

**Министерство образования и науки Российской Федерации**  
федеральное государственное автономное образовательное учреждение  
высшего образования  
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

---

Школа ИШИТР

Направление подготовки 09.03.04 Программная инженерия

Отделение информационных технологий

**БАКАЛАВРСКАЯ РАБОТА**

Тема работы
<b>Разработка алгоритмического и программного обеспечения модуля автоматизированного извлечения контента веб-ресурса</b>

УДК 004.415.2:004.774:316.653

Студент

Группа	ФИО	Подпись	Дата
8К4Б	Евграфов Денис Андреевич		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев Алексей Олегович	К. Т. Н.		

**КОНСУЛЬТАНТЫ:**

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Петухов Олег Николаевич	К. Э. Н., доцент		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ОКД ИШКНБ	Авдеева Ирина Ивановна			

**ДОПУСТИТЬ К ЗАЩИТЕ:**

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Чердынцев Евгений Сергеевич	К. Т. Н., доцент		

Томск – 2018 г.

**Министерство образования и науки Российской Федерации**  
федеральное государственное автономное образовательное учреждение  
высшего образования  
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

---

Школа ИШИТР  
Направление подготовки 09.03.04 Программная инженерия  
Отделение информационных технологий

УТВЕРЖДАЮ:  
Руководитель ООП

\_\_\_\_\_  
(Подпись)      (Дата)      (Ф.И.О.)

**ЗАДАНИЕ**  
**на выполнение выпускной квалификационной работы**

В форме:

бакалаврской работы
---------------------

Студенту:

Группа	ФИО
8К4Б	Евграфов Денис Андреевич

Тема работы:

Разработка алгоритмического и программного обеспечения модуля автоматизированного извлечения веб-ресурса	
Утверждена приказом директора (дата, номер)	

Срок сдачи студентом выполненной работы:	
--	--

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ:**

<p><b>Исходные данные к работе</b></p> <p><i>(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.).</i></p>	<p>– литературные источники.</p>
---	----------------------------------

<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>	<ul style="list-style-type: none"> <li>– аналитический обзор существующих методов извлечения веб-контента;</li> <li>– исследование методов синтаксического анализа;</li> <li>– обзор имеющихся решений на рынке;</li> <li>– реализация алгоритма извлечения контента с новостных сайтов</li> <li>– реализация алгоритма извлечения контента из социальной сети twitter.</li> <li>– реализация алгоритма предварительной обработки контента</li> </ul>
--	---

<p><b>Перечень графического материала</b></p> <p><i>(с точным указанием обязательных чертежей)</i></p>	
--	--

**Консультанты по разделам выпускной квалификационной работы**

Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Доцент ОСГН ШБИП Петухов Олег Николаевич
Социальная ответственность	Ассистент ОКД ИШКНБ Авдеева Ирина Ивановна

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

--

<p><b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b></p>	
--	--

**Задание выдал руководитель:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев Алексей Олегович	к. т. н., доцент		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8К4Б	Евграфов Денис Андреевич		

**Планируемые результаты обучения по направлению 09.03.04  
«Программная инженерия»**

Код результатов	Результат обучения (выпускник должен быть готов)	Требования ФГОС, критерии АИОР
P1	Применять базовые и специальные естественнонаучные и математические знания в области информатики и вычислительной техники, достаточные для комплексной инженерной деятельности.	Требования ФГОС (ОК-1, 10, ПК-4, 5, 6), критерий 5 АИОР (п. 1.1)
P2	Применять базовые и специальные знания в области современных информационных технологий для решения инженерных задач.	Требования ФГОС (ОК-11, 12, 13, ПК-1, 2, 11), критерий 5 АИОР (п.1.1, 1.2)
P3	Ставить и решать задачи комплексного анализа, связанные с созданием аппаратно-программных средств информационных и автоматизированных систем, с использованием базовых и специальных знаний, современных аналитических методов и моделей.	Требования ФГОС (ОК-1, 8, ПК-2, 4, 6), критерий 5 АИОР (п. 1.2)
P4	Разрабатывать программные и аппаратные средства (системы, устройства, блоки, программы, базы данных и т. п.) в соответствии с техническим заданием и с использованием средств автоматизации проектирования.	Требования ФГОС (ОК-2, 3, ПК-3, 4, 5), критерий 5 АИОР (п. 1.3)
P5	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретация полученных данных, в области создания аппаратных и программных средств информационных и автоматизированных систем.	Требования ФГОС (ОК-6, ПК-6, 7), критерий 5 АИОР (п.1.4)
P6	Внедрять, эксплуатировать и обслуживать современные программно-аппаратные	Требования ФГОС (ОК-4,

	комплексы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды.	15, 16, ПК-9, 10, 11), критерий 5 АИОР (п. 1.5)
P7	Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности.	Требования ФГОС (ОК-1, 4, ПК-1, 6, 7), критерий 5 АИОР (п. 2.1)
P8	Владеть иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты комплексной инженерной деятельности.	Требования ФГОС (ОК-14, ПК-7), критерий 5 АИОР (п. 2.2)
P9	Эффективно работать индивидуально и в качестве члена группы, состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре организации.	Требования ФГОС (ОК-2, 3, 4), критерий 5 АИОР (п. 2.3, 2.4)
P10	Демонстрировать знания правовых, социальных, экономических и культурных аспектов комплексной инженерной деятельности.	Требования ФГОС (ОК-1, 5, 9), критерий 5 АИОР (п. 2.5)
P11	Демонстрировать способность к самостоятельному обучению в течение всей жизни и непрерывному самосовершенствованию в инженерной профессии.	Требования ФГОС (ОК-6, 7), критерий 5 АИОР (п. 2.6)

**Министерство образования и науки Российской Федерации**  
федеральное государственное автономное образовательное учреждение  
высшего образования  
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

---

Школа ИШИТР

Направление подготовки 09.03.04 Программная инженерия

Уровень образования бакалавриат

Отделение информационных технологий

Период выполнения осенний / весенний семестр 2017/2018 учебного года

Форма представления работы:

бакалаврская работа

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

Срок сдачи студентом выполненной работы:

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
15.05.2018 г.	Основная часть	60
20.05.2018	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	20
25.05.2018	Социальная ответственность	20

Составил преподаватель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Савельев Алексей Олегович	к. т. н.		

**СОГЛАСОВАНО:**

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Чердынцев Евгений Сергеевич	к. т. н., доцент		

## Реферат

Выпускная квалификационная работа: 74 с., 8 рис., 26 табл., 20 источников, 3 приложения.

Ключевые слова: извлечение контента веб-ресурсов, синтаксический анализ.

Объектом является система исследования общественного мнения.

Цель работы – разработка модуля, программного обеспечения, предназначенного для извлечения контента с веб-ресурса.

В процессе прохождения практики: произведена разработка модулей системы, модуль извлечения контента веб-ресурсов и модуль предварительной обработки контента.

В результате прохождения практики: определены дальнейшие действия по совершенствованию модуля извлечения контента.

Область применения: данная система может применяться фирмами которым необходимо производить исследование общественного мнения.

Экономическая эффективность работы в высокой научной значимости и актуальности поставленной задачи.

В будущем планируется проводить исследование и улучшение алгоритмов на основе реализованной системы анализа общественного мнения.

## Определения

В данной работе используются следующие термины соответствующими определениями:

**API (программный интерфейс приложения)** - набор готовых классов, процедур, функций, структур и констант, предоставляемых приложением или операционной системой для использования во внешних программных продуктах;

**Атрибут** – переменная, связанная с классом или объектом;

**URL** – единый указатель ресурса;

**RSS** – обогащенная сводка сайта;

**XML** – расширяемый язык разметки;

**Web Content Mining (извлечение контента веб-ресурсов)**– это использование методов интеллектуального анализа данных для автоматического обнаружения веб-документов и услуг, извлечения информации из веб-ресурсов и выявления общих закономерностей в Интернете;

**Контент** – содержимое, информационное наполнение;

**DFD** – диаграммы потоков данных.

**DOM** – объектная модель документа

## Нормативные ссылки

В данной работе используются ссылки на следующие стандарты:

1. ГОСТ 12.0.003-74. ССБТ. Опасные и вредные производственные факторы. Классификация.
2. ГОСТ 12.1.003–83 ССБТ. Шум. Общие требования безопасности.
3. ГОСТ 12.1.045–84 ССБТ. Электростатические поля. Допустимые

- уровни на рабочих местах и требования к проведению контроля.
4. ГОСТ 12.1.038–82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов.
  5. ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования.
  6. ГОСТ 12.1.010-76 ССБТ. Взрывобезопасность. Общие требования.
  7. ГОСТ 19.402-78. ЕСПД. Описание программы.

## Оглавление

### Оглавление

Реферат.....	7
Определения .....	8
Нормативные ссылки.....	8
Оглавление .....	10
Введение .....	13
1.1 Требования к модулю извлечения контента .....	14
2. Обзор предметной области .....	16
2.1 Описание различных подходов web mining .....	15
2.1.1 Анализ DOM дерева.....	15
2.1.2 Использование регулярных вырожений .....	16
2.1.3 XML-парсинг .....	17
2.2 Выводы.....	18
3 Описание алгоритмов модуля извлечения контента.....	19
3.1 Описание источников первичной информации .....	19
3.2 Алгоритмизация метода XML-парсинга .....	20
3.3 Предварительная обработка данных и выгрузка в базу данных.....	23
3.4 Выводы.....	24
4 Программная реализация.....	26
4.1 Структура модуля .....	26
4.2 Пользовательский интерфейс .....	27
4.3 Тестирование.....	28
Общий вывод по разделу .....	32
5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение.....	33
5.2 SWOT-анализ.....	34

<b>5.3</b>	<b>Определение возможных альтернатив проведения научных исследований</b>	<b>36</b>
5.3.1	Структура работ в рамках научного исследования	36
5.3.2	Определение трудоемкости выполнения работ	37
5.3.3	Разработка графика проведения научного исследования	41
5.3.4	Бюджет научно-технического исследования (НТИ)	45
5.3.4.1	Расчет материальных затрат НТИ	45
5.3.4.2	Расчет затрат на специальное оборудование для научных (экспериментальных) работ	47
5.3.4.3	Основная заработная плата исполнителей темы	47
5.3.4.3	Дополнительная заработная плата исполнителей темы	49
5.3.4.5	Отчисления во внебюджетные фонды (страховые отчисления)	49
5.3.4.6	Расчет затрат на научные и производственные командировки	51
5.3.4.7	Контрагентные расходы	51
5.3.4.8	Накладные расходы	51
5.3.4.9	Формирование бюджета затрат научно- исследовательского проекта	52
5.3.5	Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	52
	Общий вывод по разделу:	57
6.1	Производственная безопасность	58
6.1.1	Параметры микроклимата	59
6.1.2	Освещенность рабочего места	60
6.1.3	Уровень шума	62
6.1.4	Умственное перенапряжение	63
6.1.5	Опасность поражения электрическим током	64
6.2	Экологическая безопасность	66
6.3	Безопасность в чрезвычайных ситуациях	67
6.4	Правовые и организационные вопросы обеспечения безопасности	68
	Заключение	70
	Список используемой литературы	71
	Приложение А	74

<b>Предложение Б.....</b>	<b>76</b>
<b>Приложение В .....</b>	<b>77</b>

## Введение

В современном мире довольно сложно представить организацию, предприятие или фирму, которая не заинтересована в изучении общественного мнения и его направленности, так как оно подвержено изменениям, на которые можно повлиять. В большинстве случаев можно пользоваться данными опросов общественного мнения, постоянно публикующимися в средствах массовой информации, а в случае необходимости получить данную информацию от коммерческих организаций, проводящих социологические исследования, но помимо этого, работникам необходимо иметь квалифицированных специалистов в области изучения общественного мнения. Знание основ проведения опросов населения, аналитического изучения настроения людей, умение интерпретировать разного рода статистические данные.

*Целью* данной работы является разработка программного модуля, обеспечения предназначенного для выполнения функции сбора данных с веб-ресурсов и их предварительной обработки.

*Объектом* исследования является система исследования общественного мнения.

*Предметом* исследования являются системы извлечения контента из веб-ресурсов и предварительной обработки данных.

