

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  
 Отделение школы (НОЦ) Отделение информационных технологий

### МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
<b>Разработка модели для определения наиболее вероятной длительности задач проекта методами нейронных сетей</b>

УДК 004.7.032.26:005.8

Студент

Группа	ФИО	Подпись	Дата
8ВМ92	Курбонов Шамил Алиевич		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Фадеев А.С.	К.Т.Н.		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Верховская М.В.	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	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Кочегурова Е. А	К.Т.Н.		

## ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ООП

Код компетенции	Наименование компетенции
	<b>Универсальные компетенции</b>
УК(У)-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий
УК(У)-2	Способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	Способен организовывать и руководить работой команды, выработывая 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
	<b>Общепрофессиональные компетенции</b>
ОПК(У)-1	Способен самостоятельно приобретать, развивать и применять математические, естественно-научные, 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
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 проектировать сложные пользовательские интерфейсы
ПК(У)-3	Способен управлять процессами и проектами по созданию (модификации) информационных ресурсов

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  
 Отделение школы (НОЦ) Отделение информационных технологий

УТВЕРЖДАЮ:  
 Руководитель ООП  
 \_\_\_\_\_ Кочегурова Е.А.  
 (Подпись) (Дата) (Ф.И.О.)

### ЗАДАНИЕ на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации (бакалаврской работы, дипломного проекта/работы, магистерской диссертации)
--

Студенту:

Группа	ФИО
8ВМ92	Курбонову Шамилу Алиевичу

Тема работы:

Разработка модели для определения наиболее вероятной длительности задач методами нейронных сетей	
Утверждена приказом директора (дата, номер)	№ 95-29/с от 05.04.2021

Срок сдачи студентом выполненной работы:	01.06.2021
--	------------

#### ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p><b>Исходные данные к работе</b>  <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>Объектом проектирования и разработки является модель для определения наиболее вероятной длительности задач методами нейронных сетей и анализа рисков по задачам.</p>
<p><b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b>  <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. Анализировать факторы, которые влияют на риски;</li> <li>2. Разработка модели для определения длительности;</li> <li>3. Реализация программного обеспечения на основании полученной модели;</li> <li>4. Работа над разделом по финансовому менеджменту, ресурсоэффективности и ресурсосбережения;</li> </ol>

	5. Работа над разделом по социальной ответственности; 6. Работа над разделом на английском языке.
<b>Перечень графического материала</b> <i>(с точным указанием обязательных чертежей)</i>	1. Параметры оценки риска 2. Матрица градации рисков задачи 3. Графики связывание слов 4. Таблицы нейросетевых моделей
<b>Консультанты по разделам выпускной квалификационной работы</b> <i>(с указанием разделов)</i>	
<b>Раздел</b>	<b>Консультант</b>
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Доцент ОСГН ШБИП, к.э.н., Верховская М.В.
Социальная ответственность	Доцент ООД ШБИП, к.б.н., Антоневиц О.А.
Английский язык	Доцент ОИЯ ШБИП, к.ф.н., Аксёнова Н.В.
<b>Названия разделов, которые должны быть написаны на русском и иностранном языках:</b>	
Раздел 1. Development of models for determining the most probable duration of with NLP	

<b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b>	01.03.2021
---	------------

**Задание выдал руководитель:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Фадеев А.С.	к.т.н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ВМ92	Курбонов Шамил Алиевич		

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  
 Уровень образования Магистратура  
 Отделение школы (НОЦ) Отделение информационных технологий  
 Период выполнения \_\_\_\_\_ (осенний / весенний семестр 2020 /2021 учебного года)

Форма представления работы:

<b>Магистерская диссертация</b> (бакалаврская работа, дипломный проект/работа, магистерская диссертация)
---

### КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	01.06.2021
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
01.06.2021	Анализировать факторы, которые влияют на риски	20
01.06.2021	Разработка модели для определения длительности	20
01.06.2021	Реализация программного обеспечения на основании полученной модели	30
01.06.2021	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
01.06.2021	Социальная ответственность	10
01.06.2021	Приложение на английском языке	10

**СОСТАВИЛ:**

**Руководитель ВКР**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Фадеев А.С.	к.т.н.		

**СОГЛАСОВАНО:**

**Руководитель ООП**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Кочегурова Е. А	к.т.н.		

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

Студенту:

<b>Группа</b>	<b>ФИО</b>
8BM92	Курбонову Шамилу Алиевичу

<b>Школа</b>	<b>ИШИТР</b>	<b>Отделение школы (НОЦ)</b>	<b>ОИТ</b>
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

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

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Оклад инженера – 22 695,68 руб. Оклад руководителя – 35 111,5 руб.
2. Нормы и нормативы расходования ресурсов	Районный коэффициент 30%; Коэффициент дополнительной заработной платы 12%; Накладные расходы 16%.
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Коэффициент отчислений на уплату во внебюджетные фонды 30%

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

1. Оценка коммерческого и инновационного потенциала НТИ	- построить диаграмму Ганта
2. Разработка устава научно-технического проекта	- анализ конкурентных технических решений
3. Планирование процесса управления НТИ: структура и график поведения, бюджет, риски и организация закупок	- накладные расходы
4. Определение ресурсной, финансовой, экономической эффективности	- определение потенциального эффекта исследования

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

1. <i>«Портрет» потребителя результатов НТИ</i>
2. <i>Сегментирование рынка</i>
3. <i>Оценочная карта конкурентных технических решений</i>
4. <i>Диаграмма FAST</i>
5. <i>Матрица SWOT</i>
6. <i>График проведения и бюджет НТИ</i>
7. <i>Оценка ресурсной, финансовой и экономической эффективности НТИ</i>
8. <i>Потенциальные риски</i>

Дата выдачи задания для раздела по линейному графику	01.03.2021
--	------------

**Задание выдал консультант:**

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Доцент ОСГН ШБИП	Верховская М.В.	К.Э.Н.		

**Задание принял к исполнению студент:**

<b>Группа</b>	<b>ФИО</b>	<b>Подпись</b>	<b>Дата</b>
8BM92	Курбонов Шамил Алиевич		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

<b>Группа</b>	<b>ФИО</b>
8ВМ92	Курбонову Шамилу Алиевичу

<b>Школа</b>	<b>ИШИТР</b>	<b>Отделение (НОЦ)</b>	<b>ОИТ</b>
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

Тема ВКР:

<b>Разработка нейросетевого алгоритма распознавания надписей на изображениях реальных сцен</b>	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
<p>1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения</p>	<p>Объектом исследования является комплекс программных продуктов для реализации выгрузки реальных задач проекта Инициатор и тестирования модели методами нейронных сетей. Разработка модели для определения наиболее вероятной длительности задач методами нейронных сетей на рабочем месте, при помощи персонального компьютера, монитора, клавиатуры и компьютерной мыши.</p>
<b>Перечень вопросов, подлежащих исследованию, проектированию и разработке:</b>	
<p><b>1. Правовые и организационные вопросы обеспечения безопасности:</b></p> <ul style="list-style-type: none"> <li>— специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства;</li> <li>— организационные мероприятия при компоновке рабочей зоны.</li> </ul>	<p>Для соблюдения всех санитарно-технических и гигиенических правил и требований рассмотрен:</p> <ul style="list-style-type: none"> <li>– Специальные правовые нормы трудового законодательства при работе с компьютером и орг.техникой: ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования, СП 2.4.3648-20 «Санитарно-эпидемиологические требования к организациям воспитания и обучения, отдыха и оздоровления детей и молодежи»</li> <li>– Санитарно-эпидемиологические требования к условиям труда СП 2.2.3670-20.</li> </ul>
<p><b>2. Производственная безопасность:</b></p> <p>2.1. Анализ выявленных вредных и опасных факторов</p> <p>2.2. Обоснование мероприятий по снижению воздействия</p>	<p>В качестве вредных факторов выделены:</p> <ul style="list-style-type: none"> <li>– Повышенная или пониженная температура и влажность воздуха рабочей.</li> <li>– Недостаточная освещенность рабочей зоны;</li> <li>– Повышенный уровень шума на рабочем месте;</li> <li>– Повышенный уровень электромагнитных излучений;</li> <li>– Психофизиологические.</li> </ul> <p>В качестве опасных факторов выделены:</p> <ul style="list-style-type: none"> <li>– повышенное значение напряжения в электрической цепи, замыкание;</li> <li>– повышенная напряженность электрического поля.</li> </ul>
<p><b>3. Экологическая безопасность:</b></p>	<ul style="list-style-type: none"> <li>– Влияния на гидросферу и атмосферу разработка не оказывает;</li> <li>– анализ воздействия на литосферу (в работе рассмотрена утилизация неисправного ПК).</li> </ul>
<p><b>4. Безопасность в чрезвычайных ситуациях:</b></p>	<p>Возможно возникновение пожара.</p>

Дата выдачи задания для раздела по линейному графику	01.03.2021
--	------------

**Задание выдал консультант:**

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

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ВМ92	Курбонов Шамил Алиевич		

## Реферат

Выпускная квалификационная работа состоит из 103 с., 23 рис., 26 таблиц, 32 источников.

