения онтологии состоит из двух функций: сначала происходит поиск слов-связок, затем на основе их роли происходит поиск зависимых и главных концептов.

В функции поиска слов связок сначала происходит морфологический анализ предложения в поиске слова, содержащегося в списке именованных сущностей. Если слово является именем собственным, то происходит итеративный анализ следующих за ним слов в поиске глаголов, вспомогательных глаголов и частичек перед ними. Также происходит анализ кратких прилагательных и глаголов.

```

def get_link_word(head_word:list,sentence_info:list):
    result=[]

    for head in head_word:
        cnt=0
        for snt in sentence_info:
            if(snt.text==head):
                break
            cnt=cnt+1

        pos=cnt
        while(pos < len(sentence_info)):
            cur_verb=sentence_info[pos]
            if(cur_verb.pos=='AUX'):
                next_verb=sentence_info[pos+1]
                if(next_verb.pos=='VERB' and next_verb.feats.get('VerbForm'=='Inf')):
                    verb=cur_verb.text+' '+next_verb.text
                    result.append([head,verb])
                    break

            if(cur_verb.pos=='VERB' and cur_verb.feats.get('Mood')!='Imp'):
                next_verb=sentence_info[pos+1]
                if(cur_verb in result):
                    break
                if(next_verb.pos=='VERB' and next_verb.feats.get('VerbForm'=='Inf')):
                    verb=cur_verb.text+' '+next_verb.text
                else:
                    verb=sentence_info[pos].text
                result.append([head,verb])
                break
            pos=pos+1

    return result

```

Рисунок 20 – Синтаксический поиск слов-связок

В качестве входных параметров выступают список полученных главных существей и морфологическая информация об анализируемом предложении. Результатом является список слов-связок.

Sentence:  
 Манчестер Сити сыграет с Интером в финале Лиги чемпионов в Стамбуле  
 link word:  
 ['сыграет']

Sentence:  
 , Sports  
 link word:  
 Sentence:  
 ru проведет текстовую онлайн-трансляцию матча  
 link word:  
 ['проведет']

Sentence:  
 Начало – в 22:00 по московскому времени  
 link word:  
 Sentence:  
 , Игру в прямом эфире покажет канал Матч ТВ  
 link word:  
 Sentence:  
 , Лига чемпионов, Финал, ПРИМЕЧАНИЕ: время начала матча – московское  
 link word:  
 Sentence:  
 , Статистика Лиги чемпионов, Опубликовал Андрей Карнаухов , Источник – Sports  
 link word:  
 ['Опубликовал']

Рисунок 21 – Результат функции поиска слов-связок

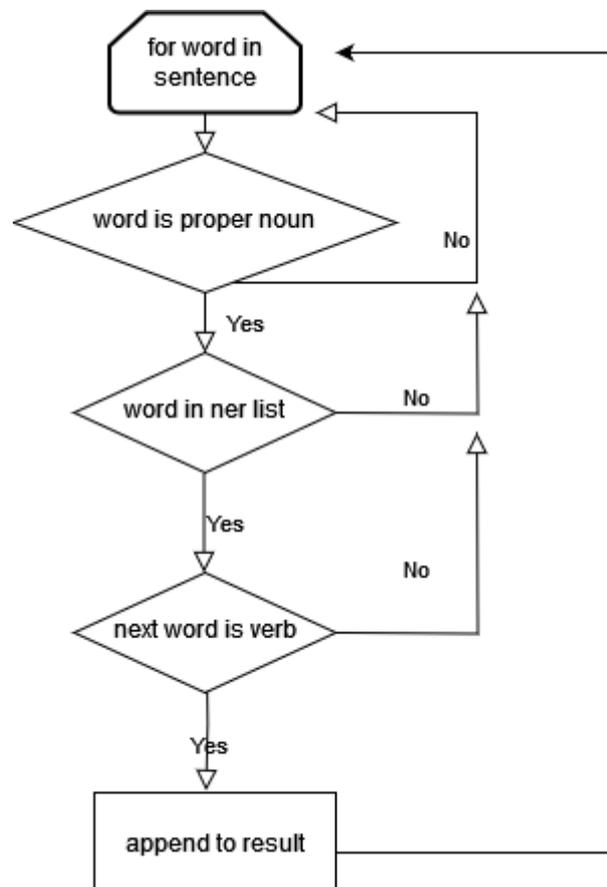


Рисунок 22 – Общая блок-схема функции поиска слова-связки

Для получения морфологической информации о предложении необходимо воспользоваться анализатором в библиотеке Natasha.

```
def get_words_info(text:str):  
    doc=init_ner(text)  
    words_info=doc.morph.tokens  
    return words_info
```

Рисунок 23 – Получение морфологической информации о предложении

Процесс выделения частей речи сопровождается инициацией распознавания при помощи функции `init_ner` из модуля распознавания именованных сущностей, а затем возвращается объект `tokens`, содержащий всю необходимую информацию о частях речи (Рисунок 23).

Входным параметром выступает корпус текста символьного типа. Выходным параметром является список информации о частях речи. Для выявления содержания списка слов в списке именованных сущностей необходимо выявить её принадлежность к списку именованных сущностей.

```
def in_ner(ner_text:list,word:str):  
    ret=False  
    for ner in ner_text:  
        if(word in ner):  
            ret=True  
            break  
  
    return ret
```

Рисунок 24 – Принадлежность слова списку именованных сущностей

Функция итеративно анализирует список именованных сущностей и проверяет, содержится ли искомое слово в этом списке (Рисунок 24).

Для поиска главного концепта необходимо проанализировать морфологическую информацию о предложении на наличие содержащихся в списке именованных сущностей существительных или имён собственных в именительном падеже.

```

def get_head_word(ner_info:list,sentence_info:list): |
    result=[]

    for i in range (len(sentence_info)):
        word=sentence_info[i]

        if(word.text in ''.join(result)):
            continue

        if((word.pos=='PROPN' or word.pos=='NOUN')
            and word.feats.get('Case')=='Nom'
            and in_ner(ner_info,word.text)):
            if(i > 0 ):
                prev_word=sentence_info[i-1]
                if(prev_word.text=='-'):
                    continue
            if(i < (len(sentence_info)-1)):
                next_word=sentence_info[i+1]
                if(next_word.pos=='PROPN' and word.pos=='PROPN'):
                    result.append(word.text + ' '+next_word.text)
                    continue

            result.append(word.text)

        if(word.pos=='PUNC'):
            break

    return result

```

Рисунок 25 – Поиск главного концепта

Входными параметрами выступают список именованных сущностей и морфологическая информация о предложении. Возвращается список слов.

```

Sentence:
Сегодня в рамках четвертого матча финальной серии плей-офф НБА 2023 года Майами дома уступил Денверу
head word:
['Майами']

Sentence:
Форвард Наггетс Аарон Гордон набрал 27 очков
head word:
['Аарон Гордон']

Sentence:
Денвер ведет в серии - 3-1
head word:
['Денвер']

Sentence:
Ты уродливый. Тебе лучше уйти. Джимми Батлер стал бездомным в 13 лет
head word:
['Джимми Батлер']

Sentence:
Это дилетантский взгляд. Эрик Споэльстра такой скрытный
head word:
['Эрик Споэльстра']

```

Рисунок 26 – Результат функции поиска главного концепта

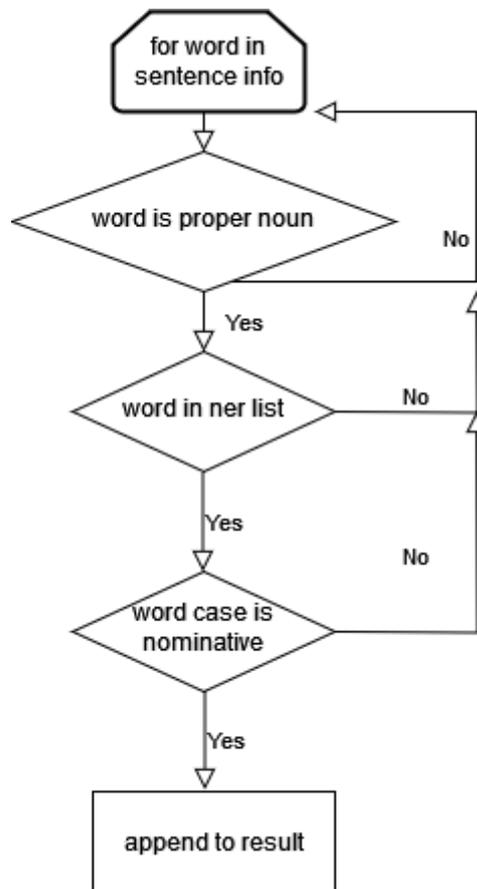


Рисунок 27 – Блок-схема функции поиска главного концепта

Аналогично для поиска зависимого концепта происходит начальный анализ списка слов-связок.

```

def get_dep_word(link_words:list,sentence_info:list):
    result=[]
    pos=0
    if(len(link_words)==0):
        return result

    for inf in link_words:
        while((pos < len(sentence_info)-1)):
            word=sentence_info[pos]
            if(word.text==inf):
                break
            pos=pos+1

        while(pos<(len(sentence_info))):
            word=sentence_info[pos]
            if(word in result):
                pos=pos+1
                continue
            #cur_result=''
            if(word.pos=='PUNCT' and (word.text=='.' or word.text==',')):
                break

            if(((word.pos=='NOUN' or word.pos=='PROPN')
                and word.feats.get('Case')!='Nom')
                or word.pos=='X' or word.pos=='ADJ'):
                result.append(word.text)

            pos=pos+1

    return result

```

Рисунок 28 – Поиск зависимого концепта

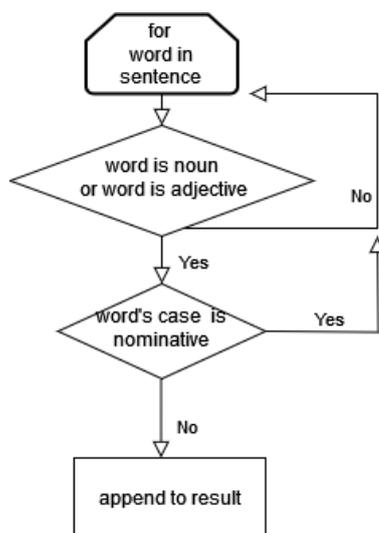


Рисунок 29 – Общая блок-схема поиска зависимого концепта

После определения позиции слова-связки происходит поиск следующих за ним существительных или имен собственных не в именительном падеже. (Рисунок 29). В качестве входных аргументов выступают слова связки и

морфологическая информация о предложении. Результат – список зависимых концептов (Рисунок 30).

---

Sentence:  
Пари НН примет Родину в ответном стыковом матче за право выступать в Мир РПЛ.  
dependent word:  
[ 'Родину' ]

Sentence:  
Также в субботу Факел сыграет дома с Енисеем.  
dependent word:  
[ 'Енисеем' ]

Рисунок 30 – Результат поиска зависимого концепта

Для итогового представления онтологической модели необходимо объединить работу данных функций.

Общий алгоритм представляет собой разбиение полученного корпуса текста на отдельные предложения с последующим морфологическим анализом каждого предложения, после чего происходит получение слова-связки. Далее идёт получение главного и зависимого концептов.

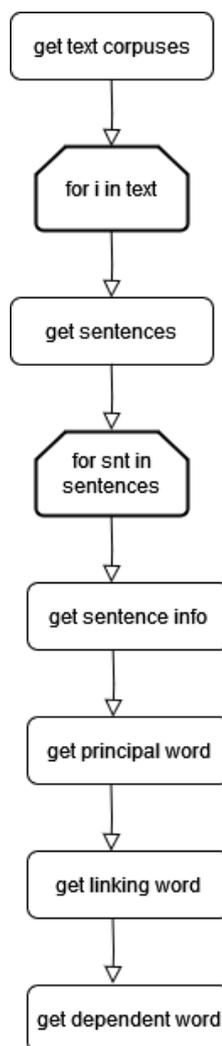


Рисунок 31 – Общая блок-схема синтаксического анализатора

В результате работы данного модуля получаются тройки связей типа  $\langle X,R,F \rangle$ , где X- главный концепт, R- слово-связка и F- зависимый концепт.

## 2.5 Проектирования и разработка модуля статистического анализа текстовых корпусов

Статистический анализ текстовых корпусов включает в себя процесс ранжирования полученных концептов и выделения среди них главных и дочерних элементов.

Как уже было сказано ранее, для ранжирования списка слов используется методика TF-IDF. Она отображает степень важности слов в тексте.