С практической точки зрения, анализ общественного мнения, является только частью возможного функционала данного программного обеспечения, поскольку данная система может реализовать методы поиска информации в сети интернет, выявления знаний веб-ресурсов, персонализации информации и поиска шаблонов в поведении пользователей. Следовательно, данное программное обеспечение может применяться как в фирмах, которым необходимо анализировать мнение людей по отношению к произошедшим событиям, так и в фирмах занимающимся анализом поведения пользователей на веб-ресурсах.

## 1. Постановка задачи

Разработка алгоритмического и программного обеспечения модуля автоматизированного извлечения контента, предназначенного для сбора данных с новостных порталов и социальных сетей.

Модель работы модуля представлена в виде диаграммы IDEF0 (Рисунок 1.1).



Рисунок 1.1 – Постановка задачи

### 1.1 Требования к модулю извлечения контента

До того, как приступить к выбору методов и подходов, на основе которых будет происходить разработка, необходимо рассмотреть требования, которые предъявляются данному программному продукту:

- разработка должна вестись на языке разработки Java;
- работа с новостными сайтами, через RSS – запросы;
- работа с социальной сетью Twitter, через функционал API Twitter4j;
- модуль должен поддерживать работу с XML файлами;

- работа со встроенным функционалом веб ресурсов для извлечения контента;
- обновление данных с новостных ресурсов до актуального состояния;
- составление общего списка новостей на основе данных с нескольких ресурсов;
- получение списка пользователей входящих в целевую выборку;
- извлечение постов людей входящих в целевую выборку;
- произведение предварительной обработки текста;
- преобразование данных в формат необходимый для их дальнейшей обработки;
- выгрузка результата работы модуля в базу данных

## **2. Обзор предметной области**

В настоящее время все пользователи веб-услуг пользуются большим количеством информации, взятой из интернета, отсюда и появилась необходимость в автоматическом анализе информации. С развитием всемирной паутины в глобальную информационную структуру, обычные пользователи стали как потребителями, создателями так и распространителями информации. Вследствие чего новое направление в методологии анализа данных, предназначенное для эффективного структурирования задач поиска, решения и анализа, беспорядочно организованной информации в сети.

Web content mining представляет из себя процесс извлечения данных из веб-ресурсов, имеющий как правило более практическую составляющую нежели теоретическую. Основной целью web content mining является извлечение знаний контента документов в сети интернет.

### **2.1 Описание различных подходов web mining**

Фактически, задача web content mining сводится к созданию HTML парсеров. В данный момент существует несколько подходов к извлечению данных:

- Анализ DOM дерева.
- Использование регулярных выражений.
- XML парсинг.
- Визуальный подход.

#### **2.1.1 Анализ DOM дерева**

Данный метод основан на технологии DOM-парсинг, данная технология действует, опираясь на вложения веб-браузера. К примеру такие программы

могут получать динамическое содержимое, сгенерированное скриптом на стороне клиента. Данные программы парсят веб-страницы в DOM-дерево.

Используя данный метод, данные получаются на прямую по идентификатору, имени или другим атрибутам элементов дерева. Также если элемент не обозначен идентификатором, то к нему можно получить доступ по уникальному пути, спускаясь по дереву вниз.

*body -> p[20] -> b[1] -> текст ссылки*

Или пройтись по коллекции однотипных элементов.

*body -> links -> 5 элемент -> текст ссылки*

Достоинства данного метода:

- возможность получать данные любого типа и любого уровня сложности.
- возможность получать значение элемента прописав путь к его расположению на прямую.

Недостатки данного метода:

- необходимость привязываться к определенному HTML/Javascript движку, поскольку различные движки по-разному генерируют DOM-деревья.
- необходимость переделывать парсер при изменении пути элемента.
- сложность и не всегда однозначность DOM-пути.

### **2.1.2 Использование регулярных выражений**

Данный метод анализа основан на использовании регулярных выражений. Регулярные выражения представляют из себя формальный язык

для поиска и манипуляциями с подстроками текста, данный метод основан на использовании метасимволов. Поиск же осуществляется с помощью строки-образца, состоящий из символов и метасимволов, задающих правила поиска.

Данный метод необходимо использовать только в тех случаях, когда имеется строгий формат извлекаемых данных. Поскольку при малейшей разнице в форматах извлекаемых данных данный метод будет выдавать некорректные данные. Обусловлено это тем, что регулярные не предназначены для разбора тегов.

### **2.1.3 XML парсинг**

Данный метод анализа основан на работе с одним из самых распространённых форматов, в которых хранятся данные в сети. XML является рекомендованным консорциумом W3C языком разметки. В свою очередь язык предназначен для хранения структурированных данных. Также преимуществом XML является его нативная поддержка во всех основных операционных системах и языках программирования.

Суть работы данного метода заключается в извлечение данных с веб-ресурсов в формате XML с последующей обработкой с помощью стандартного функционала используемой среды разработки.

Преимуществом данного метода является современная распространённость формата XML для хранения информации в сети интернет. Также неоспоримым преимуществом является, то что именно в данном формате приходят данные в ответ на RSS-запросы, которые являются распространённым способом извлечения контента из веб-ресурсов.

Особенности XML:

- удобен для понимания машиной, но неудобный для понимания человеком.
- XML может быть использован в любых типах приложения.

- наличие инструментов для создания, редактирования, разбора XML файлов присутствует для всех операционных систем и всех актуальных языков программирования.
- схемы XML файлы можно проверять с помощью XSD схем, что предотвращает попадание файла в неправильном формате.

## **2.2 Выводы**

Рассмотренные подходы Web content mining имеют свои недостатки и достоинства. Выбор метода, на котором будет основана система извлечения контента, был сделан на основе требований, выявленных на этапе постановки задачи. Поскольку на веб-ресурсах выбранных в качестве источников данных реализована поддержка RSS-запросов, то наш выбор пал на метод парсинга XML файлов, поскольку результатом RSS-запроса является XML файл. Помимо этого, формат XML имеет простой синтаксис удобный для понимания машиной.

### 3. Описание алгоритмов модуля извлечения контента

#### 3.1. Описание источников первичной информации

Основными объектами системы извлечения контента из веб-ресурсов являются посты на новостных сайтах и сообщения людей в социальной сети Twitter. Для извлечения контента нам необходимы RSS-запросы для извлечения данных с новостных сайтов и Twitter4j для извлечения постов людей, входящих в целевую выборку.

Рассмотрим эти методы и результаты их работы. RSS-запросы представляют из себя URL-запрос к сайту (Рисунок 3.1), ответом на данный запрос является XML-файл (Рисунок 3.2).

<https://lenta.ru/rss/news>

Рисунок 3.1 – RSS-запрос извлекающий все новости за последние трое суток

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0" xmlns:atom="http://www.w3.org/2005/Atom">
  <channel>
    <language>ru</language>
    <title>Lenta.ru : Новости</title>
    <description>Новости, статьи, фотографии, видео. Семь дней в неделю, 24 часа в
сутки</description>
    <link>https://lenta.ru</link>
    <image>
      <url>https://assets.lenta.ru/small_logo.png</url>
      <title>Lenta.ru</title>
      <link>https://lenta.ru</link>
      <width>134</width>
      <height>22</height>
    </image>
    <atom:link rel="self" type="application/rss+xml" href="http://lenta.ru/rss/news"/>
  <item>
    <guid>https://lenta.ru/news/2018/07/10/kartofan/</guid>
    <title>Россельхознадзор рассказал о праве россиян сажать любую картошку</title>
    <link>https://lenta.ru/news/2018/07/10/kartofan/</link>
    <description>
      <![CDATA[Россельхознадзор опроверг слухи о том, что за выращивание картофеля из своих семян
взимается административный штраф. В службе пояснили, что ответственность за продажу посадочного
материала ненадлежащего качества и неизвестного происхождения предусмотрена только в отношении
лиц, ведущих коммерческую деятельность.]]>
    </description>
    <pubDate>Tue, 10 Jul 2018 06:13:00 +0300</pubDate>
    <enclosure
url="https://icdn.lenta.ru/images/2018/07/10/06/20180710061349251/pic_c534806cc7a78844af6fc43d114d
5ffe.jpg" length="93077" type="image/jpeg"/>
    <category>Россия</category>
  </item>
</rss>
```

Рисунок 3.2 – XML-файл, являющийся результатом работы RSS-запроса

Twitter4j является фреймворком для работы с социальной сетью Twitter с помощью Java. Ниже приведён пример метода по извлечению списка постов пользователя.