Ключевые слова: разработка, модель, машинное обучение, обработка текста, текстовые области, word2vec, статистика.

Объектом проектирования и разработки является модель для определения наиболее вероятной длительности задач методами нейронных сетей и анализа рисков по задачам.

Целью данной работы является исследование и анализ задач методами машинного обучения, для решения вопроса схожести и дальнейшего предсказания трудозатрата по задаче. Для её достижения необходимо выполнение следующих задач: анализировать факторы, которые влияют на риски, проанализировать существующие нейросетевые модели, разработать модель для определения длительности, реализовать программу на основе полученной модели, произвести обучение и тестирование разработанной модели.

В процессе исследования был проведён анализ рисков и разработан сервис для генерации факторов риска по задачам, на основании которого был произведён выбор модели для дальнейшей реализации.

В результате был предложен модель поверх нейросетевого алгоритма word2vec с возможностью добавления и редактирования факторов риска. В ходе проведения тестирования были подобраны такие параметры, которые показывают наиболее вероятные длительности задач с возможностью ограничить модель в рамках какой-то компании либо проекта. Было проведено увеличение исходной базы слов за счет национального корпуса русского языка и дампа википедии за 2018 год. В результате нейронная модель помогает принять решения при оценке длительности задачи.

В результате наиболее вероятная длительность задачи, рассчитанная программой, служит ориентиром при выставлении трудозатрат задачи. Функция помогает более точно назначить срок задачи, ориентируясь среднюю длительность выполнения задачи, а также на минимальный и максимальный срок выполнения.

Значительное отклонение от предлагаемой программой длительности может стать поводом для того, чтобы скорректировать установленные сроки.

## Содержание

Введение .....	13
1 Определение и оценка рисков задач .....	14
1.1 Параметры оценки риска .....	14
1.2 Градация рисков .....	18
1.3 Управление рисками задач .....	19
2 Разработка модели.....	20
2.1 Входные данные и нормализация.....	20
2.2 Качество выборки данных .....	21
2.3 Алгоритм word2vec .....	22
2.4 Подсчеты, совпадения и значение.....	24
2.5 Точечная взаимная информация и сжатие .....	25
2.6 Сходство слов и близость векторов .....	27
2.7 Аналогии и линейное пространство .....	28
2.8 Различия и прогнозы .....	30
2.9 Реализация алгоритма с помощью библиотеки gensim.....	30
3 Функция оценки длительности задач.....	34
3.1 Статистический анализ модели .....	34
3.2 Реализация оценки длительности задач.....	37
3.3 Интерфейс функции оценка длительности задачи .....	39
3.4 Интерфейс самые рискованные задачи .....	41
4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение .....	42
4.1 Предпроектный анализ .....	42
4.2 Планирование научно-исследовательских работ .....	51
4.3 Оценка сравнительной эффективности исследования .....	60
4.4 Вывод по разделу .....	63
5 Социальная ответственность .....	65
5.1 Правовые и организационные вопросы обеспечения безопасности .....	65
5.2 Производственная безопасность .....	69

5.3. Экологическая безопасность .....	79
5.4. Безопасность в чрезвычайных ситуациях .....	83
5.5 Выводы по разделу.....	86
Заключение .....	87
Список использованных источников .....	88
Приложение А .....	92
Приложение Б.....	102

## **Введение**

Благодаря методам машинного обучения мы наделяем компьютерам способности обрабатывать большие объемы данных, которые в противном случае стояли бы непроницаемой стеной, даем им возможность обучаться на этих данных и извлекать из них практические выводы.

Целью данной работы является исследование и анализ задач методами машинного обучения, для решения вопроса схожести и дальнейшего предсказания трудозатрата по задаче.

Для достижения данной цели необходимо:

- анализировать факторы, которые влияют на риски
- проанализировать существующие нейросетевые модели;
- разработать модель для определения длительности;
- реализовать программу на основе полученной модели;
- произвести обучение и тестирование разработанной модели.