Функция получения метрики TF принимает в качестве входных аргументов два параметра: корпус текстов, где производится поиск слова и само слово.

```
def get_tf(text:list,word:str):
    wrd_occur=0
    for t in text:
        if(t == word):
            wrd_occur=wrd_occur+1

    sz=len(text)

    return round(float(wrd_occur/sz),3)
```

Рисунок 32 – Функция получения метрики TF

В качестве результата функция возвращает округлённое значение типа переменной с плавающей точкой (рисунок 32).

Функция получения метрики IDF также принимает два параметра: список всех корпусов текста и анализируемое слово.

```
def get_idf(text_corpuses:list,word:str):
    text_cnt=len(text_corpuses)
    word_cnt_in_text=0

    for t in text_corpuses:
        for l in t.split():
            if(l == word or (l in word)):
                word_cnt_in_text=word_cnt_in_text+1
                break

    if(word_cnt_in_text==0):
        return 0
    result=math.log((text_cnt/word_cnt_in_text),10)

    return round(float(result),7)
```

Рисунок 33 – Функция получения метрики IDF

Функция получения метрики IDF возвращает округлённое до семи знаков значение типа с плавающей точкой.

Для получения конечного результата метрики TF-IDF необходимо перемножить полученные ранее значения.

```
def get_tf_idf(tf:float,idf:float):  
    return round(float(tf*idf),7)
```

Рисунок 34 – Функция получения метрики TF-IDF

В результате возвращается округленная до семи знаков переменная (рисунок 35).

```
Денвера 0.0002097  
Никола 0.0001031  
Йокич 0.0002097  
НБА 0.0054471  
Наггетс 0.0008078  
Майами 0.0005543  
Хит 0.0008078  
Йокич 0.0002097  
Йокичпризнан 0.0001031  
Денвер 0.0002097  
Опубликовал 0.0  
Валерий 0.0003203  
Левкин 0.0  
Источник 0.0  
ESPN 0.007  
Выберите 0.0
```

Рисунок 35 – Результат получения метрики TF-IDF

Ранжирование метрик TF-IDF всех слов происходит путем постепенного добавления их значений в список слов и их последующей сортировки.

В итеративно происходит поиск наименьшего значения полученных метрик TF-IDF, после чего они добавляются в ранжированный список.



```

def get_word_importance(ner_word:str,tf_idf:list,tf:list,text_corpuses:list):

    tf_idf_pos=tf_idf[0].index(ner_word) #get words pos
    r2=float(1/tf_idf_pos)

    tf_pos=tf[0].index(ner_word)
    r1=float(1/tf_pos)

    q=get_text_corpuses_word_cnt(text_corpuses,ner_word)

    max_wrd=get_max_text_corpuses_word(text_corpuses,ner_word)
    min_wrd=get_min_text_corpuses_word(text_corpuses,ner_word)

    r1_r2=abs(r1-r2)
    if(r1_r2 == 0):
        k_0=0
    else:
        k_0=round(float(1/(r1_r2)),7)
    max_min=(max_wrd-min_wrd)

    if(max_min ==0):
        k_1=0
    else:
        k_1=round(float(q/(max_min)),7)

    k=round(float(k_0*k_1),7)
    return k

```

Рисунок 38 – функция получения метрики важности слова

Функция применяет упомянутую заранее формулу, и проводит оценку важности каждого слова в тексте. Входными аргументами выступают анализируемое слово, список результата метрики TF, список результата метрики IDF и текстовые корпуса.

В результате функция получает переменную, отражающую степень важности слова в тексте.

Получив метрику важности слова, необходимо ранжировать эти слова метрики важности слов и выделить главные концепты. Главными концептами являются слова с наибольшим значением  $k$ , следовательно, необходимо их ранжировать и выделить слова с наибольшим значением.

Конечным шагом данного модуля будет получение метрики актуальности полученной онтологической тройки  $\langle X,R,F \rangle$  с применением следующей формулы:

$$P_{place} = \frac{k * \left( \left( \frac{|m - n|}{\min(m, n)} + 2 \right) + 1 \right)}{n + m + 1}$$

```
def get_actuality_metric(word_1:str,word_2:str,link_word:str,sentence:list):
    n=get_distance(w1,link_word,sentence)
    m=get_distance(w2,link_word,sentence)

    result=k*((abs(m-n)/min(m,n)+1)n+m+1)
    return result
```

Рисунок 39 – Функция получения метрики актуальности

В данной функции сначала происходит получение расстояния от первого концепта до слова-связки, от второго концепта до слова-связки и далее применяется вышеизложенная формула.

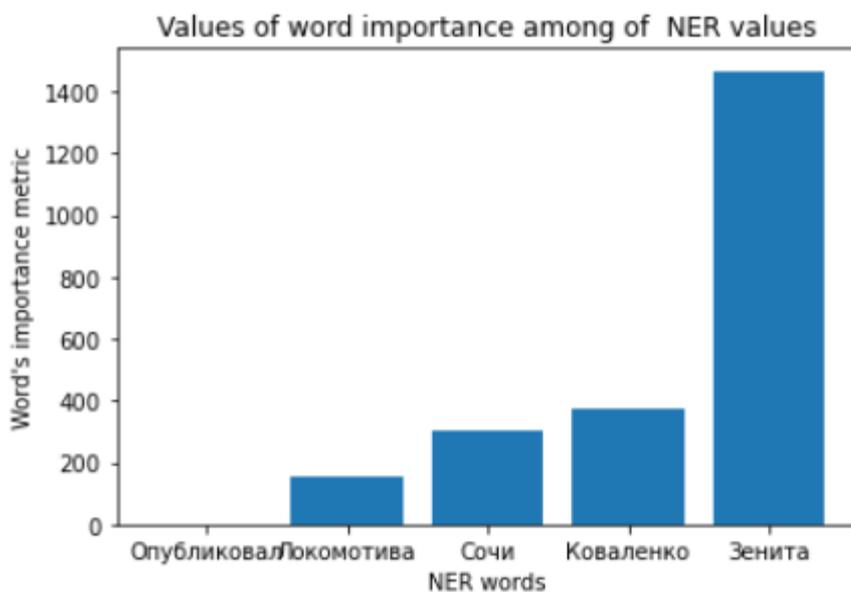


Рисунок 40 – График результата вывода функции оценки важности именованных сущностей

Итак, алгоритм данного модуля представляет собой получение метрик TF и IDF, его ранжирования, получения главных концептов и извлечение метрики актуальности каждой тройки онтологии.

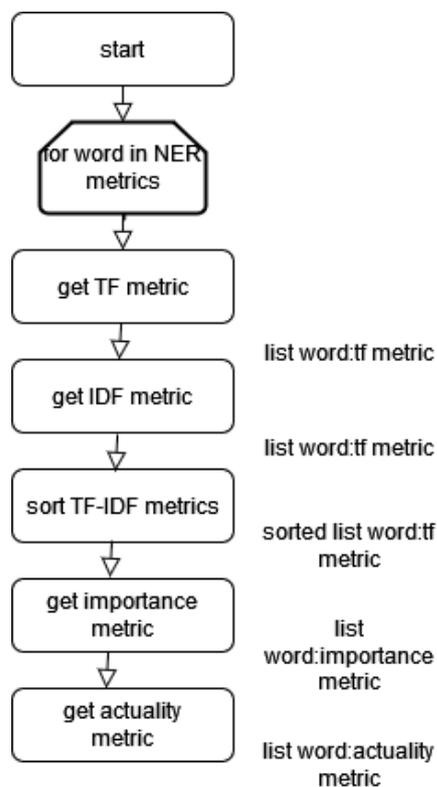


Рисунок 41 – Общая блок-схема модуля статистического анализатора  
 Результатом данного модуля является список метрик актуальности для троек онтологий.

## 2.6 Проектирование и разработка модуля представления результата

Для отображения полученного результата будет использована библиотека networkx. Полученные модели онтологии будут представлены как графовые объекты. В функции отображения результата происходит постепенное добавление рёбер от главного концепта к слову-связке и от слова-связки до зависимого концепта.



Рисунок 42 – Результат работы программы

Таким образом, в результате работы программы были отображены полученные онтологии в графовом виде.

Отказ от использования языков разметки обусловлен необходимостью дальнейшего развития данного программного обеспечения как с точки зрения его архитектурного строения, так и с точки зрения функциональной составляющей.

## ГЛАВА 3. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

Данная работа посвящена разработке программного обеспечения для автоматического построения онтологических моделей на основе анализа текстовых корпусов. В контексте функциональной полезности, программное обеспечение способствует облегчению восприятия специфической информации в определённой предметной области, а также формализации усваиваемой информации. Возможными потребителями могут выступать образовательные учреждения, бизнес организации, государственные организации.

### 3.1 Предпроектный анализ

В целях определения потенциальных потребителей необходимо провести анализ целевого рынка. Потенциальными потребителями могут выступать:

- 1) Частные организации.
- 2) Образовательные организации.
- 3) Независимые частные лица.

Теперь необходимо провести анализ секторов рынка. По причине неоднородности категории потребителей, критерии сегментов рынка также будут отличаться.

Таблица 1 – Сегментирование рынка услуг

		Области применения		
Категория потребителей		Создание дорожных карт(mind map)	Проектирование бизнес-процессов	Индексация поисковых запросов
	Частные организации			
	Образовательные организации			



В результате построения таблицы сегментирования выявлена корреляция потребителей и области применения программного обеспечения. Программное решение построение онтологических моделей предметной области может быть применено в сегменте рынка частных организаций, однако приоритетными остаются независимые физические лица.

### 3.2 Анализ конкурентных технических решений

Данное программное решение призвано решать проблемы визуализации сущностных характеристик предметных областей.

Данный проект даёт пользователям эффективное представление информации и позволяет им лучше воспринимать данные.

Для анализа конкурентоспособных технических решений воспользуемся следующей формулой:

$$K = \sum V_i * B_i$$

Где  $V_i$  – вес показателя,  $B_i$  – балл  $i$ -го показателя.

Сначала представим оценочную карту для сравнения конкурентных технических решений

Таблица 2 – Сравнение конкурентных технических решений (разработок)

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		$B_{\phi}$	$B_{K1}$	$B_{K2}$	$K_{\phi}$	$K_{K1}$	$K_{K2}$
1	2	3	4	5	6	7	8
<b>Технические критерии оценки ресурсоэффективности</b>							
1.Повышение производительности труда пользователя	0,2	0,12	0,15	0,11	0,10	0,1	0,02
2. Простота использования	0,15	0,19	0,11	0,12	0,10	0,14	0,1

3. Эффективность выполнения задачи	0,15	0,02	0,23	0,21	0,35	0,24	0,11
4. Производительность с точки зрения использования технических ресурсов	0,15	0,15	0,10	0,11	0,11	0,15	0,1
5. Функциональность	0,15	0,19	0,17	0,07	0,19	0,16	0,12
6. Конкурентоспособность	0,2	0,11	0,05	0,10	0,09	0,15	0,18
<b>Итого</b>	<b>1</b>	<b>0,78</b>	<b>0,81</b>	<b>0,72</b>	<b>0,84</b>	<b>0,94</b>	<b>0,63</b>

В таблице 2 происходит сравнение трёх конкурирующих разработок по следующим критериям: Повышение производительности труда пользователя, простота использования, эффективность выполнения задачи, производительность с точки зрения использования технических ресурсов, функциональность, конкурентоспособность.

В качестве показателя научно -технического проекта программы построения онтологических моделей на основе анализа текстовых корпусов, – баллы за критерии ресурсоэффективности конкурирующего программного обеспечения для построения онтологической модели с использованием языка нотаций RDF,  $B_{к2}$  – баллы за критерии ресурсоэффективности конкурирующего программного обеспечения отображения моделей онтологий protege.

$K_{ф}$  является общая конкурентоспособность данной разработки, и  $K_{к2}$  – конкурентоспособности первого и второго аналогичного проектов соответственно.

Значение средневзвешенного показателя  $K$  составляет 78%, что является неплохим результатом перспективы функционирования программного обеспечения на рынке.

### 3.3 SWOT-анализ

Для выявления недостатков и достоинств программного обеспечения нужно провести SWOT анализ.