```
Twitter twitter = new TwitterFactory().getInstance();
```

```

try {
    List<Status> statuses;
    String user;
    if (args.length == 1) {
        user = args[0];
        statuses = twitter.getUserTimeline(user);
    } else {
        user = twitter.verifyCredentials().getScreenName();
        statuses = twitter.getUserTimeline();
    }
} catch (TwitterException te) {
    te.printStackTrace();
    System.out.println("Failed to get timeline: " + te.getMessage());
    System.exit(-1);
}

```

Результатом работы данного метода является массив объектов класса Status в котором находятся 20 последних постов пользователя за прошедшие сутки.

Для решения поставленной задачи, а именно извлечение контента с целью его последующего хранения и обработки нам необходимо произвести предварительную обработку полученных данных

### **3.2 Алгоритмизация метода XML-парсинга**

Для реализации модуля извлечения контента из веб-ресурсов был выбран метод XML-парсинг. Рассмотрим алгоритм работы метода путем декомпозиции диаграммы IDEF0 из раздела 1 (рисунок 3.2)

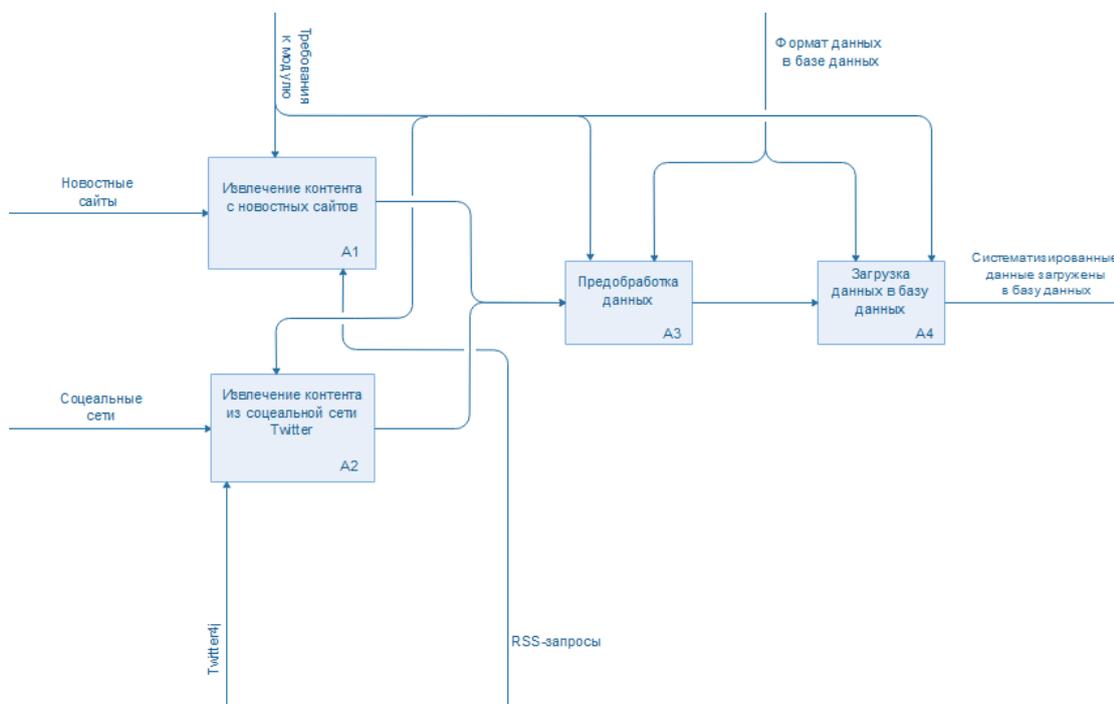


Рисунок 3.2 – Декомпозиция поставленной задачи

Как мы видим, алгоритм состоит из четырёх основных этапов, на выходе из которого мы должны получить систематизированные данные, загруженные в базу данных. На вход алгоритм получает адреса новостных сайтов и социальной сети. На первом этапе модуль производит извлечение контента из веб ресурсов, а именно в первую очередь получает список пользователей, попавших в целевую выборку, после чего производит извлечение новостных постов и постов пользователей. После чего в процессе “Предобработке данных” происходит синтаксический анализ текста следствием которого является текст из которого исключены все символы, не влияющие на морфологическое значение текста.

Рассмотрим декомпозицию процесса “Извлечение контента с новостных сайтов” (Рисунок 3.3)

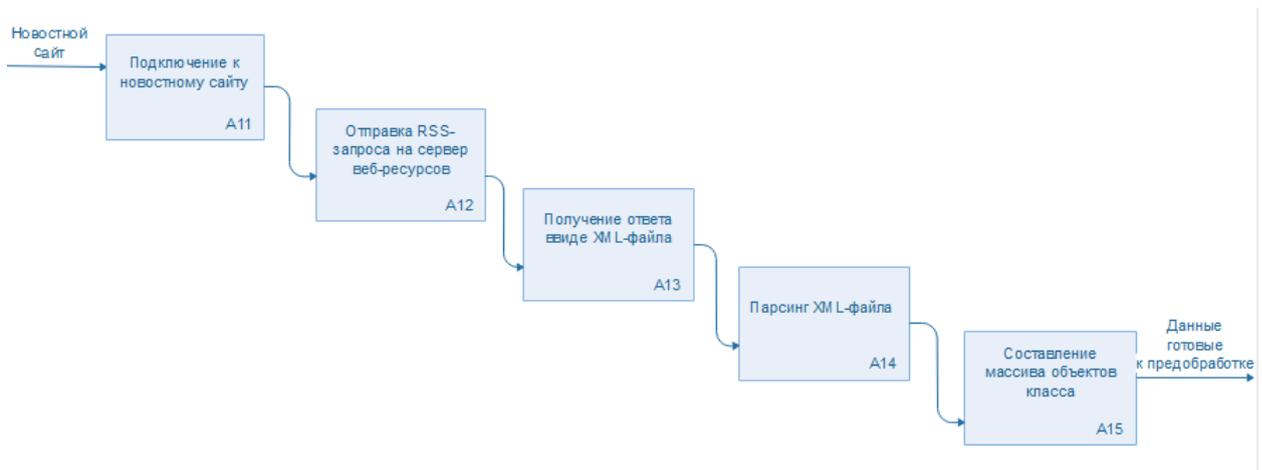


Рисунок 3.3 – Декомпозиция процесса А3

На рисунке выше изображена последовательность работы метода извлечения контента с новостных сайтов. На вход подаётся адрес новостного сайта, после чего происходит подключение к сайту с целью отправки RSS-запроса на сервер новостного ресурса. После получения ответа на запрос в виде XML-файла происходит его парсинг и результатам парсинга заполняются поля объектов класса News. Последним действием выполняемым данным модулем является составление массива объектов класса News с целью его передачи в модуль предварительной обработки.

### 3.3 Предварительная обработка данных и выгрузка в базу данных

Результатом работы модулей изъятия контента являются XML файлы. Для дальнейшей работы с полученными данными необходимо произвести синтаксический анализ для выделения необходимой информации. Анализ производится с помощью стандартного фреймворка XMLDOMImplementation. После выделения необходимая информация из XML файла выгружается в базу данных. Общение с базой данных происходит с помощью JDBC драйверов. Данный модуль задействует несколько таблиц базы данных Oracle. Рассмотрим эти таблицы (Рисунок 3.4)

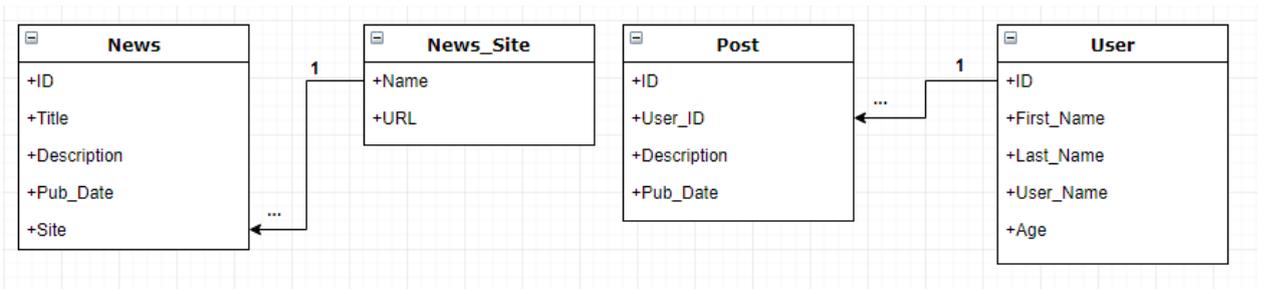


Рисунок 3.4 – Модель используемых таблиц базы данных

На рисунке выше изображена частичная модель базы данных. По ней мы видим разработанный модуль использует четыре таблицы базы данных, в этих таблицах хранится как информация о пользователях и новостных сайтах, также в данные таблицы происходит выгрузка данных прошедших предварительную обработку.

Рассмотрим схему модели потоков данных (Рисунок 3.5)

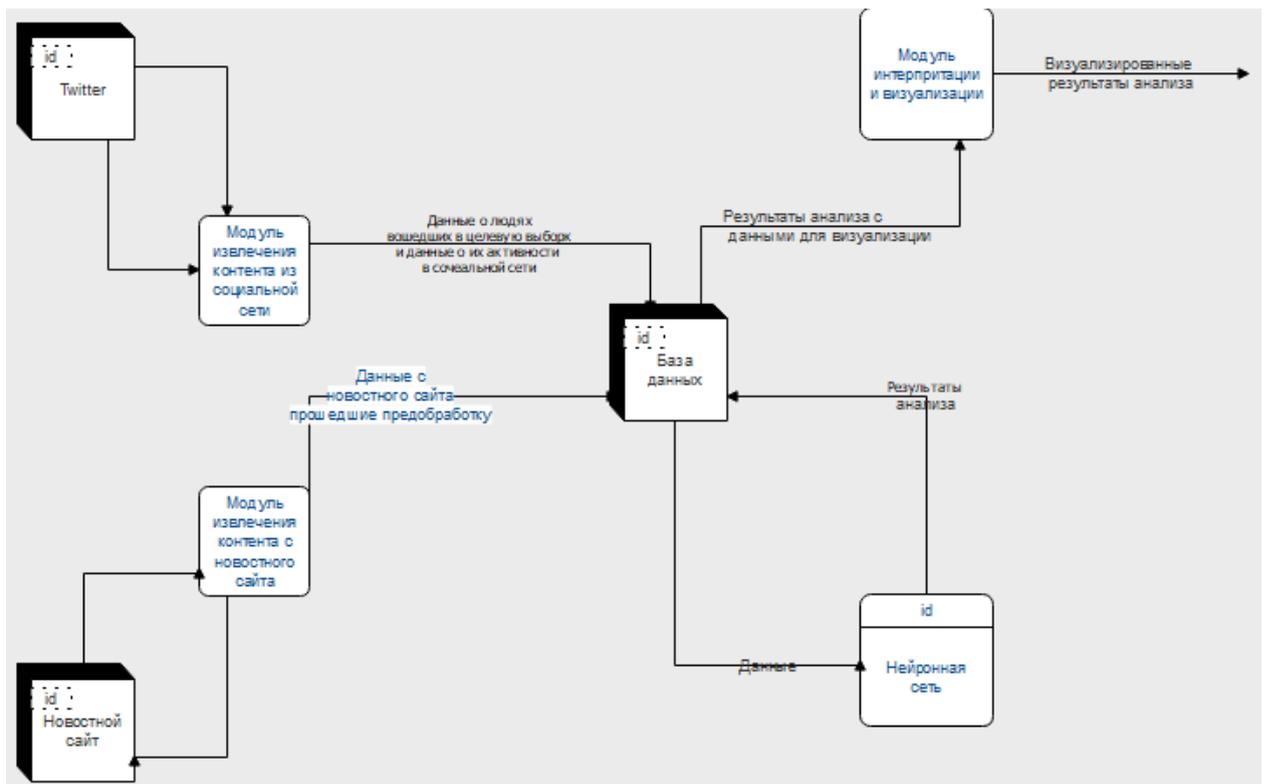


Рисунок 3.5 – Модель потоков данных

Выше приведенная модель отображает общую модель потоков данных разрабатываемого продукта, с декомпозицией модуля извлечения контента из веб-ресурса. Из декомпозиции мы видим, что модули извлечения контента с новостного сайта и социальной сети производят обмен данными, далее модули после обработки полученных данных производят выгрузку полученных

данных в базу данных.

### **3.4 Выводы**

Выше были рассмотрены источники первичной информации, алгоритмы извлечения и предварительной обработки контента, модель потоков данных и частичную модель базы данных. Для описания приведенных алгоритмов были использованы UML диаграммы IDEF0 и её декомпозиции, DFD. Результатом работы над данным подразделом являются алгоритмы на которых основывается работа данного модуля.

## 4 Программная реализация

### 4.1 Структура модуля

Для реализации данного модуля была разработана структура модуля, состоящая из пяти самостоятельно написанных классов, используемых в четырёх фреймворках .

Ниже приведена модель диаграммы классов, написанных самостоятельно (Рисунок 4.1)

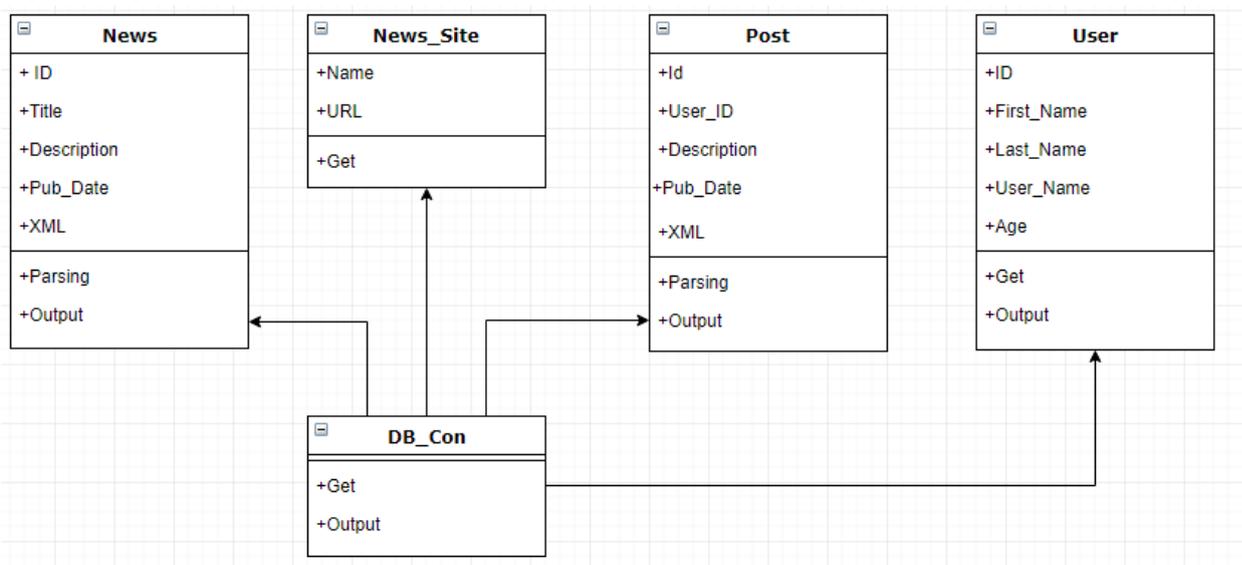


Рисунок 4.1 – Диаграмма классов

На диаграмме находятся пять классов, участвующих в работе модуля, ниже приведено описание этих классов:

1. Класс News, в данном классе хранятся статьи с новостных сайтов в формате XML и результат их предварительного анализа, также в данном классе имеются два метода. Метод Parsing реализовывает предварительный анализ статьи и заполняет остальные поля объекта класса. Метод Output производит выгрузку данных в базу данных, данный метод является наследуемым от интерфейса DB\_Con.
2. Класс Post, в данном классе хранятся посты пользователей социальной сети Twitter, входящих в целевую выборку, в формате XML и результат их предварительного анализа, также в данном классе имеются два метода. Метод Parsing реализовывает

предварительный анализ постов и заполняет остальные поля объекта класса. Метод Output производит выгрузку данных в базу данных, данный метод является наследуемым от интерфейса DB\_Con.

3. Класс News\_Site, в данном классе хранятся новостные сайты которые выбраны в качестве источников новостных статей. В данном классе имеются поля в которых хранятся название новостного сайта и его URL адрес. В данном классе имеется один метод Get, который наследуется от интерфейса DB\_Con и производит извлечение контента с новостных сайтов.
4. Класс Users, в данном классе хранится целевая выборка пользователей социальной сети Twitter. В данном классе имеются несколько методов Get реализующий извлечение данных о пользователях из социальной сети, данный метод является переопределённым методом из интерфейса DB\_Con, и метод Output реализующий выгрузку полученных данных в базу данных, данный метод является наследуемым от интерфейса DB\_Con.
5. Класс DB\_Con является интерфейсом в котором реализованы два метода Get для получения данных из веб-ресурсов и метод Output реализующие выгрузку данных в базы данных.

Также в разработанном модуле имеются классы, включенные в используемые фреймворки диаграмма данных классов приведена Приложении А.

#### **4.2 Пользовательский интерфейс**

Модуль автоматизированного извлечения веб-ресурсов является частью системы исследования общественного мнения. Данная система реализована как веб-приложение, в следствии чего пользовательский интерфейс представляет собой веб-страницы. Ниже приведена схематичная модель пользовательского интерфейса.

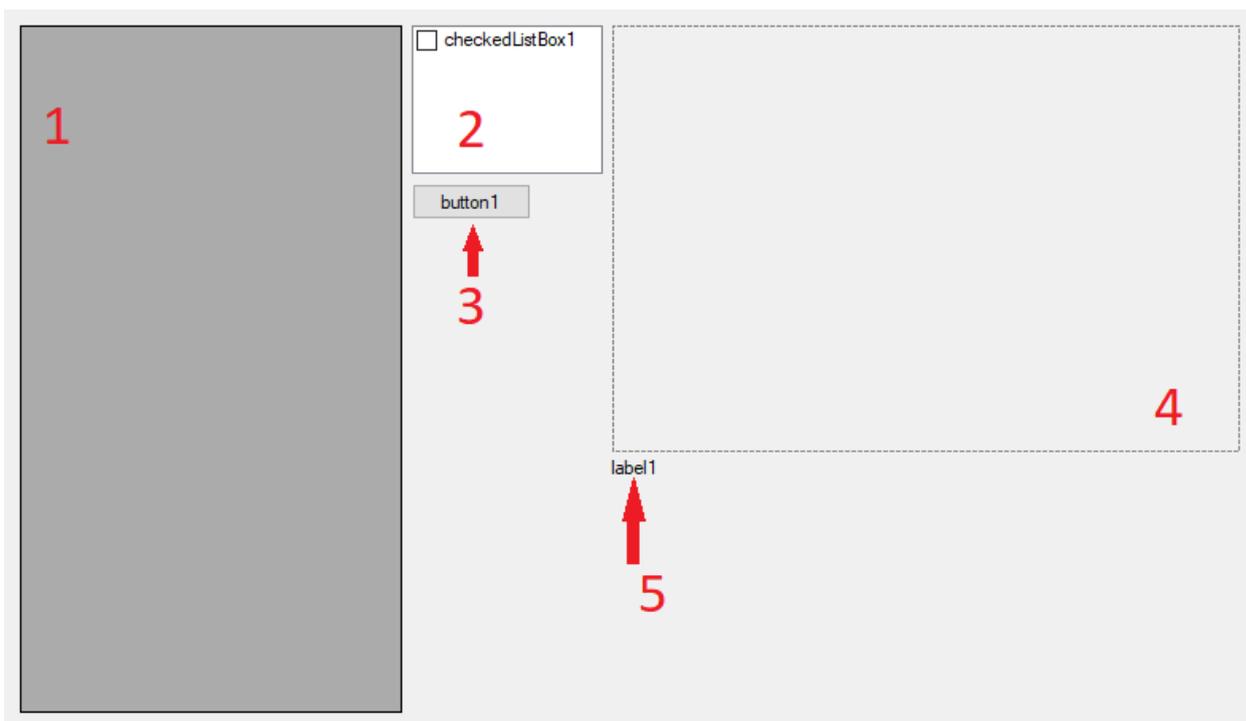


Рисунок 4.2 – Схематичная модель интерфейса

На модели находятся пять основных элементов, ниже приведено описание данных элементов:

1. Список новостей, выгружаемых из базы данных в соответствии с выбранной категорией.
2. Список категорий, на которые разбиты новости с новостного сайта.
3. Кнопка инициализирующая начало работы модуля.
4. Модуль, который производит визуальную интерпретацию данных полученных в результате работы программного продукта.
5. Модуль, производящий текстовую интерпретацию полученных данных для дополнения визуальной интерпретации данных.

### 4.3 Тестирование

Тестирование разработанного модуля проводилось с помощью тест-кейсов. Были разработаны тест-кейсы, часть из которых положительные, а вторая часть негативные. Ниже будут приведена часть тест-кейсов, разработанных в процессе тестирования модуля.

Вывод списка новостей в соответствии с категорией		
Действие	Ожидаемый результат	Результат теста
1) Запуск приложения	<ul style="list-style-type: none"> <li>• Окно “Main” открыто</li> <li>• “checkedListBox” корректно заполняется списком категорий</li> </ul>	
2) Выбираем необходимую категорию	<ul style="list-style-type: none"> <li>• В “список новостей” производится выгрузка списка новостей соответствующих выбранной категории</li> </ul>	

Таблица 4.1 –Позитивный тест-кейс для проверки выгрузки списка новостей из базы данных

С помощью данного тест-кейса производится проверка работоспособности части программного кода отвечающего за выгрузку списка новостей в выбранной категории. Данный тест-кейс является позитивным, это означает что в нем находятся корректная последовательность действий и корректный набор данных. Результатом проведения теста в соответствии с данным кейсом является верная работа программы, без каких-либо ошибок.

Запуск основного функционала программы без выбора новости		
Действие	Ожидаемый результат	Результат теста
1) Запуск приложения	<ul style="list-style-type: none"> <li>• Окно “Main” открыто</li> <li>• “checkedListBox” корректно заполняется списком категорий</li> </ul>	
2) Выбираем необходимую категорию	<ul style="list-style-type: none"> <li>• В “список новостей” производится выгрузка списка новостей соответствующих выбранной категории</li> </ul>	
3) Нажимаем на кнопку ”Выполнить”, не выбирая новость	<ul style="list-style-type: none"> <li>• Валидационное сообщение со всеми ошибками выведено на экран “Для начала работы программы необходимо выбрать новость, относительно которой происходит исследование”</li> </ul>	

Таблица 4.2 –Негативный тест-кейс проверки защиты от ошибки оператора

С помощью данного тест-кейса производится проверка работоспособности части программного кода на предотвращение возникновения критической ошибки, приводящей к аварийной остановке программы. Данный тест-кейс является негативным, это означает что данный тест-кейс оперирует как корректными, так и некорректными данными и ставит целью проверку исключительных ситуаций, а также проверяет, что вызываемая приложением функция не выполняется при срабатывании валидатора. Результатом проведения теста в соответствии с данным кейсом является верная работа программы, с предупреждением оператора о не корректной последовательности действий при работе с программой.

## **Общий вывод по разделу**

В ходе выполнения работы, были разработаны модули извлечения и предварительной обработки контента веб-ресурсов, выполняющие следующие задачи:

- проверка новостных сайтов на публикацию новых статей;
- извлечение новых статей в реальном времени;
- составление общего списка новостей на основе данных с нескольких ресурсов;
- получение списка пользователей входящих в целевую выборку;
- извлечение постов людей входящих в целевую выборку;
- производство предварительной обработки текста;
- приведение данных к формату необходимому для выгрузки в базу данных;
- выгрузка результата работы модуля в базу данных

Данный модуль создан для использования как в специализированных системах, которые занимаются анализом общественного мнения, так и в крупных системах, занимающихся поиском данных в сети интернет.

## 5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

### 5.1 Анализ конкурентных технических решений

Прямых конкурентов разработка в силу своей специфики не имеет, однако наиболее близкими по характеру и области применения являются NeOn и RDF-Translator (первый и второй конкурент соответственно). Результаты анализа представлены в виде оценочной карты в таблице 2.