Данная задача была поставлена в рамках проекта Инициатор для лучшего анализа и управление проектным предприятием. Для этого нам нужно учитывать все факторы, которые влияют на задачи и особенно оказывают значимое влияние на трудозатраты.

## 1 Определение и оценка рисков задач

Анализ рисков задачи предназначена для определения вероятности превышения сроков выполнения задачи. Для оценки рисков задачи используется количественная (балльная) система оценки. Итоговый балл оценки риска задачи определяется на основании анализа нескольких параметров задачи. Однако, точность определения трудозатрат может быть уменьшена если к примеру названия задачи, состоит всего лишь из одного слова, в таких случаях модель по большей части ориентируется на риски задач и прогнозирует трудозатраты относительно них. [31]

### 1.1 Параметры оценки риска

Для оценки рисков задачи используется несколько объективных и субъективных параметров. Каждому параметру присваивается значение от 1 (минимальное) до 5 (максимальное).

По формуле средневзвешенного значения рассчитывается оценка степени риска в баллах (1.1).

Формула расчета риска задачи:

$$\text{Риск задачи} = \frac{\left(\sum_{i=1}^n P_i\right) * (n - 1)}{n} - (n - 2) \quad (1.1)$$

$P_i$  - параметр,  $i = (1; n)$ ,  $n$  - количество параметров оценки

*Объективные параметры:*

**Вероятность** – этот параметр оценивает вероятность наступления негативного события, в данном случае - превышение установленного срока задачи. Если срок задачи не установлен или проект еще не начат (статус проекта - Инициация) или уже завершается (статус проекта - Завершение), то вероятность

просрочки минимальна [31].

**Влияние** – оценка риска с точки зрения его важности и влияния на проект в случае наступления негативного события (как повлияет превышение сроков выполнения задачи на весь проект в целом). Чем больше резерв задачи (см. сетевой план) - тем ниже влияние. У задачи, находящейся на критическом пути, влияние максимальное, поэтому такая задача получает максимальный балл [31].

**Текущее влияние** – оценка риска с точки зрения его влияния на раздел с учетом степени готовности раздела и сроков окончания раздела. Чем ближе срок начала выполнения задачи к окончанию раздела, тем выше степень риска [31].

*Субъективный параметр:*

**Рейтинг доверия исполнителя** – оценка надежности исполнителя исходя из процента просроченных им задач к общему числу задач этого исполнителя. Программа рассчитывает рейтинг доверия исполнителя на основании анализа всех его задач, имеющихся в базе программы Инициатор [31].

Таблица 1.1 – Параметры оценки риска

Параметр оценки	Значение	Расшифровка
Вероятность	1	Новая задача. Статус проекта - Инициация / Закрытие проекта. Сроки выполнения задачи могут быть не установлены. Задача еще не выполняется или уже выполнена.
	2	Новая задача. Проект находится в статусе Выполнение. Сроки выполнения задачи могут быть не еще установлены.
	3	Срок выполнения задачи истек на 50% от общей длительности. Задача не выполнена и/или не проверена (не установлен статус Выполнено/Проверено).  Коэффициент завершения задачи, рассчитанный по формуле Длительность задачи* в часах/2 ≤ 1. Сроки задачи переносились и/или увеличивались.  *Длительность задачи рассчитывается от Точки текущего времени
	4	Срок выполнения задачи истек на 66% от общей длительности, на выполнение задачи осталось всего 34% от общей длительности.  Задача не выполнена и/или не проверена (не установлен статус Выполнено/Проверено).