Таблица 3 – матрица SWOT-анализ

	<p><b>Сильные стороны научно-исследовательского проекта:</b></p> <p>С1. Возможность динамического подключения модулей</p> <p>С2. Новый подход в проектировании и отражении моделей онтологий</p> <p>С3. Возможность расширения функционала программного обеспечения</p> <p>С4. Кроссплатформенность</p>	<p><b>Слабые стороны научно-исследовательского проекта</b></p> <p>Сл1. Долгое время выполнения программного обеспечения</p> <p>Сл2. Отсутствие решения проектирования ООП</p> <p>Сл 3. Отсутствие возможности работы с текстами разных источников</p>
<p><b>Возможности</b></p> <p>В1. Построение модели бизнес-процесса на основе его текстового описания</p> <p>В2. Динамическое построение дорожных карт учебного материала (ограничено областью знаний)</p> <p>В3. Индексация поисковых запросов</p>	<p>Сильные стороны данного проекта характеризуют его как динамически подключаемый проект с возможностями дальнейшего развития функционала.</p> <p>Кроссплатформенность позволяет использовать проект на любых платформах.</p> <p>Возможности проекта позволяют ему быть конкурентноспособным на рынке программного обеспечения.</p> <p>Сильные стороны проекта коррелируют с его возможностями, что позволяет говорить об их единой природе.</p>	<p>Слабые стороны данного проекта говорят о его недостаточной оптимизации с точки зрения времени выполнения и некоторых недостатках архитектуры.</p> <p>Для их исправления необходимо провести оптимизацию и перепроектирование проекта.</p>
<p><b>Угрозы</b></p> <p>У1. Отсутствие спроса в среде промышленного производства</p> <p>У2. Ограничение эксплуатации</p>	<p>Основными негативными тенденциями данного проекта являются ограничение области эксплуатации ввиду</p>	<p>Слабые стороны и угрозы проявляют некоторую корреляцию между собой. Так, долгое выполнение</p>

УЗ. Отсутствие технической поддержки	ограниченности области знания данного проекта, отсутствие технической поддержки что, подразумевает невозможность консультации пользователей по возникшим проблемам.	программного обеспечения или отсутствие архитектуры ООП могут повести за собой отсутствие спроса в промышленной среде и отсутствие возможности работы с текстами разных источников может выступать причиной ограничению эксплуатации
--------------------------------------	---	--

В целом, данный проект способен составить конкуренцию на рынке. Рассматривая его сильные стороны и возможности, стоит учесть, что Сильные стороны данного проекта характеризуют его как динамически подключаемый проект с возможностями дальнейшего развития функционала. Кроссплатформенность позволяет использовать проект на любых платформах. Возможности проекта позволяют ему быть конкурентноспособным на рынке программного обеспечения.

Сильные стороны проекта коррелируют с его возможностями, что позволяет говорить об их единой природе.

Слабые стороны и угрозы проявляют некоторую корреляцию между собой. Так, долгое выполнение программного обеспечения или отсутствие архитектуры ООП могут повести за собой отсутствие спроса в промышленной среде и отсутствие возможности работы с текстами разных источников может выступать причиной ограничению эксплуатации.

Основными негативными тенденциями данного проекта являются ограничение области эксплуатации ввиду ограниченности области знания данного проекта, отсутствие технической поддержки что, подразумевает невозможность консультации пользователей по возникшим проблемам.

### **3.4 Оценка готовности проекта к коммерциализации**

Вне зависимости от степени внедрения программного обеспечения важно оценить степень её коммерциализации. Для оценки готовности проекта

к процессу коммерциализации построим таблице анализа коммерческого потенциала программного обеспечения.

Таблица 4 – Оценка степени готовности научного проекта к коммерциализации

№ п/п	Наименование	Степень проработанности научного проекта	Уровень имеющихся знаний у разработчика
1.	Определен имеющийся научно-технический задел	4	1
2.	Определены перспективные направления коммерциализации научно-технического задела	3	1
3.	Определены отрасли и технологии (товары и услуги) для предложения на рынке	2	2
4.	Определена товарная форма научно-технического задела для представления на рынок	2	1
5.	Определены авторы и осуществлена охрана их прав	1	2
6.	Проведена оценка стоимости интеллектуальной собственности	2	3
7.	Проведены маркетинговые исследования рынка сбыта	3	1
8.	Разработан бизнес-план коммерциализации научной разработки	1	1
9.	Определены пути продвижения научной разработки на рынок	2	2

10	Разработана стратегия (форма) реализации научной разработк	3	2
11.	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	3	3
12.	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	2	2
13.	Проработаны вопросы финансирования коммерциализации научной разработки	1	1
14.	Имеется команда для коммерциализации научной разработки	1	2
15.	Проработан механизм реализации научного проекта	2	1
	Итого баллов	32	23

В результате оценки степени готовности научного проекта к коммерциализации перспектива данного проекта оценивается как средняя.

### **3.5 Методы коммерциализации результатов научно-технического исследования**

К методам коммерциализации результатов научно-технического данного исследования можно отнести:

1. Торговлю патентными лицензиями.
2. Инжиниринг.
3. Организация собственного предприятия.
4. Организация совместного предприятия.

### 3.6 Инициация проекта

#### 3.6.1 Цели и результаты проекта

Цель данного проекта – создание программного обеспечения для построения онтологических моделей предметной области.

Результат позволит организовать более эффективное восприятие информации, что поможет организовать процессы обучения в образовательных организациях, построения бизнес-процессов в промышленности.

Данный проект применим к профильным, государственным и частным организациям различной направленности и может удовлетворить такие потребности как потребность в более эффективном усвоении информации, формализации знаний.

Таблица 5– Цели и результат проекта

Цели проекта	Ожидаемые результаты проекта
Разработка программного обеспечения для построения онтологических моделей предметной области	Эффективное достижение целей программного обеспечения. Внедрение на рынок услуг

#### 3.6.2 Организационная структура проекта.

В организационную структуру данного проекта входят: Кайда Анастасия Юрьевна, которая является руководителем проекта, Ковин Роман Владимирович является консультантом по техническим и проектировочным вопросам, Михайлов Никита Игоревич, являющийся исполнителем данного проекта.

Таблица 6 – Рабочая группа проекта

ФИО, основное место работы, должность	Роль в проекте	Функции
Кайда А.Ю, старший преподаватель ТПУ	Руководитель	Контроль выполнения проекта. Постановка технических

		заданий
Ковин Р.В, преподаватель ТПУ	Консультант по техническим вопросам и вопросам проектирования	Консультирование по вопросам разработки
Михайлов Н.И, студент ТПУ	Исполнитель	Проектирование и разработка проекта

### 3.6.3 Ограничения и допущения проекта.

Ограничения проекта – это все факторы, которые могут послужить ограничением степени свободы участников команды проекта, а также «границы проекта» - параметры проекта или его продукта, которые не будут реализованных в рамках данного проекта.

Таблица 7 – Ограничения проекта

Фактор	Ограничения/допущения
3.1. Бюджет проекта	Ограничение финансовых средств в рамках работы проекта
3.1.1. Источник финансирования	НИ ТПУ
3.2. Сроки проекта	12.03.2023 - 27.06.2023
3.2.1. Дата утверждения плана управления проектом	12.03.2023
3.2.2. Дата завершения проекта	26.03.2023
3.3. Прочие ограничения и допущения	-

### 3.7 Планирование управления научно-техническим проектом

Для осуществления разработки, был сформирован ряд работ и назначены исполнители для каждого этапа работы

Таблица 8 – Структура работ в рамках научного исследования

Но работы	Наименование работы	Исполнители работы
1	Выбор научного руководителя	Исполнитель

	магистерской работы	
2	Составление и утверждение темы выпускной квалификационной работы магистранта	Исполнитель Руководитель
3	Составление календарного плана-графика выполнения выпускной квалификационной работы магистранта	Руководитель
4	Подбор и изучение литературы по теме магистерской работы	Исполнитель
5.	Анализ предметной области	Исполнитель Руководитель
6.	Проектирование программного обеспечения	Исполнитель Руководитель Консультант по техническим вопросам и вопросам проектирования
7.	Разработка технической документации и проектирование	Исполнитель Руководитель
8.	Тестирование проекта	Исполнитель Руководитель
9.	Согласование выполненной работы с научным руководителем	Исполнитель Руководитель
10	Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	Исполнитель

Представим иллюстрацию календарного плана проекта при помощи диаграммы Ганта.

Таблица 9 – Диаграмма Ганта

Код работ (из ИСР)	Вид работ	Исполнители	Тк, кал, дн	Продолжительность выполнения работ															
				февр			март			апрель			май		июнь				
				2	3	1	2	3	1	2	3	1	2	3	1	2			

1	Выбор научного руководителя магистерской работы	Исполнитель	3																
2	Составление и утверждение темы выпускной квалификационной работы магистранта	Исполнитель	5		 														
3	Составление календарного плана-графика выполнения выпускной квалификационной работы магистранта	Руководитель	2																
4	Подбор и изучение литературы по теме магистерской работы	Исполнитель	17																
5	Анализ предметной области	Исполнитель, Руководитель	26		 														
6	Проектирование программного обеспечения	Исполнитель Руководитель Консультант по техническим вопросам и вопросам проектирования	6		  														



где  $t_{\min}$  – минимальная трудоемкость работ, чел/дн,  $t_{\max}$  – максимальная трудоемкость работ, чел/дн.

Следующая формула, определяет продолжительность каждой работы в рабочих днях  $T_{pn} \cdot T_{pn} = \frac{t_{ож}}{ч_n}$

где  $T_{pn}$  – продолжительность одной работы, раб.дни,

$t_{ож}$  – ожидаемая трудоемкость выполнения одной работы, чел.-дни,

$ч_n$  – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел

Расчет продолжительности этапа работ в календарных днях ведется по формуле:

$$T_{кд} = T_{рд} * K_{кл}$$

где  $T_{кд}$  – продолжительность выполнения  $i$ -й работы в календарных днях,

$T_{рд}$  – продолжительность выполнения  $i$ -й работы в рабочих днях,

$K_{кл}$  – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$K_{кл} = \frac{T_{кал}}{T_{кд} - T_{вд} - T_{пр}}$$

где – количество календарных дней в году,

$T_{вд}$  – количество выходных дней в году,

$T_{пр}$  – количество праздничных дней в году.

Построим таблицу показателей осуществления разработки

Таблица 10 – Временные показатели осуществления разработки

No	Наименование работы	Исполнители работы	Трудоемкость работ, чел.-дни			Длительность работ, дни	
			$t_{\min}$	$t_{\max}$	$t_{ож}$	$T_{pn}$	$T_k$
1	Выбор научного руководителя магистерской работы	Исполнитель	1	3	1,8	1,8	1,8
2	Составление и утверждение темы	Руководитель	2	7	4	4	4,2

	выпускной квалификационной работы магистранта						
3	Составление календарного плана-графика выполнения выпускной квалификационной работы магистранта	Руководитель	2	11	5,6	5,6	6,3
4	Подбор и изучение литературы по теме магистерской работы	Исполнитель	4	3	3,6	3,6	3,8
5	Анализ предметной области	Исполнитель, Консультант по техническим вопросам и вопросам проектирования	20	30	24	24	56
6	Проектирование программного обеспечения	Исполнитель	20	87	46,8	46,8	57
7	Разработка технической документации и проектирование	Исполнитель Руководитель Консультант по техническим вопросам и вопросам проектирования	19	78	42,6	21,3	26
8	Тестирование проекта	Исполнитель	12	35	21,2	21,2	27
9	Согласование выполненной работы с научным руководителем	Исполнитель Руководитель	11	56	29	14,5	15
10	Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	Исполнитель Руководитель	37	89	57,8	57,8	98

### **3.8. Бюджет научно-технической разработки.**

**Сырье, материалы, покупные изделия и полуфабрикаты (за вычетом отходов)**

В эту статью включаются затраты на приобретение всех видов материалов, комплектующих изделий и полуфабрикатов, необходимых для выполнения работ по данной теме. Количество потребных материальных ценностей определяется по нормам расхода.

Таблица 11 – Группировка затрат по статьям

Вид работ	Статьи				
	Специальное оборудование для научных (экспериментальных) работ	Основная заработная плата	Дополнительная заработная плата	Отчисления на социальные нужды	Итого плановая себестоимость
Написание научно-технического проекта	45789	1314,32	10562,5664	16617,85	74283,7364

Материальные затраты включают в себя затраты на покупку канцелярских предметов на сумму 1462 рубля. Общая сумма материальных затрат составляет 1462 рубля

Оборудованием выступает персональный компьютер, чья стоимость составляет 45789 рублей. Планируемый срок использования компьютера – 4 года. Планируется использовать компьютер для написания работы в течении 4 месяцев. Тогда норма амортизации:

$$Ah = \frac{1}{n} * 100\% = \frac{1}{4} = 25\%$$

годовые амортизационные отчисления:

$$Ar = j * Ah = 45789 * 0.25 = 11447,25$$

ежемесячные амортизационные отчисления:

$$Am = \frac{Ar}{12} = \frac{11447,25}{12} = 953,9$$

итоговая сумма амортизации основных средств:

$$A = Am * d = 953,9 * 4 = 3815,6$$

Итоговая сумма затрат на амортизацию составила 3815,6.