Таблица 2. Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	Вес критерия	Вес критерия			Конкурентоспособность		
		Б <sub>ф</sub>	Б <sub>к1</sub>	Б <sub>к2</sub>	К <sub>ф</sub>	К <sub>к1</sub>	К <sub>к2</sub>
1	2	3	4	5	6	7	8
<b>Технические критерии оценки ресурсоэффективности</b>							
1. Надежность	0,1	3	4	1	0,3	0,4	0,1
2. Соответствие требованиям потребителей	0,25	5	1	2	1,25	0,25	0,5
3. Потребность в ресурсах памяти	0,15	5	3	4	0,75	0,45	0,6
4. Дополнительные утилиты	0,1	1	4	2	0,1	0,4	0,2
<b>Экономические критерии оценки эффективности</b>							
1. Конкурентоспособность продукта	0,15	3	2	2	0,45	0,3	0,3
2. Цена	0,05	2	4	4	0,1	0,2	0,2
3. Предполагаемый срок эксплуатации	0,15	4	2	1	0,6	0,3	0,15
4. Уровень проникновения на рынок	0,05	1	4	2	0,05	0,2	0,1
<b>Итого</b>	<b>1</b>	<b>24</b>	<b>24</b>	<b>18</b>	<b>3,6</b>	<b>2,5</b>	<b>2,15</b>

Согласно результатам, разрабатываемый модуль в наибольшей степени и по техническим, и по экономическим критериям соответствует требованиям разработчиков базы знаний.

## 5.2 SWOT-анализ

SWOT-анализ представляет собой комплексный анализ научно-исследовательского проекта. Он проводится в два этапа. По результатам проведения анализа представлена матрица в таблице 3.

Таблица 3. Матрица SWOT

	<p><b>Сильные стороны научно-исследовательского проекта:</b>  С1. Ускорение процесса обработки метаданных  С2. Отсутствие прямых аналогов скриптов-конвертеров из-за редкости использования требуемого стандарта С3.  Востребованность С4.  Объединение знаний в области онтологий и программирования на высокоуровневом языке  С5. Разработка на основе реальных данных</p>	<p><b>Слабые стороны научно-исследовательского проекта:</b>  Сл1. Сложность обработки входных данных при изменении их структуры  Сл2. Синтаксическая ригидность модуля  Сл3. Специфика метаданных научного эксперимента  Сл4. Сложность встраивания в общую систему  Сл5. Отсутствие контекстных пояснений к некоторым исходным данным</p>
<p><b>Возможности:</b>  В1. Поточковая автоматизация процесса импорта метаданных  В2. Использование для всех научных экспериментов сообщества  В3. Подключение новых источников метаданных  В4. Возможность перевода данных в RDF-триплеты  В5. Появление дополнительного спроса на продукты с используемым</p>	<p>За счет своей уникальности и специфики области применения, аналоги данного модуля, подстроенные под другую узкоспециализированную область знаний, могут найти применение и в других проектах.</p>	<p>Сложность базы знаний научного эксперимента и ее структуры требует дополнительных знаний в области онтологий для возможности подключения к базе новых источников данных.</p>
<p><b>Угрозы:</b>  У1. Неизвестны будущие источники метаданных  У2. Изменение стандартов языка, на котором написаны скрипты  У3. Длительность и сложность современных исследований  У4. Рост требований к вычислительным</p>	<p>Для дальнейшей перспективы работы с другими текстовыми источниками данных требуется унификация написанных элементов модуля.</p>	<p>Сложность поставленной задачи и работа с разнородными источниками входных данных ставит под угрозу возможность использования разработанного модуля с новыми источниками</p>

мощностям оборудования У5. Многоуровневое согласование на всех этапах работы		данных.
--	--	---------

Результаты, представленные в матрице, получены в ходе составления интерактивных матриц проекта, представленных в таблицах 4-7.

Таблица 4. Интерактивная матрица проекта

Сильные стороны проекта						
Возможности проекта		С1	С2	С3	С4	С5
	В1	+	0	+	-	0
	В2	+	+	+	-	0
	В3	+	0	+	+	+
	В4	0	+	+	+	+
	В5	-	+	+	+	-

Таблица 5. Интерактивная матрица проекта

Слабые стороны проекта						
Возможности проекта		Сл1	Сл2	Сл3	Сл4	Сл5
	В1	+	+	-	+	-
	В2	+	+	+	+	-
	В3	0	+	+	+	+
	В4	-	+	+	-	+
	В5	-	-	+	-	0

Таблица 6. Интерактивная матрица проекта

Сильные стороны проекта						
Угрозы проекта		С1	С2	С3	С4	С5
	У1	+	-	0	-	+
	У2	-	+	0	+	0
	У3	0	-	+	-	+
	У4	+	0	+	-	+
	У5	-	0	+	-	+

Таблица 7. Интерактивная матрица проекта

Слабые стороны проекта						
Угрозы проекта		Сл1	Сл2	Сл3	Сл4	Сл5
	У1	+	+	+	+	+
	У2	+	-	-	+	-
	У3	+	-	+	+	-
	У4	+	+	+	+	+
	У5	-	+	0	+	+

### 5.3 Определение возможных альтернатив проведения научных исследований

В рамках поставленной задачи возможных вариантов реализации было два: написание скриптов на Python или Java-приложения. Результаты морфологического анализа представлены в таблице 8.

Таблица 8. Морфологическая матрица для модуля импорта метаданных

	1	2
А. Среда разработки	IDLE (Python 3.6 64-bit)	Eclipse
Б. Редактор онтологий	Protégé	
В. Язык программирования	Python	Java
Г. Реализация	Скрипт	Консольное приложение

Как видно из выше приведенной таблицы, сформированы варианты: А1Б1В1Г1, А2Б1В2Г2

#### 5.3.1 Структура работ в рамках научного исследования

Перечень этапов выполнения проекта указаны в таблице 9. Таблица

9. Перечень этапов, работ и распределение исполнителей

Основные этапы	№ раб	Содержание работ	Должность исполнителя
Разработка технического задания	1	Составление и утверждение технического задания	Руководитель
Выбор направления исследований	2	Подбор материалов по теме	Руководитель
	3	Изучение материалов по теме	Бакалавр
	4	Выбор направления	Руководитель, бакалавр
	5	Календарное планирование работ по теме	Руководитель, бакалавр
Проектирование структуры и разработка ИС	6	Проектирование структуры модуля	Бакалавр
	7	Разработка модуля	Бакалавр
	8	Тестирование модуля	Бакалавр
Обобщение и оценка результатов	9	Оценка эффективности полученных результатов	Руководитель, бакалавр
Оформление отчета по НИР (комплекта документации по	10	Составление пояснительной записки	Бакалавр

ОКР)	(эксплуатационно-технической документации)	
------	--	--

### 5.3.2 Определение трудоемкости выполнения работ

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников научного исследования.

Трудоемкость выполнения научного исследования оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к. зависит от множества трудно учитываемых факторов. Для определения ожидаемого (среднего) значения трудоемкости  $t_{ож\ i}$  используется следующая формула:

$$t_{ож\ i} = \frac{3t_{\min\ i} + 2t_{\max\ i}}{5}, \quad (1)$$

где  $t_{ож\ i}$  – ожидаемая трудоемкость выполнения  $i$ -ой работы чел.-дн.;

$t_{\min\ i}$  – минимально возможная трудоемкость выполнения заданной  $i$ -ой работы (оптимистическая оценка: в предположении наиболее благоприятного стечения обстоятельств), чел.-дн.;

$t_{\max\ i}$  – максимально возможная трудоемкость выполнения заданной  $i$ -ой

работы (пессимистическая оценка: в предположении наиболее неблагоприятного стечения обстоятельств), чел.-дн.

Для выполнения перечисленных в таблице 9, указанной выше, работ требуются специалисты:

- бакалавр (Б);
- научный руководитель (Р).

Исходя из ожидаемой трудоемкости работ, определяется продолжительность каждой работы в рабочих днях  $T_p$ , учитывающая параллельность выполнения работ несколькими исполнителями. Такое вычисление необходимо для обоснованного расчета заработной платы, так как удельный вес зарплаты в общей сметной стоимости научных исследований составляет около 65 %.

$$T_{pi} = \frac{t_{ожi}}{Ч_i}, \quad (2)$$

где  $T_{pi}$  – продолжительность одной работы, раб. дн.;

$t_{ожi}$  – ожидаемая трудоемкость выполнения одной работы, чел.-дн.

$Ч_i$  – численность исполнителей, выполняющих одновременно одну и

ту же работу на данном этапе, чел.

### 5.3.3 Разработка графика проведения научного исследования

Для удобства построения графика, длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни. Для этого необходимо воспользоваться следующей формулой:

$$T_{ki} = T_{pi} \cdot k_{кал}, \quad (3)$$

где  $T_{ki}$  – продолжительность выполнения  $i$ -й работы в календарных днях;

$T_{pi}$  – продолжительность выполнения  $i$ -й работы в рабочих днях;

$k_{кал}$  – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}}, \quad (4)$$

где  $T_{\text{КАЛ}}$  – календарные дни ( $T_{\text{КАЛ}} = 366$ );

$T_{\text{ВД}}$  – выходные дни ( $T_{\text{ВД}} = 92$ );

$T_{\text{ПД}}$  – праздничные дни ( $T_{\text{ПД}} = 26$ ).

---

Временные показатели проведения научного исследования представлены в таблице 10.

Таблица 10. Временные показатели проведения научного исследования

№ работ	Трудоёмкость работ						Исполнители		Длительность работ в рабочих днях $T_{pi}$		Длительность работ в календарных днях $T_{ki}$	
	$t_{min}$ , чел-дни		$t_{max}$ , чел-дни		$t_{ожгi}$ , чел-дни				Испол.1	Испол.2	Испол.1	Испол.2
	Испол.1	Испол.2	Испол.1	Испол.2	Испол.1	Испол.2						
1	10	10	15	15	12	12	Р	Р	12	12	14	14
2	8	8	12	12	9,6	9,6	Р	Р	9,6	9,6	12	12
3	10	10	20	20	14	14	Б	Б	14	14	20	20
4	1	1	2	2	1,4	1,4	Р,Б	Р,Б	0,7	0,7	3	3
5	1	1	2	2	1,4	1,4	Р,Б	Р,Б	0,7	0,7	1	1
6	5	5	8	8	6,2	6,2	Б	Б	6,2	6,2	9	9
7	15	20	20	30	17	24	Б	Б	17	24	23	32
8	1	1	2	2	1,4	1,4	Б	Б	1,4	1,4	1	1
9	3	3	5	5	3,8	3,8	Р,Б	Р,Б	1,9	1,9	2	2
10	20	20	23	23	21,2	21,2	Б	Б	21,2	21,2	12	12
Итого	Всего								84,7	91,7	97	106
	Руководитель								24,9	24,9	32	32
	Бакалавр								63,1	70,1	71	80

На основании таблицы 10 строится календарный план-график. График строится для максимального по длительности исполнения работ в рамках научно-исследовательского. План-график приведен в таблице 11.

Таблица 11. Календарный план-график

№ работ	Вид работ	Исполнители	$T_{ki}$ , кал. дн.	Продолжительность выполнения работ											
				февраль			март			апрель			май		
				1	2	3	1	2	3	1	2	3	1	2	3
1	Составление и утверждение технического задания.	Руководитель	14	■	■	■									
2	Подбор материалов по теме	Руководитель	12		■	■									
3	Изучение материалов по теме	Бакалавр	20				▨	▨	▨						
4	Выбор направления	Руководитель, бакалавр	3						▨						
5	Календарное планирование работ по теме	Руководитель, бакалавр	1						▨						
6	Проектирование структуры ИС	Бакалавр	9						▨						
7	Разработка ИС	Бакалавр	23							▨	▨	▨			
8	Тестирование ИС	Бакалавр	1									▨			
9	Оценка эффективности и полученных результатов	Руководитель, бакалавр	2									▨			
10	Составление пояснительной записки	Бакалавр	12										▨	▨	▨

- руководитель



- бакалавр

### **5.3.4 Бюджет научно-технического исследования (НТИ)**

При планировании бюджета НТИ должно быть обеспечено полное и достоверное отражение всех видов расходов, связанных с его выполнением. В процессе формирования бюджета НТИ используется следующая группировка затрат по статьям:

- материальные затраты НТИ;
- затраты на специальное оборудование для научных (экспериментальных) работ;
- основная заработная плата исполнителей темы;
- дополнительная заработная плата исполнителей темы;
- отчисления во внебюджетные фонды (страховые отчисления);
- затраты научные и производственные командировки;
- контрагентные расходы;
- накладные расходы.

#### **5.3.4.1 Расчет материальных затрат НТИ**

Произведем расчет всех материалов, используемых при разработке проекта:

- приобретаемые со стороны сырье и материалы, необходимые для создания научно-технической продукции;

- покупные материалы, используемые в процессе создания научно-технической продукции для обеспечения нормального технологического процесса и для упаковки продукции или расходуемых на другие производственные и хозяйственные нужды (проведение испытаний, контроль, содержание, ремонт и эксплуатация оборудования, зданий, сооружений, других основных средств и прочее), а также запасные части для ремонта оборудования, износа инструментов, приспособлений, инвентаря, приборов, лабораторного оборудования и других средств труда, не относимых к основным средствам, износ спецодежды и других малоценных и быстроизнашивающихся предметов;