		<p>Коэффициент завершения задачи, рассчитанный по формуле  Длительность задачи* в часах/3 ≤ 1.  Задача не выполнена и/или не проверена (не установлен статус  Выполнено/Проверено).</p>
	5	<p>Срок выполнения задачи истек на 80% от общей длительности,  на выполнение задачи осталось всего 20% от общей  длительности.</p> <p>Задача не выполнена и/или не проверена (не установлен статус  Выполнено/Проверено).</p> <p>Коэффициент завершения задачи, рассчитанный по формуле  Длительность задачи* в часах/5 ≤ 1.  Задача не выполнена и/или не проверена (не установлен статус  Выполнено/Проверено).</p>
Влияние	1	Резерв задачи выше 31 трд.
	2	Резерв задачи от 21 до 30 трд.
	3	Резерв задачи от 11 до 20 трд.
	4	Резерв задачи от 1 до 10 трд.
	5	Резерв задачи 0 трд. (задача находится Критическом пути)
Текущее влияние (в стадии)	1	Проект находится в статусе Инициация/Закрытие. Задача еще не выполняется или уже выполнена
	2	Дата начала задачи приходится на первую половину раздела
	3	Дата начала задачи приходится на середину и вторую половину раздела
	4	Сдача альбомов, закрытие работ по разделу через 15 дней и менее*
	5	Сдача альбомов, закрытие работ по разделу через 5 дней и менее*
Рейтинг доверия исполнителя	1	Количество просроченных задач относительно к общему количеству задач, выполняемых на текущий момент = 0%
	2	Количество просроченных задач относительно к общему количеству задач, выполняемых на текущий момент - от 1% до 10%
	3	Количество просроченных задач относительно к общему количеству задач, выполняемых на текущий момент - от 11% до 24%
	4	Количество просроченных задач относительно к общему количеству задач, выполняемых на текущий момент - от 25% до 80%
	5	Количество просроченных задач относительно к общему количеству задач, выполняемых на текущий момент - от 81% до 100%

Если исполнитель задачи не назначен, то риски задачи все равно просчитываются, при оценке риска используется три параметра.

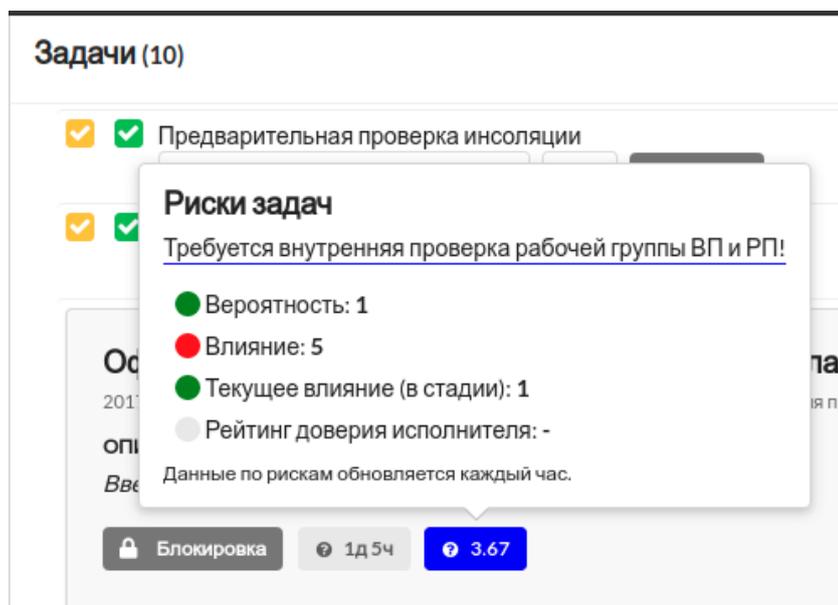


Рисунок 1.1 – Исполнитель задачи не назначен (при расчете риска используются только три параметра)

Пример расчета риска задачи:

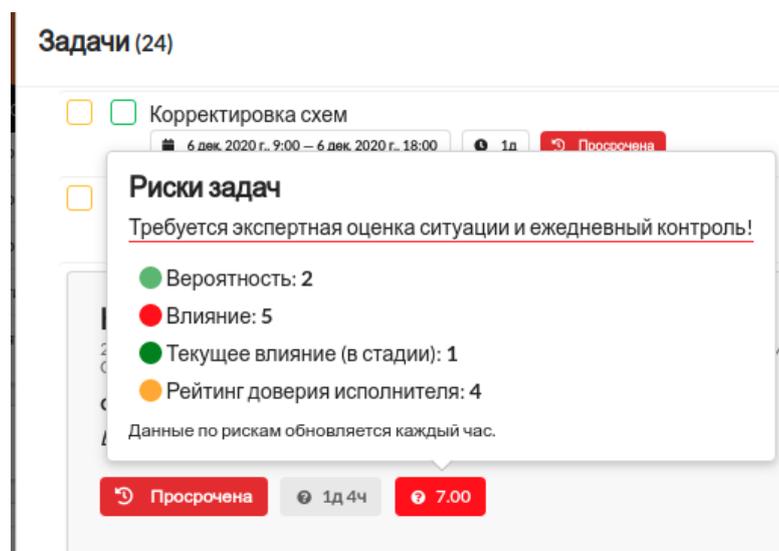


Рисунок 1.2 – Интерфейс функции Оценка риска задач

Частная формула для расчета риска задачи (1.2):

$$\text{Риск задачи} = \frac{(\text{Вероятность} + \text{Влияние} + \text{Тек. влияние} + \text{рейтинг доверия}) * (n - 1)}{n} - (n - 2) \quad (1.2)$$