Статья расходов «Основная заработная плата исполнителей темы» включает основную заработную плату с учетом премий и доплат для исполнителей проекта: исполнителя и руководителя.

Оклад руководителя составляет 23567 рублей, исполнителя – 7143 рубля (стипендия). Количество отпускных и выходных дней составляет 10,4. (с учетом длительности отпуска в 48 дней).

Рассчитаем среднедневную заработную плату:

$$Z_{рук} = \frac{Z_m * M}{F} = \frac{23567 * 10,4}{243} = 1008,62 \text{ руб}$$

$$Z_{студ} = \frac{Z_m * M}{F} = \frac{7143 * 10,4}{243} = 305,70 \text{ руб}$$

Расчет основной заработной платы осуществляется по формуле:

$$Z_{осн} = Z_{дн} * T_p * (1 + K_{пр} + K_d) * K_p$$

Где

$Z_{осн}$  – среднедневная заработная плата, руб.,

$T_p$  – продолжительность работ, выполняемых работником, раб. дни,

$K_{пр}$  – премиальный коэффициент,

$K_d$  – коэффициент доплат и надбавок,

$K_p$  – районный коэффициент.

Результаты соответствующих расчетов приведены в таблице.

Таблица 12 – Расчет основной заработной платы

Исполнители	З <sub>дн</sub> руб.	K <sub>пр</sub>	K <sub>д</sub>	K <sub>р</sub>	T <sub>р</sub>	K <sub>осн</sub> руб
Исполнитель	305,70	0,3	0,2	1,3	98	58519,27
Руководитель	1008,62	0,3	0,2	1,3	15	29502,135
Консультант по техническим вопросам и вопросам проектирования	1030,5	0,3	0,2	1,3	9	18085,27
Итого						106106,81

Зная основную заработную плату, можно рассчитать дополнительную заработную плату в размере 12 % от основной.

$$Z_{\text{доп}} = k_{\text{доп}} * Z_{\text{осн}}$$

где  $k_{\text{доп}}$  – коэффициент дополнительной заработной платы,  $Z_{\text{осн}}$  – основная заработная плата.

Таблица 13 – Расчет дополнительной заработной платы

Исполнители	$k_{\text{доп}}$	Z <sub>осн</sub> руб	Z <sub>доп</sub> , руб
Исполнитель	0,12	58519,27	7022,3124
Руководитель	0,12	29502,135	3540,2562
Консультант по техническим вопросам и вопросам проектирования	0,12	18085,27	2170,2324
Итого			12170,2324

Отчисления во внебюджетные фонды рассчитываются как

$$Z_{\text{внеб}} = k_{\text{внеб}} * (Z_{\text{осн}} + Z_{\text{доп}})$$

где  $Z_{\text{внеб}}$  – коэффициент внебюджетные фонды; в 2022 г., в соответствии с Федеральным законом для учреждений, осуществляющих образовательную и научную деятельность, используется пониженная ставка – 30%,  $Z_{\text{осн}}$  – основная заработная плата,  $Z_{\text{доп}}$  – дополнительная заработная плата.

Таблица 14 – Расчет страховых отчислений

Исполнители	$k_{\text{внеб}}$	$Z_{\text{доп}}$ руб	$Z_{\text{осн}}$ , руб	$Z_{\text{внеб}}$
Исполнитель	0,30	7022,3124	58519,27	19662,47
Руководитель	0,30	3540,2562	29502,135	9912,71
Консультант по техническим вопросам и вопросам проектирования	0,30	2170,2324	18085,27	6076,65
Итого				29575,18

Рассчитаем накладные расходы, учитывающие прочие затраты, не попавшие в предыдущие статьи расходов. Их величина определяется согласно следующей формуле:

$$Z_{\text{накл}} = k_{\text{накл}} * S$$

,где  $S$  определяется как сумма статей расходов, а  $k_{\text{накл}}$  - коэффициент накладных расходов, принятый за 16 %

Таблица 15 – Расчет накладных расходов

Статьи затрат	Сумма,руб
Материальные затраты	1462
Затраты на амортизацию	3815,6
Затраты на основную заработную плату	88021,405
Затраты на дополнительную заработную плату	10562,5686
Накладные расходы	16617,85

Результаты итогового бюджета отображены в следующей таблице

Таблица 16 – Бюджет затрат на разработку

Наименование	Сумма,руб	Удельный вес, %	Сумма,руб	Удельный вес, %
	Текущая разработка		Разработка конкурента	
Материальные затраты	1462	0,9	2211	0,4
Затраты на амортизацию	3815,6	2,5	3815	0,68
Затраты на основную заработную плату	88021,405	58,6	85412	15
Затраты на дополнительную заработную плату	10562,5686	7	91121	16
Затраты на отчисления во внебюджетные фонды	29575,18	19,7	354128	63
Накладные расходы	16617,85	11	23154,7	4
Итого	150054,6	100	559841,7	100

### 3.9 Оценка сравнительной эффективности исследования

Определение эффективности происходит на основе расчёта интегрального показателя эффективности научного исследования. Его нахождение связано с определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности. Интегральный финансовый показатель разработки определяется по следующей формуле:

$$l_{\text{исп}} = \frac{\Phi_{pi}}{\Phi_{\text{max}}}$$

где:  $l_{\text{исп}}$  – интегральный финансовый показатель разработки;  $\Phi_{pi}$  – стоимость  $i$ -го варианта исполнения;

$\Phi_{\text{max}}$  – максимальная стоимость исполнения научно- исследовательского проекта (в т.ч. аналоги).

$$l_{\text{исп}} = \frac{150054,6}{559841,7} = 0,26$$

Интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования можно определить по следующей формуле:

$$l_{pi} = \sum a_i * b_i$$

где  $l_{pi}$ : – интегральный показатель ресурсоэффективности для  $i$ -го варианта исполнения разработки;

$a_i$  – весовой коэффициент  $i$ -го варианта исполнения разработки;

$b_i$  – балльная оценка  $i$ -го варианта исполнения разработки, устанавливается экспертным путём по выбранной шкале оценивания;

Расчёт интегрального показателя отображён в таблице.

Таблица 17 – Сравнительная оценка характеристик вариантов исполнения проекта

Критерии ПО	Весовой коэффициент критерия	Программа построения онтологических моделей на основе анализа текстовых корпусов	Конкурентная разработка программного обеспечения protege для отображения онтологических моделей
Простота использования	0,15	5	3
Эффективность выполнения задачи	0,15	5	4
Производительность с точки зрения использования технических ресурсов	0,15	3	4
Функциональность	0,15	4	5
Конкурентоспособность	0,2	4	4
Повышение производительности труда пользователя	0,2	5	5
Итого	1	26	25

$$l_{pm} = 5 * 0,15 + 5 * 0,15 + 3 * 0,15 + 4 * 0,15 + 4 * 0,2 + 5 * 0,2 = 4,35$$

$$l_a = 3 * 0,15 + 4 * 0,15 + 4 * 0,15 + 5 * 0,15 + 4 * 0,2 + 5 * 0,2 = 3,6$$

Интегральный показатель эффективности разработки текущего проекта  $l_{p \text{ финр}}$  и проекта конкурента  $l_{a \text{ финр}}$  определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$l_{p \text{ финр}} = \frac{I_{p \text{ м}}}{I_{p \text{ ф}}} \quad l_{a \text{ финр}} = \frac{I_{a \text{ м}}}{I_{a \text{ ф}}}$$

Сравнение интегрального показателя эффективности текущего проекта и проекта конкурента позволит определить сравнительную эффективность проекта. Сравнительная эффективность проекта определяется по формуле:

$$\mathcal{E}_{\text{сред}} = \frac{I_{p \text{ финр}}}{I_{a \text{ финр}}}$$

где:

$\mathcal{E}_{\text{сред}}$  – сравнительная эффективность проекта;

$I_{p \text{ м}}$  – интегральный показатель разработки;

$I_{p \text{ ф}}$  – интегральный технико-экономический показатель конкурентной разработки

Таблица 18 – Сравнительная эффективность разработки

№ п/п	Показатели	Программа построения онтологических моделей на основе анализа текстовых корпусов	Конкурентная разработка программного обеспечения protege для отображения онтологических моделей
1	Интегральный финансовый показатель разработки	0,26	1
2	Интегральный показатель ресурсоэффективности разработки	4,35	3,6
3	Интегральный показатель	16,7	3,6

	эффективности		
4	Сравнительная эффективность вариантов исполнения	1	1

Таким образом, анализ конкурентных решений данного проекта показал способность интеграции проекта в коммерческую деятельность. Готовность к коммерциализации оценена как средняя. SWOT анализ определил потенциальные пути улучшения решения, представляющие собой оптимизацию исходного кода программного обеспечения, своевременной дополнение функционала и добавление технической поддержки. Общий бюджет затрат на разработку данного проекта, включающий в себя основную и дополнительную заработные платы, расхода на амортизацию и внебюджетные расходы составляет 150054,6 рублей. Сравнение эффективности текущего проекта и его аналога показало, что текущий проект достаточно эффективен

## **ГЛАВА 4. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ**

Данная работа посвящена разработке программного обеспечения для построения онтологической модели предметной области на основе анализа текстовых корпусов. Данное программное обеспечение может быть применено в таких областях, как оптимизация поисковых систем, формализация задач бизнес-процессов, то есть отображение задач и направлений развития какого-либо предприятия, в учебных заведениях для создания дорожных карт. Функциональные возможности данного программного обеспечения позволят облегчить восприятие информации в контексте определённой предметной области, повысить эффективность усвоения новых знаний.

Работа выполнялась с использованием персональной электронно-вычислительной машины на базе отделения информационных технологий Инженерной школы информационных технологий и робототехники Томского политехнического университета

### **4.1 Правовые и организационные вопросы обеспечения безопасности**

Основным документом, регулирующим отношения между работником и работодателем, является Трудовой Кодекс Российской Федерации (ТК РФ). По трудовому кодексу, нормальная продолжительность рабочего времени не может превышать 40 часов в неделю.

При заключении трудового договора в нем может быть предусмотрен испытательный срок с целью проверки его соответствия получаемой работе.

В период испытания на работника распространяются положения трудового законодательства и иных нормативных правовых актов, содержащих нормы трудового права, коллективного договора, соглашений, локальных нормативных актов.

Испытательный срок не может быть больше трёх месяцев, для руководителей организаций, их заместителей и руководителей филиалов испытательный срок не должен превышать шести месяцев, если иное не установлено федеральным законом.

При заключении трудового договора на срок от двух до шести месяцев испытание не может превышать двух недель. В срок испытания не засчитываются период временной нетрудоспособности работника и другие периоды, когда он фактически отсутствовал на работе. Заработная плата – является вознаграждением за работу в соответствии с квалификацией работника, сложности, количества трудовых часов и условий выполняемой работы.

Выплата заработной платы может проводиться в денежной форме, в валюте Российской Федерации (в рублях). В случаях, предусмотренных законодательством Российской Федерации в валютном регулировании и валютном контроле, выплата заработной платы может производиться в иностранной валюте.

Согласно ГОСТ 12.2.033-78 «Рабочее место при выполнении работ сидя», рабочее место должно обеспечивать выполнение трудовых операций в пределах зоны досягаемости моторного поля.

При проектировании оборудования и организации рабочего места следует учитывать антропометрические показатели женщин (если работают только женщины) и мужчин (если работают только мужчины); если оборудование обслуживают мужчины и женщины — общие средние показатели мужчин и женщин.

Организация рабочего места и конструкция оборудования должны обеспечивать прямое и свободное положение корпуса тела работающего или наклон его вперед не более чем на 15°.

Конструкцией производственного оборудования и организацией рабочего места должно быть обеспечено оптимальное положение работающего, которое достигается регулированием: высоты рабочей поверхности.

В тех случаях, когда невозможно осуществить регулирование высоты рабочей поверхности и подставки для ног, допускается проектировать и

изготавливать оборудование с нерегулируемой высотой рабочей поверхности и подставки для ног.

Для обеспечения удобного, возможно близкого подхода к столу, станку или машине должно быть предусмотрено пространство для стоп размером не менее 150 мм по глубине, 150 мм по высоте и 530 мм по ширине.

Согласно Федеральному закону от 28 декабря 2013 г. N 426-ФЗ «О специальной оценке условий труда», Регулирование специальной оценки условий труда осуществляется Трудовым кодексом Российской Федерации, настоящим Федеральным законом, другими федеральными законами и иными нормативными правовыми актами Российской Федерации.