- покупные комплектующие изделия и полуфабрикаты, подвергающиеся в дальнейшем монтажу или дополнительной обработке;

- сырье и материалы, покупные комплектующие изделия и полуфабрикаты, используемые в качестве объектов исследований (испытаний) и для эксплуатации, технического обслуживания и ремонта изделий – объектов испытаний (исследований);

В материальные затраты, помимо вышеуказанных, включаются дополнительно затраты на канцелярские принадлежности, диски, картриджи и т.п. Однако их учет ведется в данной статье только в том случае, если в научной организации их не включают в расходы на использование оборудования или накладные расходы. В первом случае на них определяются соответствующие нормы расхода от установленной базы. Во втором случае их величина учитывается как некая доля в коэффициенте накладных расходов.

Расчет материальных затрат осуществляется по следующей формуле:

$$Z_m = (1 + k_T) \cdot \sum_{i=1}^m C_i \cdot N_{расxi} \quad , \quad (5)$$

где  $m$  – количество видов материальных ресурсов, потребляемых при выполнении научного исследования;

$N_{расxi}$  – количество материальных ресурсов  $i$ -го вида, планируемых к использованию при выполнении научного исследования (шт., кг, м, м<sup>2</sup> и т.д.);

$C_i$  – цена приобретения единицы  $i$ -го вида потребляемых материальных ресурсов (руб./шт., руб./кг, руб./м, руб./м<sup>2</sup> и т.д.);

$k_T$  – коэффициент, учитывающий транспортно-заготовительные расходы.

Материальные затраты представлены в таблице 12.

Таблица 12 Материальные затраты

Наименование	Единица измерения	Количество		Цена за ед., руб.		Затраты на материалы, (З <sub>м</sub> ), руб.	
		Исп.1	Исп.2	Исп.1	Исп.2	Исп.1	Исп.2
<b>Офисные принадлежности</b>							
Бумага для принтера А4	уп	1	1	200	200	200	200
Картридж для принтера	шт	1	1	600	600	600	600
<b>Электроэнергия</b>							
Электроэнергия	кВт	370	384	4	4	1480	1536
<b>Итого</b>						<b>2280</b>	<b>2336</b>

#### 5.3.4.2 Расчет затрат на специальное оборудование для научных (экспериментальных) работ

Дополнительное оборудование для научных (экспериментальных) работ не приобреталось.

#### 5.3.4.3 Основная заработная плата исполнителей темы

Рассчитаем основную заработную плату работников, непосредственно занятых выполнением НИТ, (включая премии, доплаты) и дополнительную заработную плату:

$$Z_{зп} = Z_{осн} + Z_{доп}, \quad (6)$$

где  $Z_{осн}$  – основная заработная плата;

$Z_{доп}$  – дополнительная заработная плата (12-20 % от  $Z_{осн}$ ).

Основная заработная плата ( $Z_{осн}$ ) руководителя (лаборанта, инженера) от предприятия (при наличии руководителя от предприятия) рассчитывается по следующей формуле:

$$Z_{осн} = Z_{дн} \cdot T_p, \quad (7)$$

где  $Z_{осн}$  – основная заработная плата одного работника;

$T_p$  – продолжительность работ, выполняемых научно-техническим работником, раб. дн. (табл. 7);

$Z_{дн}$  – среднедневная заработная плата работника, руб. Среднедневная заработная плата рассчитывается по формуле:

$$Z_{\text{дн}} = \frac{Z_{\text{м}} \cdot M}{F_{\text{д}}}, \quad (8)$$

где  $Z_{\text{м}}$  – месячный должностной оклад работника, руб.;

$M$  – количество месяцев работы без отпуска в течение года: при отпуске в 24 раб. дня  $M = 11,2$  месяца, 5-дневная неделя;

$F_{\text{д}}$  – действительный годовой фонд рабочего времени научно-технического персонала, показанный в таблице 13.

Таблица 13. Баланс рабочего времени

Показатели рабочего времени	Руководитель	Студент
Календарное число дней	365	365
Количество нерабочих дней	107	107
- выходные дни		
- праздничные дни		
Потери рабочего времени	24	24
- отпуск		
- невыходы по болезни		
Действительный годовой фонд рабочего времени	234	234

Месячный должностной оклад работника:

$$Z_{\text{м}} = Z_{\text{тс}} \cdot (1 + k_{\text{пр}} + k_{\text{д}}) \cdot k_{\text{р}}, \quad (9)$$

где  $Z_{\text{тс}}$  – заработная плата по тарифной ставке, руб.;

$k_{\text{р}}$  – районный коэффициент, равный 1,3 (для Томска). Расчёт основной заработной платы приведён в таблице 14.

Таблица 14. Расчёт основной заработной платы

Исп.	Исполнители	Разряд	$Z_{те}$ , руб.	$k_{пр}$	$k_{д}$	$k_{р}$	$Z_{м}$ , руб	$Z_{дн}$ , руб.	$T_{р}$ , раб. дн.	$Z_{осн}$ , руб.
Исполнение 1	Руководитель	Ведущий программист	20800	-	-	1,3	27040	1294,2	24,9	32225,58
	Бакалавр	-	2206,75	-	-	-	2206,75	106,1	63,1	6694,91
	<b>Итого</b>									
Исполнение 2	Руководитель	Ведущий программист	20800	-	-	1,3	27040	1294,2	24,9	32225,58
	Бакалавр	-	2206,75	-	-	-	2206,75	106,1	70,1	7437,61
	<b>Итого</b>									

#### 5.3.4.3 Дополнительная заработная плата исполнителей темы

Расчет дополнительной заработной платы ведется по следующей формуле:

$$Z_{доп} = k_{доп} \cdot Z_{осн}, \quad (10)$$

где  $k_{доп}$  – коэффициент дополнительной заработной платы (на стадии проектирования принимается равным 0,12 – 0,15). Примем коэффициент равный 0,12. Расчет дополнительной заработной платы указан в таблице 15.

Таблица 15. Расчёт дополнительной заработной платы

Исполнитель	Основная заработная плата, руб.		$k_{доп}$	Дополнительная заработная плата, руб.	
	Испол.1	Испол.2		Испол.1	Испол.2
Руководитель	32225,58	32225,58	0,12	3867,06	3867,06
<b>Итого</b>				3867,06	3867,06

#### 5.3.4.5 Отчисления во внебюджетные фонды (страховые отчисления)

Величина отчислений во внебюджетные фонды определяется исходя из следующей формулы:

$$З_{\text{внеб}} = k_{\text{внеб}} \cdot (З_{\text{осн}} + З_{\text{доп}}), \quad (11)$$

где  $k_{\text{внеб}}$  – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

На 2017 г. В соответствии с Федеральным законом от 24.07.2009 №212-ФЗ установлен размер страховых взносов равный 30%. На основании пункта 1 ст.58 закона №212-ФЗ для учреждений осуществляющих образовательную и научную деятельность в 2014 году водится пониженная ставка – 27,1%.

Отчисления во внебюджетные фонды представлены в таблице 16.

Таблица 16. Отчисления во внебюджетные фонды

Исполнитель	Основная заработная плата, руб.		Полная заработная плата, руб.	
	Испол.1	Испол.2	Испол.1	Испол.2
Руководитель	32225,58	32225,58	36092,64	36092,64
Коэффициент отчислений во внебюджетные фонды	0,271			
<b>Итого</b>				
Исполнение 1	<b>9781,1</b>			
Исполнение 2	<b>9781,1</b>			

#### **5.3.4.6 Расчет затрат на научные и производственные командировки**

На данном этапе в научных и производственных командировках нет необходимости.

#### **5.3.4.7 Контрагентные расходы**

В качестве контрагентных расходов учитывается услуга по предоставлению исходных данных – сторонними участниками коллаборации, не входящими в кадровый состав ТПУ. Приблизительная оценка – 5000 рублей.

#### **5.3.4.8 Накладные расходы**

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов исследования, оплата услуг связи, электроэнергии, почтовые и телеграфные расходы, размножение материалов и т.д. Их величина определяется по следующей формуле:

$$Z_{\text{накл}} = (\text{сумма статей } 1 \div 7) \cdot k_{\text{нр}}, \quad (12)$$

где  $k_{\text{нр}}$  – коэффициент, учитывающий накладные расходы.

Величину коэффициента накладных расходов можно взять в размере 16%.

### 5.3.4.9 Формирование бюджета затрат научно-исследовательского проекта

Определение бюджета затрат на научно-исследовательский проект по каждому варианту исполнения приведен в таблице 17.

Таблица 17. Расчет бюджета затрат НИИ

Наименование статьи	Сумма, руб.		Примечание
	Испол.1	Испол.2	
1. Материальные затраты НИИ	2280	2336	Пункт 5.4.3.1.
2. Затраты на специальное оборудование для научных (экспериментальных) работ	-	-	Пункт 5.4.3.2.
3. Затраты по основной заработной плате исполнителей темы	38920,49	39663,19	Пункт 5.4.3.3.
4. Затраты по дополнительной заработной плате исполнителей темы	3867,06	3867,06	Пункт 5.4.3.4.
5. Отчисления во внебюджетные фонды	9781,1	9781,1	Пункт 5.4.3.5.
6. Затраты на научные и производственные командировки	0	0	Пункт 5.4.3.6.
7. Контрагентские расходы	5000	5000	Пункт 5.4.3.7.
8. Накладные расходы	9575,78	9703,58	16 % от суммы ст. 1-7
9. Бюджет затрат НИИ	69424,43	70350,93	Сумма ст. 1- 8

### 5.3.5. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

*Интегральный показатель финансовой эффективности* научного исследования получают в ходе оценки бюджета затрат трех (или более) вариантов исполнения научного исследования. Для этого наибольший интегральный показатель реализации технической задачи принимается за базу расчета (как знаменатель), с которым соотносятся финансовые значения по всем вариантам исполнения.

**Интегральный финансовый показатель** разработки определяется как:

$$I_{\text{исп.}i}^{\text{финр}} = \frac{\Phi_{pi}}{\Phi_{\text{max}}} \quad (13)$$

где  $I_{\text{финр}}^{\text{исп.}i}$  – интегральный финансовый показатель разработки;

$\Phi_{pi}$  – стоимость  $i$ -го варианта исполнения;

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

Полученная величина интегрального финансового показателя разработки отражает соответствующее численное увеличение бюджета затрат разработки в разгах (значение больше единицы), либо соответствующее численное удешевление стоимости разработки в разгах (значение меньше единицы, но больше нуля).

Максимальная стоимость составляет 70350,93рублей, следовательно:

$$I_{\text{исп.1}}^{\text{финр}} = \frac{\Phi_{p1}}{\Phi_{\text{max}}} = 0,987$$

$$I_{\text{исп.2}}^{\text{финр}} = \frac{\Phi_{p2}}{\Phi_{\text{max}}} = 1$$

**Интегральный показатель ресурсоэффективности** вариантов исполнения объекта исследования можно определить следующим образом:

$$I_{pi} = \sum a_i \cdot b_i, \quad (14)$$

где  $I_{pi}$  – интегральный показатель ресурсоэффективности для  $i$ -го

варианта исполнения разработки;

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

$b^a$ ,  $i$   $b_i^p$  — бальная оценка  $i$ -го варианта исполнения разработки,

устанавливается экспертным путем по выбранной шкале оценивания;

$n$  – число параметров сравнения. Таблица 18. Сравнительная оценка характеристик вариантов исполнения проекта

Объект исследования Критерии	Весовой коэффициент параметра	Испол.1	Испол.2
1. Удобство внедрения	0,25	5	4
2. Функциональная мощность	0,20	4	3
3. Удобство в эксплуатации	0,25	5	4
4. Потребность в ресурсах памяти	0,15	5	3
5. Надежность	0,15	4	4
ИТОГО	1		

$$I_{p-исп1} = 5*0,25+4*0,20+5*0,25+5*0,15+4*0,15 = 4,65;$$

$$I_{p-исп2} = 4*0,25+3*0,20+4*0,25+3*0,15+4*0,15 = 3,65;$$

**Интегральный показатель эффективности вариантов исполнения разработки**

( $I_{исп.i}$ ) определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$I_{исп.1} = \frac{I_{p-исп1}}{I_{финр.1}}, \quad I_{исп.2} = \frac{I_{p-исп2}}{I_{финр.2}} \text{ и т.д.} \quad (15)$$

$$I_{исп.1} = \frac{4,65}{0,987} = 4,71; \quad I_{исп.2} = \frac{3,65}{1} = 3,65;$$

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволит определить сравнительную эффективность проекта и выбрать наиболее целесообразный вариант из предложенных. Сравнительная эффективность проекта ( $\mathcal{E}_{cp}$ ):

$$\mathcal{E}_{cp} = \frac{I_{исп.1}}{I_{исп.2}} = \frac{4,71}{3,65} = 1,28 \quad (16)$$

3,65 ; 4,71

*ср2*

;

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволит определить сравнительную эффективность проекта. Сравнительная эффективность разработки представлена в таблице 19.