Таким образом, подставив значения (см. рис.2), получим следующее выражение:

$$\text{Риск задачи} = \frac{(2+5+1+4) \cdot (4-1)}{4} - (4-2) = 7,0$$

## 1.2 Градация рисков

Классификация рисков – это подразделение рисков на категории с применением определенных критериев. Выбор критериев зависит от целей и особенностей процедуры анализа и управления риском. Число возможных критериев классификации риска может быть очень большим, поэтому сами критерии классификации необходимо группировать [7].

Основная цель классификации — выделение конкретных рисков, а каждый из конкретных рисков, измеряемый частотой возникновения и размером неблагоприятных последствий (ущерба), описывается его стандартными характеристиками:

- опасность, связанная с риском;
- подверженность риску;
- уязвимость (чувствительность к риску);
- другими дополнительными параметрами (характеристиками), такими как взаимодействие с другими рисками, степень однородности риска, степень прогнозируемости и др.

Степень риска оценивается в соответствии с Матрицей градации рисков.

n - количество учитываемых факторов

Таблица 1.2 – Матрица градации рисков задачи

<b>Степень риска</b>	<b>n=4</b>	<b>n=3</b>
Низкий	1 - 3	1 - 3,67
Средний	4 - 7	3,67 - 5,67
Высокий	7 - 10	5,67 - 7
Очень высокий	10 - 13	7 - 9

### 1.3 Управление рисками задач

После того, как риск задачи рассчитан и оценен в баллах (см. Оценка рисков задачи. Матрица градации риска) программа Инициатор предлагает действия в данной ситуации.

Таблица 1.3 – Рекомендуемые действия

<b>Степень риска</b>	<b>Рекомендуемые действия</b>
Низкий	Анализ ситуации не требуется
Средний	Требуется внутренняя проверка рабочей группы ВП и РП
Высокий	Требуется оценка ситуации и еженедельный контроль
Очень высокий	Требуется экспертная оценка ситуации и ежедневный контроль

## 2 Разработка модели

### 2.1 Входные данные и нормализация

Для реализации программного обеспечения был выбран язык программирования Python, по той причине что для него реализованы библиотеки для работы с текстовыми нейронными сетями. Были использованы функции открытой библиотеки Gensim. В качестве математических вычислений был выбран библиотека NumPy для работы с основными функциями статистики.

Программный код был написан в редакторе PyCharm, также частично использовался WebStorm.

В качестве глобальной выборки на данный момент, мы имеем почти миллион задач, и в рамках этих задач есть задачи со следующими статусами (рис. 2.1).

```
const
    STATUS_NEW = 1,           // Запланирована
    STATUS_CANCELLED = 2,    // Отменена
    STATUS_NEED_EDIT = 3,   // Требуется изменения
    STATUS_ISSUED = 4,       // Возникла проблема
    STATUS_DONE = 5,         // Выполнена
    STATUS_VERIFIED = 6,    // Принята
```

Рисунок 2.1 – Статусы задач

Исходя из статусов задач, мы видим, что у задач, которые со статусом от 1 до 4, никаких точных данных по трудозатратам нет, конечно мы можем исходить из каких-то факторов, к примеру, если статус отменено значит можно ориентироваться и давать меньше рейтинга, если совпадают названия задачи по смыслу с тем для которого мы пытаемся предсказать, но для первого этапа такие малозначимые факторы не учитываются, т.к. есть более значимые факторы и постепенно можно улучшать алгоритм и добавлять эти факторы.

Ситуация со статусом выполнено очень интересная, с одной стороны это задачи где есть для нас все необходимые значения, трудозатраты,

исполнители и прочее, с другой в рамках проекта возможно будут доработки по данному статусу, по этой причине оставим этот статус.

В качестве основной выборки, мы получаем задачи со статусом Принято, а их в рамках проекта почти 40 тыс. таким образом основные факторы на которых стоит ориентироваться в рамках первого релиза, это дедлайны задач из которых мы получим общий период выполнения, владелец и название задачи.

Таким образом, мы получили основную выборку для анализа задач в csv формате. Для автоматического обновления данного списка и получения новых успешных задач был создан отдельная команда.