Нормы, регулирующие специальную оценку условий труда и содержащиеся в федеральных законах и иных нормативных правовых актах Российской Федерации, должны соответствовать нормам Трудового кодекса Российской Федерации и настоящего Федерального закона.

Специальная оценка условий труда является единым комплексом последовательно осуществляемых мероприятий по идентификации вредных и (или) опасных факторов производственной среды и трудового процесса (далее также - вредные и (или) опасные производственные факторы) и оценке уровня их воздействия на работника.

По результатам проведения специальной оценки условий труда устанавливаются классы (подклассы) условий труда на рабочих местах.

В соответствии с ГОСТ Р 50948-2001. «Средства отображения информации индивидуального пользования. Общие эргономические требования и требования безопасности.», для безопасности к визуальным параметрам, яркость знака должна быть не менее 20 кд/м<sup>2</sup> для плоских дискретных экранов.

Неравномерность яркости рабочего поля экрана должна быть не более 20%.

Неравномерность яркости элементов знака должна быть не более 20%.

Яркостный контраст изображения должен быть не менее 3:1. Яркостный контраст внутри знака и между знаками быть не менее 3:1. Ширина контура знака должна быть от 0,25 до 0,5 мм. Временная нестабильность изображения на плоских дискретных экранах не должна быть зафиксирована. Частота обновления должна составлять для плоских дискретных экранов не менее 60 Гц.

#### **4.2 Производственная безопасность**

Согласно ГОСТ 12.0.002-2014 деятельность человека – это активное взаимодействие человека с окружающим миром, в ходе которого человек воздействует на определённый объект с целью удовлетворения своих потребностей.

Производственной деятельностью является вид деятельности человека, направленный на создание экономического продукта с целью создания необходимых условий для своего существования и развития социально-трудовых отношений.

В соответствии с ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация» неблагоприятные производственные факторы на организм работника подразделяются на следующие группы: 1) вредные производственные факторы, то есть такие факторы, которые приводят к заболеванию, а также усугубляющие уже имеющиеся заболевания. 2) опасные производственные факторы, то есть факторы, приводящие к травмам, в том числе и смертельным.

Так же вредные производственные факторы классифицируются на: 1) факторы, приводящие к хроническим заболеваниям 2) факторы, приводящие к острым заболеваниям или травмам за счёт относительно воздействия высокой интенсивности.

В данной работе целесообразно рассмотреть вредные факторы, влияющие на работу инженера-программиста как разработчика изучаемого в работе программного обеспечения.

Таблица 1 – Возможные опасные и вредные производственные факторы на рабочем месте

Факторы (ГОСТ 12.0.003-2015)	Нормативные документы
Опасные и вредные производственные факторы, связанные с аномальными микроклиматическими параметрами воздушной среды на месте работающего: температурой, влажностью воздуха, скоростью движения воздуха относительно тела работника, тепловым излучением окружающих предметов	ГОСТ 12.1.005.88 «Система стандартов безопасности труда. Общие санитарно-гигиенические требования к воздуху рабочей зоны»
Производственные факторы, обладающие свойствами психофизиологического воздействия на организм человека, а именно нервно-психические перегрузки, связанные с напряженностью трудового процесса	МР 2.2.9.2311 – 07 «Профилактика стрессового состояния работников при различных видах профессиональной деятельности»
Опасные и вредные производственные факторы, связанные повышенным уровнем шума на рабочем месте	ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности.
Опасные и вредные производственные факторы, связанные с отсутствием и недостатком естественного освещения	СП 52.13330.2016 Естественное и искусственное освещение

#### **4.2.1 Опасные и вредные производственные факторы, связанные с аномальными микроклиматическими параметрами воздушной среды**

Для установления допустимых норм микроклиматических параметров чистоты воздуха на рабочем месте в помещении необходимо периодически проводить вентиляцию пространства.

Согласно ГОСТ 12.1.005.88 «Система стандартов безопасности труда. Общие санитарно-гигиенические требования к воздуху рабочей зоны» показателями, характеризующими микроклимат, являются:

1. Температура воздуха;

2. Относительная влажность воздуха;

3 Скорость движения воздуха;

Оптимальные показатели микроклимата распространяются на всю рабочую зону, допустимые показатели устанавливаются дифференцированно для постоянных и непостоянных рабочих мест.

При обеспечении оптимальных показателей микроклимата температура внутренних поверхностей конструкций, ограждающих рабочую зону (стен, пола, потолка и др.), или устройств (экранов и т. п.). А также температура наружных поверхностей технологического оборудования или ограждающих его устройств не должны выходить более чем на 2 °С за пределы оптимальных величин температуры воздуха, для отдельных категорий работ. При температуре поверхностей ограждающих конструкций ниже или выше оптимальных величин температуры воздуха рабочие места должны быть удалены от них на расстояние не менее 1 м. Температура воздуха в рабочей зоне, измеренная на разной высоте и в различных участках помещений, не должна выходить в течение смены за пределы оптимальных величин, указанных для отдельных категорий работ.

В соответствии с ГОСТ 12.1.005-88. «Межгосударственный стандарт. Система стандартов безопасности труда. Общие санитарно-гигиенические требования к воздуху рабочей зоны», в кабинах, на пультах и постах управления технологическими процессами, в залах вычислительной техники и других производственных помещениях при выполнении работ операторского типа, связанных с нервно-эмоциональным напряжением, должны соблюдаться оптимальные величины температуры воздуха 22—24 °С, его относительной влажности 40—60 % и скорости движения (не более 0.1 м/с).

При обеспечении допустимых показателей микроклимата температура внутренних поверхностей конструкций, ограждающих рабочую зону (стен, пола, потолка и др.). или устройств (экранов и т.п.) не должна выходить за пределы допустимых величин температуры воздуха, установленных в для

отдельных категорий работ. Перепад температуры воздуха по высоте рабочей зоны при всех категориях работ допускается до 3 °С.

Колебания температуры воздуха по горизонтали в рабочей зоне, а также в течение смены допускаются до 4 °С, при этом абсолютные значения температуры воздуха, измеренной на разной высоте и в различных участках помещений в течение смены, не должны выходить за пределы допустимых величин.

При обеспечении оптимальных и допустимых показателей микроклимата в холодный период года следует применять средства защиты рабочих мест от радиационного охлаждения от остекленных поверхностей оконных проемов, в теплый период года — от попадания прямых солнечных лучей.

#### **4.2.2 Производственные факторы, обладающие свойствами психофизиологического воздействия на организм человека**

Согласно МР 2 2 9 2311-07 основными факторами психофизиологического воздействия на человека являются: 1) умственная нагрузка вследствие ненормированного рабочего дня 2) Зрительная нагрузка в связи с координацией зрительных органов с системой органов движения.

Основными причинами стресса для работников умственного труда являются недостаточная двигательная активность, длительное пребывание в неудобной позе.

Формирование производственного стресса сопровождается постепенным переходом функционального состояния от напряжения к утомлению. Для снижения стрессовых факторов необходимо придерживаться рекомендации к организации режимов труда и отдыха работников умственного вида деятельности.

Таблица 2 – Рекомендуемый рациональный режим труда и отдыха для лиц умственного труда при 8-часовом рабочем дне

Перерыв	Время проведения	Продолжительность		Заполнение перерыва
		1 группа	2 группа	
2	3	4	5	6

Утренняя смена				
Регламентированный перерыв	Через 2 ч от начала работы	5—7 мин	7— 10 мин	Комплекс физических упражнений на координацию, точность и быстроту движений, концентрацию внимания
Обеденный перерыв	Через 4 ч от начала работы	30 мин	60 мин	Обед в столовой, отдых
Регламентированный перерыв	Через 6 ч от начала работы	5—7 мин	10— 15 мин	Психологическая разгрузка
Микропаузы	Индивидуально по мере необходимости	20—40 с	40с— 3 мин	Микропаузы активного отдыха

С учетом непрерывности процесса работы рекомендуется включать в режим рабочего дня 2 регламентированных перерывов и микропауз активного отдыха. Второй перерыв рекомендуется использовать для психологической разгрузки.

Для предупреждения преждевременной утомляемости за работой с ПЭВМ необходимо чередовать работу с ПЭВМ и без него.

При возникновении зрительной нагрузки, не смотря на соблюдение санитарно-эпидемиологических норм, рекомендуется индивидуально ограничить время за проведением за ПЭВМ.

При работе с ПЭВМ в ночную смену независимо от вида трудовой деятельности продолжительность перерывов рекомендуется увеличить на 30%.

Продолжительность непрерывной работы с ВДТ без регламентированного перерыва в соответствии с санитарными нормами и правилами не должна превышать 1 ч.

Таблица 3 – Суммарное время регламентированных перерывов в зависимости от продолжительности работы, вида и категории трудовой деятельности с ПЭВМ

Категория работы ПЭВМ	Уровень нагрузки за смену при работе с ПЭВМ			Суммарное время перерывов, мин	
	группа А, количество знаков	группа Б, количество знаков	группа В, количество знаков	при 8-часовой смене	при 12-часовой смене
I	До 20 000	До 15 000	До 2	50	80
II	До 40 000	До 30 000	До 4	70	110
III	До 60 000	До 40 000	До 6	90	140

Для снижения эмоционального напряжения необходимо включить в распорядок дня индивидуальные сеансы «психологической разгрузки». Их можно начинать проводить только после овладения основами и техникой их выполнения под руководством инструктора или психоневролога. Для этого необходимо в плане оздоровительных мероприятий предусмотреть достаточное количество занятий с привлечением специалистов соответствующего профиля. Сеансы «психологической разгрузки» продолжительностью 10—20 мин рекомендуется проводить ежедневно в одно и то же время, например, в часы обеденного перерыва, после приема пищи или во второй половине рабочего дня, когда начинает снижаться работоспособность. Место проведения должно быть постоянным или в комнате «психологической разгрузки», или в собственном рабочем кабинете.

#### **4.2.3 Опасные и вредные производственные факторы, связанные повышенным уровнем шума на рабочем месте**

В соответствии с ГОСТ 12.1.003 «Система стандартов безопасности труда. Шум. Общие требования безопасности», шумом являются звуковые колебания в диапазоне слышимых частот, способные оказать вредное воздействие на безопасность и здоровье работника. Повышенный шум на

рабочем месте оказывает вредное влияние на организм работника в целом, вызывая неблагоприятные изменения в его органах и системах. Длительное воздействие такого шума способно привести к развитию у работника потери слуха, увеличению риска артериальной гипертензии, болезней сердечно-сосудистой, нервной системы и др. При этом специфическим клиническим проявлением вредного действия шума является стойкое нарушение слуха (тугоухость), рассматриваемое как профессиональное заболевание.

Различают три основных вида тугоухости в зависимости от того, в какой из систем слухового тракта наблюдаются патологические изменения:

1. Звукопроводения (кондуктивная тугоухость).
2. Звуковосприятия (нейросенсорная или перцептивная тугоухость) или в обоих видов (смешанный вид тугоухости).
3. Кондуктивная тугоухость обусловлена изменением подвижности барабанной перепонки и цепи слуховых косточек.
4. Нейросенсорная тугоухость развивается при повреждении чувствительных нервных клеток внутреннего уха, слухового нерва и центральных образований слуховой системы.

От своевременного диагностирования тугоухости на начальной стадии ее развития зависит эффективность профилактических мероприятий, предупреждающих развитие профессионального заболевания.

Для выявления негативного воздействия данного фактора необходимо оценить уровень шума на рабочем месте. Он заключается в сопоставлении результата измерения нормируемой величины с гигиеническим нормативом и принятия на основе этого решения о соответствии или несоответствии условий труда на данном рабочем месте безопасным с точки зрения шумового воздействия на работника. При сопоставлении с гигиеническим нормативом результат измерения и параметр неопределенности должны быть представлены в децибелах с точностью до одного десятичного знака после запятой.

Характер воздействия шума на работника может сильно зависеть от рабочего места. Для некоторых рабочих мест шум может быть хорошо определен, стабилен в течение рабочей смены и мало изменяться день ото дня. В этом случае неопределенность измерения нормируемой величины в наибольшей степени определяется инструментальной составляющей и установкой измерительного микрофона. В других случаях шум может быть непостоянен, существенно зависеть от дня недели, сезона, рабочего задания и многих других факторов. Часто для таких рабочих мест бывает сложно выделить и описать характерные рабочие операции, а аналогичные рабочие места, по которым можно было бы провести усреднение, отсутствуют или число их невелико.

Согласно ГОСТ 12.1.003-2014 «Система стандартов безопасности труда. Шум. Общие требования безопасности Допустимые уровни шума на рабочем месте», основными мерами по защите работника от шума является выполнение работодателем ответственности по обеспечению защиты работника от шумового воздействия.