Таблица 19. Сравнительная эффективность разработки

№ п/п	Показатели	Испол.1	Испол.2
1	Интегральный финансовый показатель исполнения	0,987	1
2	Интегральный показатель ресурсоэффективности исполнения	4,65	3,65
3	Интегральный показатель эффективности вариантов исполнения	4,71	3,65
4	Сравнительная эффективность вариантов исполнения	1,28	0,77

**Общий вывод по разделу:**

В результате работы по разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» выявили и сравнили два варианта исполнения научно-исследовательской работы. Бюджет затрат первого варианта исполнения равен 69424,43рублей, второго 70350,93. Произвели сравнительную оценку эффективности разработки и, исходя из полученных результатов, можно сделать вывод, что наиболее эффективным вариантом решения поставленной в бакалаврской работе технической задачи с позиции финансовой и ресурсной эффективности является 1 вариант исполнения – написание скриптов на языке Python.

## 6. Социальная ответственность

Объектом исследования выступает рабочее место статиста, разрабатывающего модуль импорта метаданных, пользователи которого могут использовать его с помощью своих персональных компьютеров или других устройств, на которых обеспечен доступ к общей информационной системе, для которой разрабатывается модуль. Модуль выполняет функцию импорта метаданных о научных экспериментах с последующим приведением данных в вид, пригодный для загрузки в общую информационную модель, для последующего решения поставленных пользователями информационной системы задач.

### 6.1 Производственная безопасность

Опасные и вредные факторы при разработке и эксплуатации проектируемого решения. Факторы по ГОСТ 12.0.003-74 «Опасные и вредные производственные факторы. Классификация» [6] представлена в таблице 20.

Таблица 20. Основные элементы производственного процесса, формирующие опасные и вредные факторы

Наименование видов работ и параметров производственного процесса	Факторы (ГОСТ 12.0.003-74 ССБТ ) [1]		Нормативные документы
	Вредные	Опасные	
Работа с ПК, устройствами ввода и вывода информации	Несоответствие параметрам микроклимата	Опасность поражения электрическим током	ГОСТ 12.1.005-88 [7], СанПиН 2.2.2/ 2.4.1340-03 [8], ГОСТ Р 50571. 17-2000 [9]
	Недостаточная освещенность рабочего места	Короткое замыкание	СанПиН 2.2.1/2.1.1.1278-03 [10]
	Повышенный уровень шума	Статическое электричество	СанПиН 2.2.4.3359-16 [11]
	Умственное перенапряжение		ГОИ Р-45-084-01 [12]

### 6.1.1 Параметры микроклимата

Оптимальное состояние воздушной среды должно обеспечивать ощущение теплового комфорта в течение 8-часового рабочего дня, не вызывать отклонений в состоянии здоровья. Энергетические затраты организма измеряются в ккал/ч (Вт) и по затраченной энергии работы разделяются на категории. Так работа программиста относится к категории Ia – интенсивность энергозатрат до 120 ккал/ч (до 139 Вт) [8].

Работы производятся в основном сидя и сопровождаются незначительным физическим напряжением. Допустимые параметры микроклимата на рабочем месте для категории Ia приведены в таблице 21.

Таблица 21. Допустимые величины показателей воздушной среды на рабочих местах производственных помещений по СанПиН 2.2.2/2.4.1340-03

Сезон года	Категория тяжести выполняемых работ	Температура, °С		Относительная влажность, %		Скорость движения воздуха, м/сек	
		Факт. значение	Допуст. значение	Факт. значение	Допуст. значение	Фактич. значение	Допуст. значение
1	2	3	4	5	6	7	8
Холодный	Ia	22-24	20-25	60	15-75	0,1	0,1
Теплый	Ia	23-25	21-28	60	15-75	0,1	0,2

Температура воздуха в рабочем помещении в холодное время года поддерживается в диапазоне (согласно измерениям термометром) от 21 до 23°C, в теплое – от 23 до 25°C. Влажность в соответствии с нормами (согласно измерениям гигрометром) колеблется около 60%. Для поддержания соответствующих микроклиматических параметров используются системы отопления и вентиляции, а также проводится кондиционирование воздуха в помещении.

### **6.1.2 Освещенность рабочего места**

Для обеспечения нормативных условий работы необходимо провести оценку освещенности рабочей зоны в соответствии с СанПиН 2.2.1/2.1.1.1278- 03.

Правильное освещение рабочих мест и помещений является важным условием для создания безопасных и благоприятных условий труда.

Все поле зрения должно быть освещено равномерно – это является основным гигиеническим требованием. Другими словами, уровень естественного освещения рабочего места и яркость дисплея компьютера должны быть приблизительно одинаковыми, т.к. яркий свет в зоне периферийного зрения заметно увеличивает глазное напряжение, что приводит к их быстрой утомляемости.

Для снижения отраженной блескости наряду с перечисленными выше рекомендуются следующие мероприятия:

Для внутренней отделки интерьера помещений с компьютерами должны использоваться диффузно отражающие материалы с коэффициентом отражения для потолка 0,7 – 0,8, для стен 0,5 – 0,6, для пола – 0,3 - 0,5.

Дизайн ПЭВМ должен предусматривать окраску корпуса в спокойные мягкие тона с диффузным рассеиванием света. Корпус ПЭВМ, клавиатура и другие блоки и устройства ПЭВМ должны иметь матовую поверхность с коэффициентом отражения 0,4 – 0,6 и не иметь блестящих деталей, способных создавать блики.

Для освещения помещений с ПЭВМ рекомендуется применять светильники с зеркальными параболическими решетками. Применение светильников без рассеивателей или экранирующих решеток нежелательно.

Персоналу, эксплуатирующему компьютеры с черными экранами, не рекомендуется использование светлой или блестящей одежды.

Фактические и требуемые параметры систем естественного и искусственного освещения показаны в таблице 22.

Таблица 22. Параметры систем естественного и искусственного освещения на рабочих местах по СанПиН

2.2.1/2.1.1.1278-03

Наименование рабочего места	Рабочая поверхность, м	Коэффициент естественной освещенности КЕО, %		Коэффициент совмещенного освещения, КСО, %		Искусственное освещение				
		при верхнем или комбинированном освещении	при боковом освещении	при верхнем или комбинированном освещении	при боковом освещении	Освещенность, лк			Показатель диска форта, М, не более	Коэффициент пульсации освещенности, %, не более
						при комбинированном освещении	при общем освещении	Всего		
Кабинет информатики	Г – 0,8; Экран дисплея: В – 1	3,5	1,2	2,1	0,7	500	300	300 200	15	10
Аудитории, учебные кабинеты	Г – 0,8;	3,5	1,2	2,1	0,7	-	-	300	40	10

Для кабинета информатики и учебных кабинетов норматив по СанПиН 2.2.1/2.1.1.1278-03 (искусственное общее освещение) минимальные значения освещённости - 300 люкс.

Равномерное освещение горизонтальной рабочей поверхности достигается при определенных отношениях расстояния между центрами светильников  $L$ (м) ( $L = 1,75 \cdot H$ ), к высоте их подвеса над рабочей поверхностью  $H$  (м). В помещении для разработки исследования высота составляет 2,5.

Количество люминесцентных ламп ЛБ 40 в количестве 24 шт. Согласно ГОСТ 6825-91. "Лампы люминесцентные трубчатые для общего освещения" номинальный световой поток люминесцентные лампы ЛБ 40 составляет 3000лк.

$$L = 1,75 \cdot 2,5 = 4,375(\text{м})$$

$$\Phi = \text{НСП} \cdot N = 3000 \cdot 24 = 72000 \text{ (люмен)}$$

НСП - номинальный световой поток лампы,  $N$  - количество ламп

$$I = \Phi / 2\pi = 72000 / 6,28 = 11465 \text{ (кандел)}$$

$$E_{\text{ф}} = I / L \cdot L = 11465 / 4,375 \cdot 4,375 = 599 \text{ (люкс)} \text{ (без учёта потерь в плафоне светильника)}$$

$\Delta E$ -показатель разности между фактическим и нормальным значением освещённости

$$\Delta E = (599 - 500) / 500 \cdot 100\% = 19,8\%$$

Отсюда можно сказать что в помещении подходящая система освещения, так как сохраняется допустимое отклонение освещённости 20%

### **6.1.3 Уровень шума**

Шум ухудшает условия труда, оказывают вредное воздействие на организм человека. Действие шума различно: он затрудняет разборчивость речи, вызывает снижение работоспособности, повышает утомляемость, вызывает необратимые изменения в органах слуха человека, ослабляет внимание, ухудшает память, снижает реакция, увеличивает число ошибок при работе.

Источником шума в учебных помещениях, эксплуатирующих вычислительную

технику, являются сами вычислительные машины (встроенные в стойки ЭВМ вентиляторы, принтеры и т.д.), центральная система вентиляции и кондиционирования воздуха и другое оборудование.

На рабочем месте статиста исследователя предельно допустимый уровень звука приведен таблице 23.

Таблица 23. Предельно допустимый уровень звука по СанПиН 2.2.4.3359-16

Рабочие места	Уровень звука, дБА
Рабочее помещение	50

При выполнении основной работы уровень звука не должен превышать 50 дБА. По субъективным ощущениям шумовая обстановка на рабочем месте исследователя соответствует норме.

Снизить уровень шума в помещениях можно использованием звукопоглощающих материалов с максимальными коэффициентами звукопоглощения в области частот 63-8000 Гц для отделки стен и потолка помещений. Дополнительный звукопоглощающий эффект создают однотонные занавески из плотной ткани, повешенные в складку на расстоянии 15-20 см от ограждения. Ширина занавески должна быть в 2 раза больше ширины окна.

#### **6.1.4 Умственное перенапряжение**

Для видов трудовой деятельности устанавливается 3 категории тяжести и напряженности работы с компьютером, которые определяются: для группы А - по суммарному числу считываемых знаков за рабочую смену (не более 60000 знаков за смену); для группы Б - по суммарному числу считываемых или вводимых знаков за рабочую смену (не более 40000 знаков за смену); для группы В - по суммарному времени непосредственной работы с компьютером за рабочую смену (не более 6 часов за смену).

Таблица 24. Категория работы по тяжести и напряженности по ТОО Р 45-084- 01

Категория работы по тяжести и напряженности	Уровень нагрузки за рабочую смену при видах работы на ПК		
	Группа А Количество знаков	Группа Б Количество знаков	Группа В Время работы, ч
III	До 60000	До 40000	До 6,0

При 8-часовой рабочей смене и работе на ПК регламентированные перерывы следует устанавливать:

Для третьей категории работ — через 1,5- 2,0 часа от начала рабочей смены и через 1,5-2,0 часа после обеденного перерыва продолжительностью 20 минут каждый или продолжительностью 15 минут через каждый час работы.

### **6.1.5 Опасность поражения электрическим током**

Согласно СанПиН 2.2.2/2.4.1340-03. «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы», рабочее место должно находиться в безопасной зоне, которое не характеризуется наличием таких условий, как повышенная влажность (относительная влажность воздуха длительно превышает 75%), высокая температура (более 35°С), токопроводящая пыль, токопроводящие полы, возможность одновременного соприкосновения к имеющим соединения с землей металлическим элементам и металлическим корпусам электрооборудования.

Электрические установки, к которым относится ПК, представляют для человека большую потенциальную опасность, так как в процессе эксплуатации или проведения профилактических работ человек может коснуться комплектующих компьютера, находящихся под напряжением.

Специфическая опасность – корпуса ПК и прочего оборудования, оказавшегося под напряжением в результате повреждения или пробоя изоляции, не подают каких-либо сигналов, которые предупреждают человека

об опасности. Причинами электропоражений являются: провода с поврежденной изоляцией, розетки сети без предохранительных кожухов.