## **2.2 Качество выборки данных**

Многие современные системы и техники НЛП рассматривают слова как элементарные единицы - нет понятия сходства между словами, поскольку они представлены в виде индексов в словаре. В этом выборе есть несколько хороших причины - простота, надежность и наблюдение, что простые модели обучаются на огромном количестве данные превосходят сложные системы, обученные на меньшем количестве данных. Примером может служить популярная модель N-грамм используется для статистического моделирования языка - сегодня N-граммы можно обучать практически на всех доступных данные (триллионы слов [3]).

Однако для многих задач простые методы ограничены. Например, количество

релевантные внутренние данные для автоматического распознавания речи ограничены - производительность обычно преобладает размер высококачественных записанных речевых данных (часто всего миллионы слов). В машинный перевод, существующие корпуса для многих языков содержат всего несколько миллиардов слов или менее. Таким образом, бывают ситуации, когда простое масштабирование основных методов не приведет к какой-либо значительный прогресс, и мы должны сосредоточиться на более продвинутой

методах. С развитием техники машинного обучения в последние годы стало возможным обучать большому количеству людей, сложные модели на гораздо большем наборе данных, и они обычно превосходят простые модели. Наверное, наиболее удачная концепция

- использование распределенных представлений слов [10]. Например, нейронные сетевые языковые модели значительно превосходят модели N-грамм [17].

*Качество выборки зависит от точности названия задач.*

Чем точнее названа задача, тем точнее рассчитывается ее ожидаемая длительность, поскольку при расчетах будут анализироваться максимально подобные задачи. Из названия задачи должно быть понятен ее содержания.

Задачи, название которых состоит из одного слова (например: “Расчеты”, “Печать”), без уточнения вида работ, в выборку не включаются.

Стандартное единообразное наименование задач позволяет сделать выборку более точной и определять ожидаемую длительность для максимально конкретной задачи.

В качестве основных требований, отметим в данном подразделе, что нужно учитывать загрузки разных факторов, которые влияют на дедлайны задач, и чтобы в любое время можно было добавлять другие входные параметры для анализа, таких как лояльность сотрудника, неудачи сотрудника при решении задачи из данной стадии и др.

### **2.3 Алгоритм Word2Vec**

Word2vec — алгоритм, который преобразует слова в векторы, так что слова с одинаковым значением в конечном итоге оказываются близко друг к другу, разработанный в 2013 году. Более того, это позволяет нам использовать векторную арифметику для работы с аналогиями, например, знаменитый король - мужчина + женщина = королева. Кроме работы со словами, некоторые его концепции оказались эффективны в разработке рекомендательных механизмов и

придании смысла данным даже в коммерческих, неязыковых задачах. Эту технологию применили в своих движках рекомендаций такие компании, как Airbnb, Alibaba, Spotify и Anghami [1].

Я попытаюсь объяснить, как это работает, с особым акцентом на значение векторных различий, в то же время опуская как можно больше технических деталей.

Пожалуй, приведу самый популярный пример (рис 2.2).

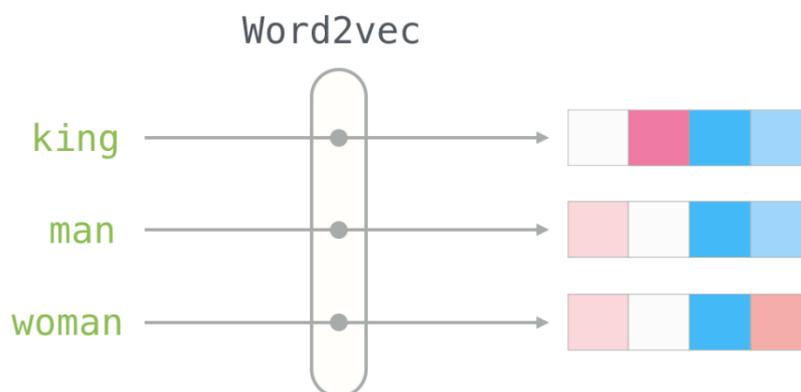


Рисунок 2.2 – Смысловое представление слов в word2vec

Также еще более наглядно, можно представить смысловую связь слов (рис. 2.3)