Эти меры могут включать в себя, в частности:

1. Оценку риска потери слуха работником;
2. Проектирование рабочих мест с учетом допустимого уровня риска;
3. Создание условий труда, при которых вредное воздействие шума не усугубляется;
4. Привлечение к работам лиц, не имеющих медицинских противопоказаний по шуму, и обеспечение прохождения ими регулярных медицинских обследований с применением средств аудиометрии;
5. Составление комплексных программ сохранения слуха работников;

#### **4.2.4 Опасные и вредные производственные факторы, связанные с отсутствием и недостатком естественного освещения**

Согласно СП 52.13330.2016 «Естественное и искусственное освещение» основными требованиями к освещению при работе с ПЭВМ являются:

1. Освещенность рабочего места оператора на рабочем столе в горизонтальной плоскости от общего искусственного освещения должна быть от 300 до 500 лк.

2. Освещенность на пюпитре в вертикальной плоскости должна быть не менее 300 лкс.

3. Для освещения зоны расположения документов допускается установка светильников местного освещения.

Отношение яркостей в зоне наблюдения (экран, документ, поверхность стола) должно быть не более 10:1.

В поле зрения оператора должны отсутствовать прямая и отраженная блёскость.

Для снижения блескости необходимо:

1. Оборудовать светопроемы солнцезащитными устройствами (шторами, регулируемыми жалюзи, внешними козырьками и т.д.);

2. Использовать для общего освещения светильники с рассеивателями и экранирующими;

3. Решетками, яркость которых в зоне углов излучения более 50° от вертикали не должна превышать 200 кд/м<sup>2</sup>;

4. Использовать для местного освещения светильники с непросвечивающим отражателем и защитным углом не менее 40°;

5. Размещать рабочий стол так, чтобы оконный проем находился сбоку (справа или слева), при этом дисплей должен располагаться на поверхности стола справа или слева от оператора;

6. Размещать рабочий стол между рядами светильников общего освещения;

7. Использовать дисплей, имеющий антибликовое покрытие экрана или антибликовый фильтр;

Рассчитаем общее равномерное искусственное освещение горизонтальной рабочей поверхности.

Для расчёта равномерного искусственного освещения воспользуемся формулой  $\Phi = \frac{E_n \cdot S \cdot K_3 \cdot Z}{N_{л.}}$

Рассчитаем индекс помещения  $i = \frac{S}{h \cdot (A+B)}$ , где  $S$  площадь помещения,  $h$  – расчетная высота подвеса светильников над рабочей поверхностью;  $A$  – длина помещения;  $B$  – ширина помещения.

Длина помещения  $A$  составляет 4 м, ширина  $B$  – 3 м, расчетная высота подвеса светильников над рабочей поверхностью  $h$  – 3,4 м.

Индекс помещения  $i$  составляет  $i = \frac{3 \cdot 4}{3,4(3+4)} = 1,15$ .

Найдем коэффициенты отражения поверхностей стен, потолка и рабочей поверхности.

Итак, нормативная освещённость помещения  $E_n$  согласно СНИП 23-05-95 210 лк, работа велась в помещении площадью  $S = 12 \text{ м}^2$ , коэффициент запаса учитывающий загрязнение источника света  $K_3$  составляет 1,5, коэффициент неравномерности освещения  $Z = 1,1$ , число ламп в помещении  $N_{л.}$  5.

Так как поверхности стен аудитории окрашены в светлый цвет, то коэффициенты отражения стен равны 50%, поверхности потолков аудитории также светлого цвета, коэффициенты отражения потолков равны 50%, рабочие поверхности средней светлости, коэффициенты отражения рабочих поверхностей равны 30%. Учитывая коэффициенты отражения поверхностей стен, потолка, рабочей поверхности и индекс помещения, определяем значения коэффициентов  $\eta_a = 34\%$ ,  $\eta_k = 41\%$

Рассчитаем световой поток источника света  $\Phi = \frac{210 \cdot 12 \cdot 1,5 \cdot 1,1}{5 \cdot 0,41} = 2028 \text{ лм}$ .

Ближайшая стандартная лампа имеет поток 2100 лм.

Проверим выполнения условия  $-10\% \leq \frac{\Phi_{\text{л.станд}} - \Phi_{\text{л.расч}}}{\Phi_{\text{л.станд}}} \cdot 100\% \leq +20\%$ . В результате получим 34%.

Найдём расстояние между светильниками  $L = \lambda \cdot h$ , где  $\lambda = 1,5$ :

$$L = 1,5 \cdot 2,3 = 3,45 \text{ м.}$$

Расстояние от крайнего ряда светильников до стены:

$$\frac{L}{3} = 3,45/3 = 1,15 \text{ м.}$$

Определим количество рядов светильников:

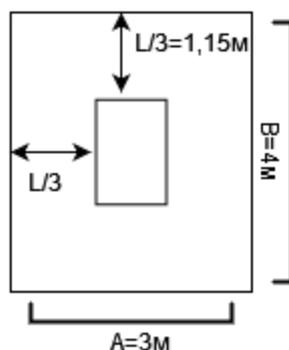
$$N_{\text{ряд}} = \frac{B - \frac{2}{3} \cdot L}{L} + 1 = \frac{4 - \frac{2}{3} \cdot 3,45}{3,45} + 1 = 1.$$

Рассчитаем количество светильников в ряду:

$$N_{\text{св}} = \frac{A - \frac{2}{3} \cdot L}{L} + 1 = \frac{3 - \frac{2}{3} \cdot 3,45}{3,45} + 1 = 1.$$

Светильники расположены в 1 ряд в количестве 1 штука.

Рисунок 1 – План помещения и размещения светильников с люминесцентными лампами



### 4.3 Экологическая безопасность

Предмет исследования не оказывает никакого влияния на окружающую среду, так как ЭВМ не производит выбросов в атмосферу вредных веществ.

При окончании срока эксплуатации ЭВМ электролитические конденсаторы (размер хотя бы одной из сторон конденсатора должен быть 25 мм или более) должны быть извлечены из ОЭЭО и направлены на размещение в соответствии с установленными требованиями безопасности.

Любой конденсатор или трансформатор, содержащий ПХБ, должен быть извлечен и размещен в соответствии с требованиями безопасности, установленными в отношении отходов класса опасности I.

Согласно с ГОСТ Р 55102- 2012 «Ресурсосбережение. Обращение с отходами. Руководство по безопасному сбору, хранению, транспортированию и разборке отработавшего электротехнического и электронного оборудования, за исключением ртутисодержащих устройств и приборов», при утилизации отходов электронного оборудования следующие элементы должны быть собраны отдельно должно быть собрано отдельно:

1. Конденсаторы.
2. Пластик.
3. Платы устройств с площадью поверхности более 10см<sup>2</sup>.
4. Элементы отработавшего электротехнического и электронного оборудования – содержат свинец, кадмий, олово.
5. Люминесцентные лампы.
6. Электронно-лучевые трубки.

Также существует определённый порядок сбора, накопления, хранения, обезвреживания и транспортировки определённой части осветительных устройств, чьё ненадлежащее выполнение может повлечь за собой причинение вреда жизни и здоровью человека, и окружающей среде.

#### **4.4 Безопасность в чрезвычайных ситуациях**

Рабочее место с ПЭВМ может представлять собой угрозу пожарной опасности. Согласно НПБ 105-03 «Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности», рабочее место относится к классу пожароопасных, так как ПЭВМ представляет собой источник, обладающий достаточной мощностью для возникновения пожара.

В соответствии с СП 1.13130 «Системы противопожарной защиты. эвакуационные пути и выходы», для предотвращения воздействия пожароопасных факторов необходимо предпринимать следующие меры:

1. Применить объемно-планировочные решения и средства, обеспечивающих ограничение распространения пожара за пределы очага.

2. Монтировать эвакуационные пути, удовлетворяющие требованиям безопасной эвакуации людей при пожаре.

3. Монтировать системы обнаружения пожара (установок и систем пожарной сигнализации), оповещения и управления эвакуацией людей при пожаре.

4. Применить первичные средства пожаротушения.

Для обеспечения безопасной эвакуации людей, эвакуационные выходы должны:

1. Быть установлены в необходимых количествах, размерах и в соответствующем конструктивном исполнении эвакуационных путей и эвакуационных выходов.

2. Обеспечивать беспрепятственное движение людей по эвакуационным путям и через эвакуационные выходы.

3. Быть организованы оповещение и управление движением людей по эвакуационным путям (в том числе с использованием световых указателей, звукового и речевого оповещения).

К противопожарным средствам, которые должны быть в наличии относятся:

1. Огнетушители.
2. Пожарные щитки.
3. Пожарные шкафы.

Сотрудники должны уметь пользоваться средствами противопожарной защиты, знать план эвакуации, расположение эвакуационных выходов и автоматических средств пожаротушения.

#### **4.5 Выводы по разделу социальная ответственность**

Значение всех производственных факторов на изучаемом рабочем месте соответствует нормам, которые также были продемонстрированы в данном разделе, за исключением фактора, обладающего свойствами психофизиологического воздействия на организм человека. Для минимизации

влияния данного фактора на организм человека, достаточно соблюдать меры, приведенные в МР 2.2.9.2311 – 07 «Профилактика стрессового состояния работников при различных видах профессиональной деятельности».

Согласно стандарту ПУЭ «Помещения по опасности поражения электрическим током» рабочее помещение относится к классу помещений без повышенной опасности.

Рабочий персонал относится к I группе безопасности, то есть без присвоения удостоверений, так как специфика работы не подразумевает соответствующую квалификацию.

Категория тяжести в аудитории ТПУ в соответствии с СанПиН 1.2.3685-21 "Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания" относится к категории Ia (работы, производимые сидя и сопровождающиеся незначительным физическим напряжением с интенсивностью энергозатрат до 120 ккал/час).

Согласно СП 12.13130.2009 «Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности» рабочее помещение относится к категории пожароопасности В3. Характеристика веществ и материалов, находящихся в помещении: горючие и трудногорючие жидкости, твердые горючие и трудногорючие вещества и материалы (в том числе пыли и волокна), вещества и материалы, способные при взаимодействии с водой, кислородом воздуха или друг с другом только гореть.

Данный рабочий объект оказывает незначительное негативное воздействие на окружающую среду.

## ЗАКЛЮЧЕНИЕ

В процессе выполнения данной работы были выполнены следующие задачи:

1. Изучена соответствующая литература на предмет исследования сущностей онтологии и онтологической модели.
2. Были выявлены методы построения онтологических моделей.
3. Спроектировано и разработано программное обеспечение автоматического построения онтологических моделей предметной области на основе анализа текстовых корпусов.

Разработка программного обеспечения включает в себя:

1. Проектирование и разработка модуля анализа текстовых корпусов (парсера).
2. Проектирование и разработка модуля распознавания именованных сущностей.
3. Проектирование и разработка модуля синтаксического анализа текстовых корпусов.
4. Проектирование и разработка модуля статистического анализа онтологий.
5. Проектирование и разработка модуля представления результата.

## СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. An ontology for search applications / R. Gupta, A. Halevy, X. Wang [и др.]. — Текст : непосредственный // Proc. VLDB Endow. — 2014. — № 1. — С. 312-370.
2. Atzori The Internet of Things: A survey / Atzori, L., G. Iera. — Текст : непосредственный // Computer Networks. — 2010. — № 1. — С. 121-132.
3. Ballou, D. Modeling data and process quality in muliti-input, multi-output information systems / D. Ballou, H. Pazer. — Текст : непосредственный // Integration of Knowledge Intensive Multi-Agent Systems. — 2001. — № 1. — С. 110-162.
4. Barrero, D. F. Automatic web data extraction based on genetic algorithms and regular expressions / D. F. Barrero, D. Moreno, M. Camacho. — Текст : непосредственный // Springer. — 2009. — № 9. — С. 112-132.
5. Berners-Lee, T. Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor / T. Berners-Lee, M. Fischetti. — Текст : непосредственный // Harper. — 1999. — № 1. — С. 1-11.
6. Bessmertny Knowledge visualization based on semantic networks / Bessmertny, I.. — Текст: непосредственный // Program Comp. Soft.. — 2010. — № 4. — С. 31-36.
7. Buitelaar, P. Ontology learning from text: methods, evaluation and applications / P. Buitelaar, C. Magnini. — Текст : непосредственный // Frontiers in Artificial Intelligence and Applications. — 2007. — № 123. — С. 32-40.
8. Compton, M. The SSN Ontology of the Semantic Sensor Networks Incubator Group / M. Compton, P. Barnaghi, L. Bermudez. — Текст : непосредственный // Journal of Web Semantic. . — 2011. — № 1. — С. 100-111.
9. Conceptual Knowledge Markup Language . — Текст : электронный // ontologos. : [сайт]. — URL: <https://www.ontologos.org/CKML/CKML%200.2.html> (дата обращения: 19.06.2023).