Для защиты от поражения электрическим током все токоведущие части должны быть защищены от случайных прикосновений кожухами, корпус устройства должен быть заземлен. Заземление выполняется изолированным медным проводом сечением 1.5 мм, который присоединяется к общей шине заземления с общим сечением 48 м при помощи сварки. Общая шина присоединяется к заземлению, сопротивление которого не должно превышать 4 Ом.

Согласно ГОСТ Р 50571.17-2000. «Требования по обеспечению безопасности электроустановок в зданиях. Выбор мер защиты в зависимости от внешних условий. Защита от пожара», питание устройства в помещении, в котором выполнялась работа, осуществляется от силового щита через автоматический предохранитель, который срабатывает при коротком замыкании нагрузки. Для снижения величин возникающих разрядов применяются покрытия из антистатического материала.

Статист исследователь работает с электроприборами: компьютером (монитор, системный блок, компьютерная мышь и клавиатура).

В данном случае существует опасность электропоражения:

- при непосредственном прикосновении к токоведущим частям во время ремонта ПК;
- при прикосновении к нетоковедущим частям, оказавшимся под напряжением (в случае нарушения изоляции токоведущих частей ПК);
- при соприкосновении с полом, стенами, оказавшимися под напряжением;
- имеется опасность короткого замыкания в высоковольтных блоках: блоке питания и блоке дисплейной развёртки.

Средства предотвращения короткого замыкания:

- Необходимо следить за состоянием проводов;

- Исключить механическое воздействие на провода

Средства предотвращения удара статическим электричеством:

- Произвести заземление
- Установить УЗО(устройство защитного отключения)
- Производить систематическую влажную уборку

Рабочие место программиста исследователя оборудовано таким образом, чтобы исключить взаимное соприкосновение кабелей и шнуров питания соседних компьютеров.

К организационно-техническим мероприятиям относится первичный инструктаж по технике безопасности. Первичный инструктаж по технике безопасности является обязательным условием для допуска к работе в данном помещении.

## **6.2 Экологическая безопасность**

Разрабатываемый проект не имеет влияния на окружающую среду, так как само решение разрабатывается и используется внутри персональных компьютеров, которые могут стать источниками различных загрязнений.

Защита литосферы. Согласно ГОСТ Р 56397-2015 «Техническая экспертиза работоспособности радиоэлектронной аппаратуры, оборудования информационных технологий, электрических машин и приборов. Общие требования» пункт 5.8.1, после проведения технической экспертизы если оборудование не ремонтпригодно, то оно признается неработоспособным и рекомендуется к списанию (замене); в случае деградационного отказа оборудования и нецелесообразности его ремонта и модернизации даются рекомендации о необходимости его списания и утилизации. [12]

Согласно «Методики проведения работ по комплексной утилизации вторичных драгоценных металлов из отработанных средств вычислительной техники», утвержденной Государственным Комитетом РФ по телекоммуникациям от 19 октября 1999г. В п.3.1.3. «Технология разборки универсальных ЭВМ» расписаны 4 этапа разборки и подготовки к утилизации

внутренних частей ПК.

В результате выполнения этапов формируется партия сырья, включающая сортировку электронного лома по типу, проведение расчета количества ячеек, соединителей, серебросодержащих кабельных изделий, ячеек и типовых элементов замены, содержащие драгоценные металлы, а также партии черных и цветных металлов и сплавов (медь, сталь, никель, латунь, бронза, алюминий, дюралюминий, свинцово - оловянные припой) направляются на переработку на заводы ВДМ, полупроводниковые приборы (диоды, транзисторы), микросхемы в металлических и металлокерамических корпусах, а также конденсаторы в металлических корпусах демонтируются с плат и сортируются по типу, интегральные микросхемы в пластмассовых корпусах (серии 155, 551 и пр.) демонтируются и собираются отдельно, керамические конденсаторы типа КМ и резисторы после демонтажа также собираются отдельно.

На рабочем месте программиста используются 16 люминесцентных ламп ЛБ40, Согласно ГОСТ 12.3.031-83 «Работы со ртутью. Требования безопасности» п.2.1. все ртутьсодержащие отходы и вышедшие из строя приборы, содержащих ртуть, подлежат сбору и возврату для последующей регенерации ртути в специализированных организациях. В п.2.2. К работе по замене и сбору отработанных ртутьсодержащих ламп допускаются только электромонтеры. Главным условием при замене и сборе отработанных ртутьсодержащих ламп является сохранение герметичности. В п.2.13. Факт сдачи ртутьсодержащих отходов подтверждается возвращением паспорта на вывоз отходов с отметкой о приеме представителя специализированного предприятия [13].

### **6.3 Безопасность в чрезвычайных ситуациях**

К наиболее вероятным ЧС можно отнести следующие: пожар (взрыв) в здании, авария на коммунальных системах жизнеобеспечения, землетрясение. Источниками возгорания может стать электропроводка,

внутренние работающие устройства ПК, взрывоопасные предметы в помещении исследователя согласно ГОСТ 12.1.044-89 «Система стандартов безопасности труда. Пожаровзрывоопасность веществ и материалов. Номенклатура показателей и методы их определения» [14].

Превентивными мерами по предупреждению ЧС могут служить системы звукового и визуального оповещения персонала лаборатории и кабинетов об опасности, обучение персонала, методам работы с компьютером, наличие средств пожаротушения и информационных досок с планами эвакуации.

В случае угрозы возникновения ЧС необходимо отключить электропитание, вызвать по телефону пожарную команду, эвакуировать людей из помещения согласно плану эвакуации. При наличии небольшого очага пламени можно воспользоваться подручными средствами с целью прекращения доступа воздуха к объекту возгорания. В качестве подручных средств можно использовать углекислотные огнетушители ОУ-5 высокого давления с зарядом жидкой двуокиси углерода (по ГОСТ 8050-85), расположение которых можно найти на плане эвакуации людей при пожаре и других ЧС из помещений

#### **6.4 Правовые и организационные вопросы**

##### **обеспечения безопасности**

Рабочие места должны соответствовать требованиям ГОСТ 12.2.032-78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования» [15] и ГОСТ 12.2.061-81 «ССБТ. Оборудование производственное. Общие требования безопасности к рабочим местам» [16]. Работа с применением персональных компьютеров сопряжена со значительными зрительными и нервно-психологическими нагрузками, что повышает требования к организации труда пользователей ПК. Конструкция рабочей мебели должна обеспечивать возможность индивидуальной регулировки, соответственно росту работающего, и создавать удобную позу.

Часто используемые предметы труда и органы управления должны находиться в оптимальной рабочей зоне. Конструкция рабочего стола должна обеспечивать оптимальное размещение на рабочей поверхности используемого оборудования с учетом его количественных и конструктивных особенностей, а также характера выполняемой работы. Высота рабочей поверхности стола должна регулироваться в пределах 680-800 мм, при отсутствии такой возможности его высота должна быть не менее 725 мм.

На поверхности рабочего стола для документов необходимо предусматривать размещение специальной подставки, расстояние которой от глаз должно быть аналогичным расстоянию от глаз до клавиатуры. Модульными размерами рабочей поверхности стола, на основании которых должны рассчитываться конструктивные размеры, следует считать: ширину 800, 1000, 1200 и 1400 мм, глубину 800 и 1000 мм при нерегулируемой его высоте, равной 725 мм. Под столешницей рабочего стола должно быть свободное пространство для ног с размерами по высоте не менее 600 мм, по ширине 500 мм, по глубине 650 мм.

Конструкция рабочего стула должна обеспечивать поддержание рациональной рабочей позы при работе, что позволит изменять позу для снижения статического напряжения мышц шейно-плечевой области и спины для предупреждения развития утомления.

## **Заключение**

В данной работе был реализован алгоритм модуля извлечения и предварительной обработки контента из веб-ресурсов, для его последующей обработки в других модулях проекта нацеленного на анализ общественного мнения.

В процессе работы над модулем были алгоритмизированы, выбранные способы извлечения веб контента и способы предварительной обработки получаемых данных

Результатом работы является прототип системы способный производить анализ общественного мнения. В дальнейшем планируется улучшение системы путём увеличения охватываемой аудитории, путём добавления возможности работы с другими социальными сетями, а также увеличением количества новостных ресурсов и качества предварительной обработки получаемых данных.

## Список используемой литературы

1. А. Ахо, Дж. Ульман. Теория синтаксического анализа, перевода и компиляции. Т. 1. Пер. с англ. В.Н. Агафонова под ред. В. М. Курочкина. М.: Мир, 1978. 614 с.
2. А. Ахо, Дж. Ульман. Теория синтаксического анализа, перевода и компиляции. Т. 2. Пер. с англ. А.Н. Бирюкова и В.А. Серебрякова под ред. В. М. Курочкина. М.: Мир, 1978. 487 с.
3. Виталий Г. Блоги и RSS: интернет-технологии нового поколения [Текст] /Петербург.: «БХВ», 2006. – 274 с.
4. Markov Z, Larose D.T. Data-mining the Web : uncovering patterns in Web content, structure, and usage, - John Wiley & Sons Inc., 2007
5. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С.И. Елизаров. – 3-е издание перераб. и доп. – СПб.: БХВ-Петербург, 2009
6. ГОСТ 12.0.003-74. ССБТ. Опасные и вредные производственные факторы. Классификация.
7. ГОСТ 12.1.003–83 ССБТ. Шум. Общие требования безопасности.
10. СанПиН 2.2.2.548-96. Гигиенические требования к микроклимату производственных помещений.

11. СанПиН 2.2.1/2.1.1.1278–03. Гигиенические требования к естественному, искусственному и совмещённому освещению жилых и общественных зданий.
12. ГОСТ 12.1.045–84 ССБТ. Электростатические поля. Допустимые уровни на рабочих местах и требования к проведению контроля.
13. ГОСТ 12.1.038–82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов.
14. ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования.
15. ГОСТ 12.1.010-76 ССБТ. Взрывобезопасность. Общие требования.
16. СанПиН 2.2.2/2.4.1340–03. Санитарно-эпидемиологические правила и нормативы «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы».
17. Об утверждении правил обращения с отходами производства и потребления в части осветительных устройств, электрических ламп, ненадлежащие сбор, накопление, использование, обезвреживание, транспортирование и размещение которых может повлечь причинение вреда жизни, здоровью граждан, вреда животным, растениям и окружающей среде: Постановление Правительства Российской Федерации от 3 сентября 2010 года № 681.
18. Федеральный классификационный каталог отходов [Электронный ресурс]. – 2013. – Режим доступа: <http://www.ecoguild.ru/faq/fedwastecatalog.htm>, свободный. – Загл. с экрана.
19. Федеральный закон от 30.12.2001 N 197-ФЗ " Право работников на подготовку и дополнительное профессиональное образование, на прохождение независимой оценки квалификации"

20. Распоряжение МПР РФ от 2 декабря 2002 г. N 483-р "Об утверждении методических рекомендаций по организации лицензирования деятельности по обращению с опасными отходами на территории Российской Федерации"

## Приложение А

### Файл RssItem.java

```
public class RssItem {

    private String title;
    private String description;
    private Date pubDate;
    private String link;

    public RssItem(String title, String description, Date pubDate, String link) {
        this.title = title;
        this.description = description;
        this.pubDate = pubDate;
        this.link = link;
    }

    public String getTitle() {
        return this.title;
    }

    public String getLink() {
        return this.link;
    }

    public String getDescription() {
        return this.description;
    }

    public Date getPubDate() {
        return this.pubDate;
    }

    @Override
    public String toString() {

        SimpleDateFormat sdf = new SimpleDateFormat("MM/dd - hh:mm:ss");

        String result = getTitle() + " (" + sdf.format(this.getPubDate()) + ")";
        return result;
    }

    public static ArrayList<RssItem> getRssItems(String feedUrl) {

        ArrayList<RssItem> rssItems = new ArrayList<RssItem>();

        try {
            URL url = new URL(feedUrl);
            HttpURLConnection conn = (HttpURLConnection) url.openConnection();

            if (conn.getResponseCode() == HttpURLConnection.HTTP_OK) {
                InputStream is = conn.getInputStream();

                DocumentBuilderFactory dbf = DocumentBuilderFactory
                    .newInstance();
                DocumentBuilder db = dbf.newDocumentBuilder();

                Document document = db.parse(is);
                Element element = document.getDocumentElement();

                NodeList nodeList = element.getElementsByTagName("item");

                if (nodeList.getLength() > 0) {
                    for (int i = 0; i < nodeList.getLength(); i++) {
```

```

Element entry = (Element) nodeList.item(i);

Element _titleE = (Element) entry.getElementsByTagName(
    "title").item(0);
Element _descriptionE = (Element) entry
    .getElementsByTagName("description").item(0);
Element _pubDateE = (Element) entry
    .getElementsByTagName("pubDate").item(0);
Element _linkE = (Element) entry.getElementsByTagName(
    "link").item(0);

String _title = _titleE.getFirstChild().getNodeValue();
String _description = _descriptionE.getFirstChild().getNodeValue();
Date _pubDate = new Date(_pubDateE.getFirstChild().getNodeValue());
String _link = _linkE.getFirstChild().getNodeValue();

RssItem rssItem = new RssItem(_title, _description, _pubDate, _link);

rssItems.add(rssItem);
    }
}

}
} catch (Exception e) {
    e.printStackTrace();
}

return rssItems;
}

```

## Предложение Б

### Метод Get

```
String url =  
"https://api.twitter.com/1.1/statuses/user_timeline.json?screen_name=twitterapi&count=2";  
  
URL obj = new URL(url);  
URLConnection connection = (URLConnection) obj.openConnection();  
  
connection.setRequestMethod("GET");  
  
BufferedReader in = new BufferedReader(new InputStreamReader(connection.getInputStream()));  
String inputLine;  
StringBuffer response = new StringBuffer();  
  
while ((inputLine = in.readLine()) != null) {  
    response.append(inputLine);  
}  
in.close();
```

## Приложение В

### метод Parse

```
try {
    DocumentBuilder documentBuilder =
DocumentBuilderFactory.newInstance().newDocumentBuilder();
    Document document = documentBuilder.parse("BookCatalog.xml");

    Node root = document.getDocumentElement();

    System.out.println("List of books:");
    System.out.println();
    NodeList books = root.getChildNodes();
    for (int i = 0; i < books.getLength(); i++) {
        Node book = books.item(i);
        if (book.getNodeType() != Node.TEXT_NODE) {
            NodeList bookProps = book.getChildNodes();
            for (int j = 0; j < bookProps.getLength(); j++) {
                Node bookProp = bookProps.item(j);
                if (bookProp.getNodeType() != Node.TEXT_NODE) {
                    System.out.println(bookProp.getNodeName() + ":" +
bookProp.getChildNodes().item(0).getTextContent());
                }
            }
        }
    }
}

} catch (ParserConfigurationException ex) {
    ex.printStackTrace(System.out);
} catch (SAXException ex) {
    ex.printStackTrace(System.out);
} catch (IOException ex) {
    ex.printStackTrace(System.out);
}
```