10. Corcho, O. Overview and analysis of methodologies for building ontologies / O. Corcho, M. Fernandez-Lopez. — Текст : непосредственный // The Knowledge Engineering Review . — 2002. — № 1. — С. 1-32.
11. Gattani, A. 37. Entity extraction, linking, classification, and tagging for social media. / A. Gattani, A. Doan, D. S. Lamba. — Текст : непосредственный // Proc. VLDB Endow. — 2013. — № 11. — С. 110-211.
12. Gómez-Pérez, A. Ontology Languages for the Semantic Web / A. Gómez-Pérez, O. Corcho. — Текст : непосредственный // International Journal on Semantic Web and Information Systems, special issue on sensor networks, Internet of Things and smart devices. — 2002. — № 1. — С. 43-60.
13. Jurafsky, D. Speech and language processing. An introduction to natural language processing, computational linguistics, and speech recognition / D. Jurafsky, J. H. Martin. — Текст : непосредственный // Upper Saddle River N.J. — 2009. — № 9. — С. 734.
14. Knowledge Annotation Initiative of the Knowledge Acquisition Community . — Текст : электронный // hcs.science.uva : [сайт]. — URL: <http://hcs.science.uva.nl/usr/richard/ka2/presentation/ppframe.htm> (дата обращения: 19.06.2023).
15. Knowledge Representation in the Internet of Things: Semantic Modelling and its Applications / W. Wang, D. Suparna, C. Gilbert, M. Klaus. — Текст : непосредственный // Knowledge Representation in IoT. — 2012. — № 1. — С. 23-56.
16. Michael Grüninger, Mark S. Fox. Methodology for the Design and Evaluation of Ontologies, 1995
17. Natasha library documentation . — Текст : электронный // github : [сайт]. — URL: <https://github.com/natasha/natasha/blob/master/README.md?ysclid=lj2ixzx61v415610743> (дата обращения: 19.06.2023).

18. Networkx Documentatio. — Текст : электронный // networkx: [сайт]. — URL: <https://networkx.org/documentation/stable/tutorial.html> (дата обращения: 13.06.2023).
19. Nugumanova, A. Applying the latent semantic analysis to the issue of automatic extraction of collocations from the domain texts, in knowledge engineering and the semantic / A. Nugumanova, I. Bessmertny. — Текст : непосредственный // Springer. — 2013. — № 7. — С. 78-101.
20. Ontology population and enrichment: state of the art – knowledge-driven multimedia information extraction / G. Petasis, V. Karkaletsis, G. Paliouras, K. Zavitsanos. — Текст : непосредственный // Springer. — 2011. — № 4. — С. 110-166.
21. Platzer, C. Web service clustering using multidimensional angles as proximity measures / C. Platzer, F. Rosenberg, S. Dustdar. — Текст : непосредственный // ACM Trans. — Internet Technol. — 2009. — № 1. — С. 1-26.
22. Proposed Specification, SHOE Project . — Текст : электронный // cs.umd.edu : [сайт]. — URL: <https://www.cs.umd.edu/projects/plus/SHOE/spec1.01.htm> (дата обращения: 16.06.2023).
23. Resource Description Framework (RDF) Schema Specification. — Текст: электронный // w3: [сайт]. — URL: <https://www.w3.org/TR/PR-rdf-schema> (дата обращения: 19.06.2023).
24. Segev Context-Based Matching and Ranking of Web Services for Composition / Segev, A., E. Toch. — Текст : непосредственный // IEEE Transactions on Services Computing.. — 2009. — № 3. — С. 222.
25. Semantics for the Internet of Things: early progress and back to the future / P. Barnaghi, W. Wang, C. Henson, TaylorK.. — Текст : непосредственный // International Journal on Semantic Web and Information Systems, special issue on sensor networks, Internet of Things and smart devices. — 2012. — № 8. — С. 5-12.
26. Uschold, M. Ontologies: principles, methods and applications / M. Uschold, M. Gruninger. — Текст : непосредственный // University of Edinburg. — 1996. — № 1. — С. 2-15.

27. Андрич, О. Ф. Исследование методов оценки качества онтологических моделей / О. Ф. Андрич, Л. А. Макушкина. — Текст: непосредственный // Современные наукоемкие технологии. — 2014. — № 5. — С. 18-19.
28. Анисов А.М. Современная логика /А.М Анисов. — М., 2002.
29. Гаврилова, Т. А. Использование онтологий в системах / Т. А. Гаврилова. — 1. — Дивноморское: Физмалит, 2001. — 150 с. — Текст: непосредственный.
30. Горшков Онтологическое моделирование / Горшков, С.. — 1. — Москва: Тринидата, 2016. — 150 с. — Текст: непосредственный.
31. ГОСТ 12. 1.0 0 5 -8 8 Общие санитарно-гигиенические требования к воздуху рабочей зоны: дата введения 2008. – Москва: Стандартформ, 2008. – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный.
32. ГОСТ 12.1.003—2014 Система стандартов безопасности труда ШУМ Общие требования безопасности: дата введения 2014. – . – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный.
33. ГОСТ 12.1.019— 2009 Электробезопасность: дата введения 2009. – . – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный.
34. ГОСТ 12.2.033-78 Рабочее место при выполнении работ стоя : дата введения 1979. – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный.
35. ГОСТ Р 50948–2001 Средства отображения информации индивидуального пользования. – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный..
36. ГОСТ Р 55102—2012 Обращение с отходами: дата введения 2012. – . – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный.
37. Грошев, А. Г. Построение онтологических моделей систем автоматизированного проектирования / А. Г. Грошев, В. Н. Фролов, Е. Д. Федорков. — Текст : непосредственный // Информационные системы и технологии. — 2020. — № 4. — С. 1-11.

38. ЕСТЕСТВЕННОЕ И ИСКУССТВЕННОЕ ОСВЕЩЕНИЕ Daylighting and artificial lighting. – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный.
39. Ефименко, И. В. Онтологическое моделирование экономики предприятий и отраслей современной России / И. В. Ефименко, В. Ф. Хорошевский. — 1. — Москва : Изд. дом Высш. шк. экономики, 2011. — 150 с. — Текст : непосредственный.
40. Исследование методов автоматической генерации онтологии на основе структурированного материала. — Текст: электронный // technology : [сайт]. — URL: <https://technology.snauka.ru/2016/10/10842> (дата обращения: 15.06.2023).
41. Кравцов, Д. В. Автоматизированная система для построения онтологий предметных областей / Д. В. Кравцов, Д. А. Коростелёв, О. Н. Юркова. — Текст: непосредственный // Информационные системы и технологии в науке и образовании. — 2009. — № 3. — С. 1190-1199.
42. Кудряшев, А. Ф. Современная онтология: актуальные проблемы / А. Ф. Кудряшев. — Текст: непосредственный // Вестник Башкирского университета. — 2009. — № 3. — С. 1190-1199.
43. Левин Проблема универсалий. Современный взгляд : [монография] / Левин, Д. Г. — 1. — Москва : "Канон+", 2006. — 222 с. — Текст : непосредственный.
44. Львович, И. Я. Основы информатики / И. Я. Львович, Ю. П. Преображенский, В. В. Ермолова. — Текст: непосредственный // Журнал. — 2014. — № 2(34). — С. 335-339.
45. Маслова, О. В. Анализ методов генерации онтологических моделей по коллекции текстовых документов / О. В. Маслова, Л. А. Макушкина. — Текст: непосредственный // Вестник магистратуры. — 2014. — № 4. — С. 85-89.
46. О введении в действие санитарно-эпидемиологических правил и нормативов СанПиН 2.2.2/2.4.1340-03 : СанПИН // Постановление Главного государственного санитарного врача РФ от 3 июня 2003 г. N 118 "О введении в действие санитарно-эпидемиологических правил и нормативов СанПиН

- 2.2.2/2.4.1340-03" (с изменениями и дополнениями). – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный.
47. Онтология. — Текст : электронный // gtmarket : [сайт]. — URL: <https://gtmarket.ru/concepts/6847?ysclid=lj2hx4iic264113738> (дата обращения: 19.06.2023).
48. Основы концепции онтологического моделирования бизнес-процессов для задач принятия решений. — Текст: электронный // top-technologies : [сайт]. — URL: <https://top-technologies.ru/ru/article/view?id=38413> (дата обращения: 17.06.2023).
49. Павлов Семантика с единственным денотатом истина и многозначные интерпретации / Павлов, С. А.. — Текст: непосредственный // Логико-философские исследования. — 2010. — № 4. — С. 1-11.
50. Павлов, С. А. Анализ семантики, онтологии и синтаксиса логики высказываний / С. А. Павлов. — Текст : непосредственный // Вестник РУДН. — 2011. — № 4. — С. 110-118.
51. Палагин, А. В. Онтологические методы и средства обработки предметных знаний: монография / А. В. Палагин, С. Л. Кривый, Н. Г. Петренко. — Текст: непосредственный // изд-во ВНУ им. В. Даля. — 2012. — № 1. — С. 320-324.
52. Палей, Е. В. Онтология и теория познания / Е. В. Палей, М. Г. Зеленцова. — 1. — Иваново: хим.-технол. ун-т., 2018. — 110 с. — Текст: непосредственный.
53. Плотников, В. И. Онтология: хрестоматия / В. И. Плотников. — 1. — Екатеринбург : М, 2004. — 150 с. — Текст : непосредственный.
54. Попов, Д. В. Исследование методов построения конвертера онтологических моделей курса / Д. В. Попов, Л. А. Макушкина. — Текст: непосредственный // Современные научные исследования и инновации. — 2014. — № 1. — С. 1-3.
55. Представление данных и знаний в Интернете. — Текст : электронный // scask : [сайт]. — URL: [https://scask.ru/m\\_book\\_bki.php?id=77&ysclid=lj2jfn8wu041984052](https://scask.ru/m_book_bki.php?id=77&ysclid=lj2jfn8wu041984052) (дата обращения: 16.06.2023).

56. Преображенский, Ю. П. Построение информационной интеллектуальной системы данных / Ю. П. Преображенский. — Текст : непосредственный // Сборник научных статей 6-й Всероссийской научно-технической конференции с международным участием. — Москва : Прогрессивные технологии и процессы. , 2019. — С. 220-224.
57. Профилактика стрессового состояния работников при различных видах профессиональной деятельности. – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный.
58. Рассел, С. Искусственный интеллект: современный подход / С. Рассел, П. Норвиг. — Текст: непосредственный // Журнал. — 2006. — № 2. — С. 1300-1408.
59. Реализации системы интеллектуального поиска в области CALS-, CAD-, CAM-, CAE технологий . — Текст : электронный // pandia : [сайт]. — URL: <http://pandia.ru/text/77/373/27841.php> (дата обращения: 19.06.2023).
60. Региональный сектор экономики знаний: проблемы формирования и управления / М. Р. Ермоленко, Р. М. Закарян, Д. В. Закарян, А. П. Ланская. — Текст: непосредственный // Монография. — 2013. — № 4. — С. 220-300.
61. Рыбанов, А. А. Автоматизированное определение количественных характеристик текста / А. А. Рыбанов. — Текст: непосредственный // Современные научные исследования и инновации. — 2014. — № 2(34). — С. 1-5.
62. Савченко, А. П. Упрощённая методика построения многослойной онтологической модели предметной области / А. П. Савченко. — Текст : непосредственный // Научный журнал КубГАУ. — 2014. — № 100. — С. 21-31.
63. Савченко, А. П. Онтологический подход к представлению знаний в интеллектуальных системах: проблемы и перспективы практического применения / А. П. Савченко, — Текст: непосредственный // Актуальные проблемы управления корпорацией и человеческим капиталом в экономике знаний. — 2011. — № 3. — С. 370-379.

64. Санитарно-защитные зоны и санитарная классификация предприятий, сооружений и иных объектов: СанПИН // Проектирование, строительство, реконструкция и эксплуатация предприятий, планировка и застройка населенных мест. – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный.
65. Типовая инструкция по охране труда при работе на персональном компьютере. – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный.
66. Трудовой кодекс Российской Федерации: Федеральный закон от 30 декабря 2001 г. N 197-ФЗ (ТК РФ) // Собрание законодательств Российской Федерации. – Доступ из Справочно-правовой системы "Кодекс" . – Текст : электронный.
67. Чалая, Л. Э. Метод автоматического построения онтологических моделей с древовидно структурой концептов / Л. Э. Чалая. — Текст : непосредственный // Журнал. — 2015. — № 1. — С. 38-42..
68. Чевозерова Система онтологических принципов журналистики / Чевозерова, В. Г. — Текст: непосредственный // ВЕСТНИК ВГУ. — 2014. — № 1. — С. 212-233.
69. Что такое онтология, онтологические функции и категории, основные понятия и онтологический смысл. — Текст: электронный // rozli : [сайт]. — URL: <https://rozli.ru/istoriya/chto-takoe-ontologiya-ontologicheskie-funktsii-i-kategorii-osnovnye-ponyatiya-i-ontologicheskij-smysl.html> (дата обращения: 19.06.2023).
70. Шаповалов Основная проблема философской онтологии / Шаповалов, Ф. В. — Текст: непосредственный // Тезисы докладов и выступлений IV Российского философского конгресса. — Москва: Философия и будущее цивилизации, 2005. — С. в 5т.
71. Шебалов, Р. Современные российские разработки в области онтологического моделирования / Р. Шебалов, С. Горшков. — 1. — Москва: Тринидата, 2016. — 150 с. — Текст: непосредственный.

**Приложение А**

(обязательное)

**The program of automatic construction of an ontological model of the subject area based on the analysis of text corpora**

Студент

Группа	ФИО	Подпись	Дата
8ВМ1И	Михайлов Никита Игоревич		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Губин Е.И	к.ф.-м.н		

Консультант-лингвист отделения иностранных языков ШБИП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель (ОИЯ, ШБИП)	Ануфриева Т. Н			

## **Essentials of the term “Ontology”**

Nowadays process of communications between peoples and organizations plays a big role in humans’ life. The lack of understanding can become a result to such phenomenon like:

1. Poor degree of communication between some people and organizations.
2. Problems in understanding requirements and purposes of informational system.
3. Inventing things that have already been invented.

The structure of ontologies includes concepts and relationships between them.

Concepts are elements of abstract or concrete nature representing some object.

They also can be description of some task or action.

Ontology is mostly understood as a common view of some domain or part of knowledge which can solve problems denoted above.

Term ontology includes also a form of understanding, representing some field of knowledge. It is mostly viewed as a set of concepts, their definitions and relations. Such structures are called conceptualization.

Ontology can have forms, but it always has set of terms and their meanings.

Ontology can be viewed in form of four degree of formality:

1. High informal – almost not expressed in natural language.
2. Some informal – represents restricted form of human language.
3. Semi-formal – presented in artificial form of human language.
4. Rigorous formal- terms are presented with formal semantics and proofs.

Ontologies are used in many domains of human activity. One of important applications of ontology is usage as a framework of integration to human natural language. As a whole, ontology can be used in three main spaces of human activity:

1. Communication.
2. Inter-operability.
3. Engineering of informational systems.

*Communication*

Ontology allows to eliminate semantical confusion by supplying with standardized framework for building informational model of knowledge system. But there are some difficulties because of lack of common understanding of essence of relationships informational system. Ontologies are supposed to discover the nature of relationships by the way of identifying logical connections.

One of important roles of ontology in communication is providing implicit definitions for terms in the field of software system. Ontology is up to supply with environment displaying various meanings for different people.

Having a system with many communication participants it is of great importance to implement various perspectives by taking distinctions of each perspective. Ontology by providing standardized model allows to achieve integration by helping participants in their communications.

#### *Inter-operability*

The issue of inter-operability is marked by many applications is important too. It represents the need of exchanging data between users of different software tools. Ontology assists in solving problems of such kind by integrating one environment for different toolkits. Although there are many software tools able to solve problems of such kinds there are no consistency among them.

Any enterprise model should be integrated into the environment of information technology. These models are repositories for various tool sets.

Ontologies can also be implemented to provide translation between different languages. There are several methods to achieve these goal, one of them is to provide unique translator for two party exchange, which requires  $O(n^2)$  translators for  $n$  ontologies. Ontologies applied as mean of internal translation can decrease the whole number of translators for  $n$  different ontologies, thus it will only need translator from native ontology inter interchange one.

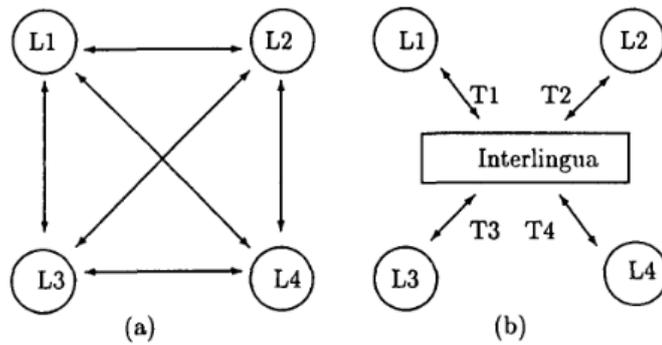


Figure 1—Shema of decreasing the number of translators

There are several methods for defining of relationships. First it is important to distinct the nature of relationships among users. Ontologies used in different organizations must be usable in different informational environments.

*Engineering of informational systems*

Ontologies can also play a key role in defining the specification of software systems. For example, the IBM Business System Development Method (IBSDM) applies ontology as a mean for organizing the basis of IT design and its development.

The role of ontologies in specifications depends on the degree of formality in system design methodology.

From the point of view of formal approach, ontology is able to assist in making the process of defining the system’s requirements more easy.

As for informal approach ontology allows to have declarative specification of software system and clarifying the ways the system supports its functionality.

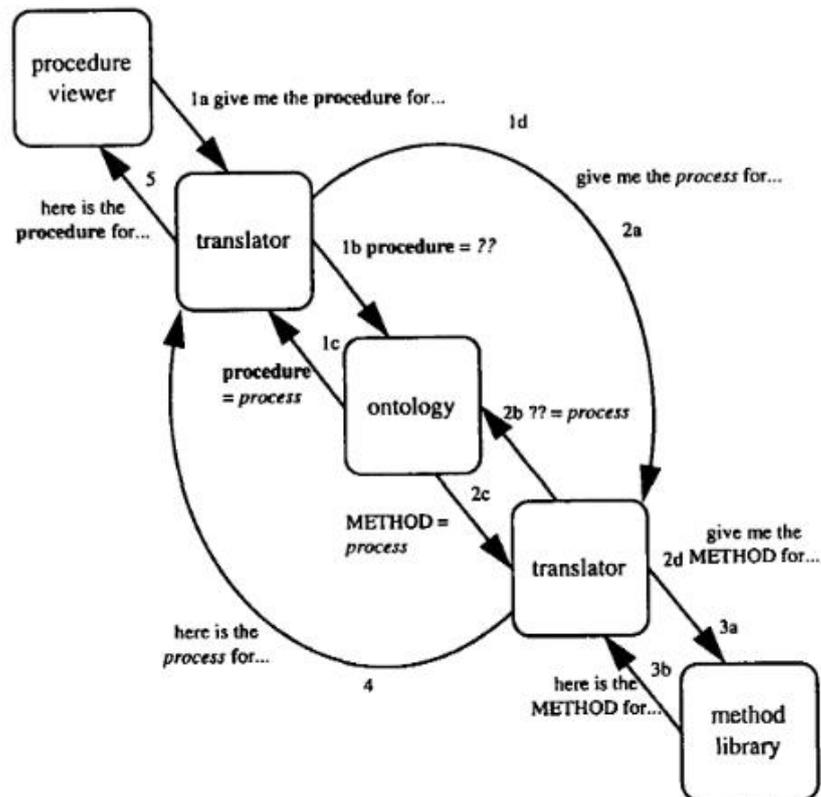


Figure 2–Schema of usage of ontology in integrating different software tools

For more effectiveness ontologies are up to be reusable, allowing an importing and exporting different modules. Thus, ontologies provide with libraries for using classes in process of modelling and displaying problems. One example of such library is Knowledge Systems Laboratory which has its own ontology server. Ontologies should also support the extending of new classes.

### Building ontology

Nowadays despite the great amount of experience on developing and using ontologies, there is a lack in standard methods of building ontologies. One of proposed of building ontology methodologies includes the following steps:

1. Capture ontologies.
2. Code ontologies.
3. Integrate existing ontologies.

#### *Capture ontologies*

Ontologies capture means the identifying of the main concepts and relationships between them, finding unambiguous definitions for them and agreeing all above.

Ontology capturing process comprises four main phases: Scoping, produce definitions, review and develop of meta ontologies.

### *Scoping*

To produce all relevant terms and definitions representing concept it is important to have some kind of brainstorming.

The next step of scoping is to categorize and group captured terms by their similarity. And define semantic cross-references between knowledges fields.

### *Produce definitions*

This process includes allowing concepts and their interrelations to define the requirements for meta-ontology.

To achieve the better control of the level of each concept detail it is important to keep a balance in terms of each level detail. A middle approach allows to reach a best balance in the way of specifying only the basic concepts. Here the most important concepts tend to be more stable.

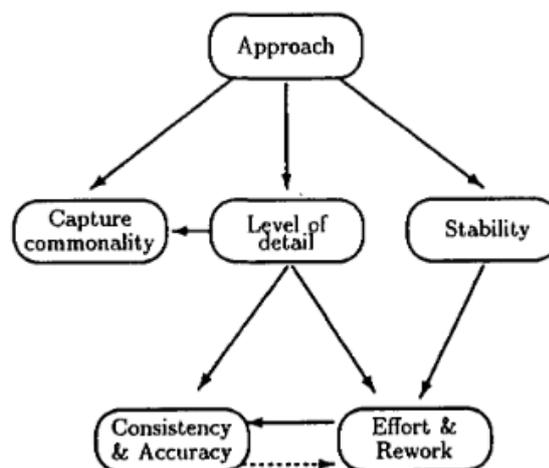


Figure 3 – Middle-level approach

### *Ontology coding*

Coding means clear representation of captured concepts in form of some standardized language. This phase comprises: defining the basic terms which will

be used in ontologies, choosing the language in which ontologies will be represented, write the code.

### *Integrating the existing ontologies*

The question of methods of integrating ontologies that already exist is not less important. This problem is managed to solve by Skuce. Its approach was to clarify all assumptions underlying the ontology.

Another approach comprises identifying synonyms and extending the ontologies in such area of knowledge where there are no concepts.

### **Ontology languages**

There are many ontology languages have been created during the years of ontology existence.

Two groups of ontology languages exist:

1. Languages based on XML (such as OML etc.).
2. Languages created by Wide Web Consortium (W3C) working groups..

#### *Simple HTML Ontology Extension (SHOE)*

Was developed at University of Maryland and is supposed to be an extension of HTML. Later it was adapted to XML.

SHOE allows to gather important information about Web pages improving search mechanisms and methods of knowledge gathering.

The gathering information has three phases:

1. Defining an ontology.
2. Providing HTML page with descriptive information.
3. Retrieving information by searching existing pages and making having updated information.

#### *Ontology Markup Language (OML)*

Was created at the University of Washington. OML is based on SHOE, because of being firstly considered to be a form of XML serialization of SHOE.

There are four levels of OML:

1. OML Core – relates to logical features of language and is included in the rest of another parts.
2. Simple OML – represents RDF.
3. Abbreviated OML – comprises conceptual graphs.
4. Standard OML – most expressive version of OML.

### *Ontology Interchange Language (OIL)*

Developed as a part of Onto Knowledge Project allows semantic combability between Web resources. It can model conceptual primitives (concepts, taxonomies, relations)

OIL comprises the layers described below:

1. Standard OIL – complete OIL model with all primitives.
2. Instance OIL – able to append instances of concept to previous model.

OIL can be presented not only in form of XML, but also in form of ASCII

### *Resource Description Framework and RDF Schema*

RDF allows to describe Web resources and specify semantic data in XML format by standardizing them. It also has means for displaying services, processes and business models by recognizing implicit information

RDF consists of three types:

1. Resources – described by RDF and named by URI plus ID.
2. Properties – represent aspects, features and relations for describing resources.
3. Statements – display a value of property.

RDF has no mechanism to specify relationships between properties and resources. This lies in responsibility of RDFS. RDFS provides primitives to find out knowledge models close to frame approaches.

To conclude, ontology represents a bunch of classes and relationships between them. It is of great importance to have some standardized representing method to display them in explicit manner to all systems of different kind. Ontologies plays a big role in humans' activity: they allow to represent information and structure of

some knowledge system in clear and obvious way, specify the design of IT systems and assist in interpreting information in more effective way.

Although there is a great amount of methods of specifying and representing ontology, they remain the important part of researches.