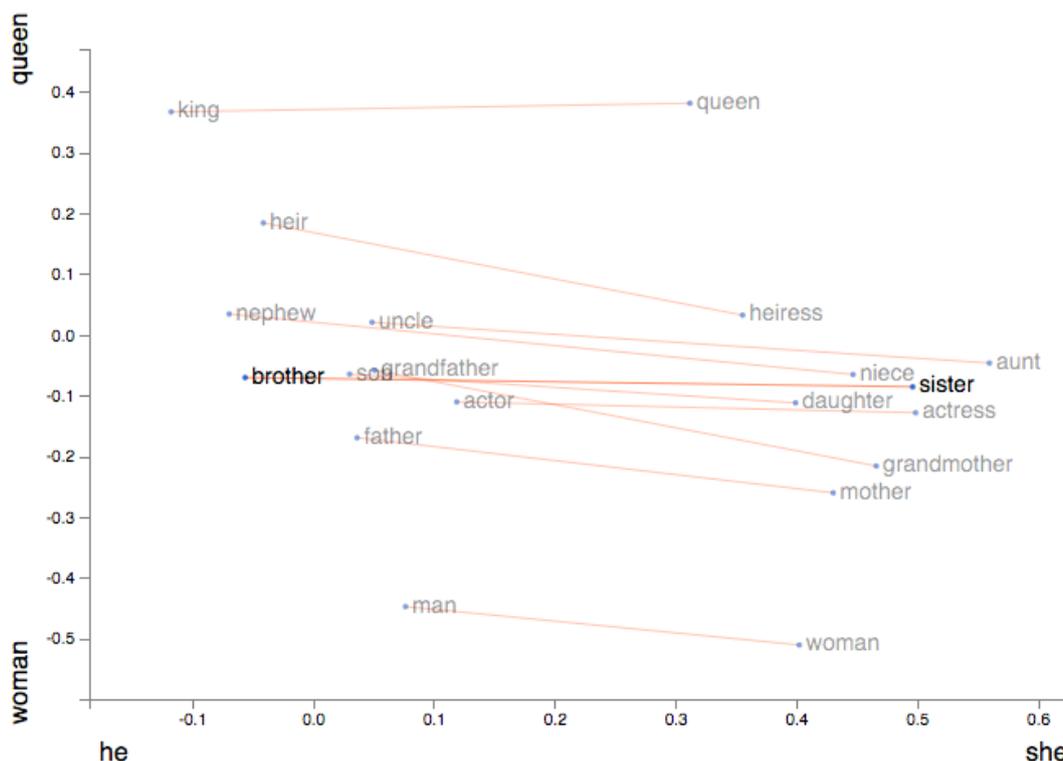


Рисунок 2.3 – Связывание слов

## 2.4 Подсчеты, совпадения и значение

*I love letter co-occurrence in the word co-occurrence.*

Иногда, казалось бы, наивная техника дает впечатляющие результаты. Простой просмотр совпадений слов при игнорировании всей грамматики и контекста может дать нам представление о значении слова. Рассмотрим это предложение:

*A small, fluffy roosety climbed a tree.*

Что за румяное? Я бы сказал, что что-то вроде белки, так как эти два слова можно легко поменять местами. Такое рассуждение называется распределительной гипотезой и может быть резюмировано следующим образом:

*a word is characterized by the company it keeps - John Rupert Firth*

Если мы хотим научить этому компьютер, самый простой приблизительный подход - заставить его смотреть только на пары слов. Пусть  $P(a / b)$  будет условной вероятностью того, что для данного слова  $b$  есть слово  $a$  на небольшом расстоянии (скажем, на расстоянии не более двух слов). Затем мы утверждаем, что два слова  $a$  и  $b$  подобны, если для каждого слова  $w$ .

Другими словами, если у нас есть это равенство, независимо от того, есть ли слово  $a$  или  $b$ , все другие слова встречаются с той же частотой (2.1).

$$P(w|a)=P(w|b)P(w|a)=P(w|b) \quad (2.1)$$

Даже простой подсчет слов, сравниваемый по источникам, может дать интересные результаты, например, что в текстах металлических песен слова (крики, вечность или пепел популярны, в то время как слова конкретно или приблизительно, особенно часто), см. *Heavy Metal and Natural Language Processing*. См. Также Гендерные роли с интеллектуальным анализом текста и N-граммы Джулии Силге.

Гораздо больше информации можно получить, глядя на совпадения (рис. 2.4). Например, один из моих проектов, дает представление о структуре программирования, основанном только на использовании тегов для

автоматической генерации сущностей. Это также показывает, что я люблю точечную взаимную информацию (<https://github.com/shamil8/nlp-tags>), что подводит нас к следующему пункту.

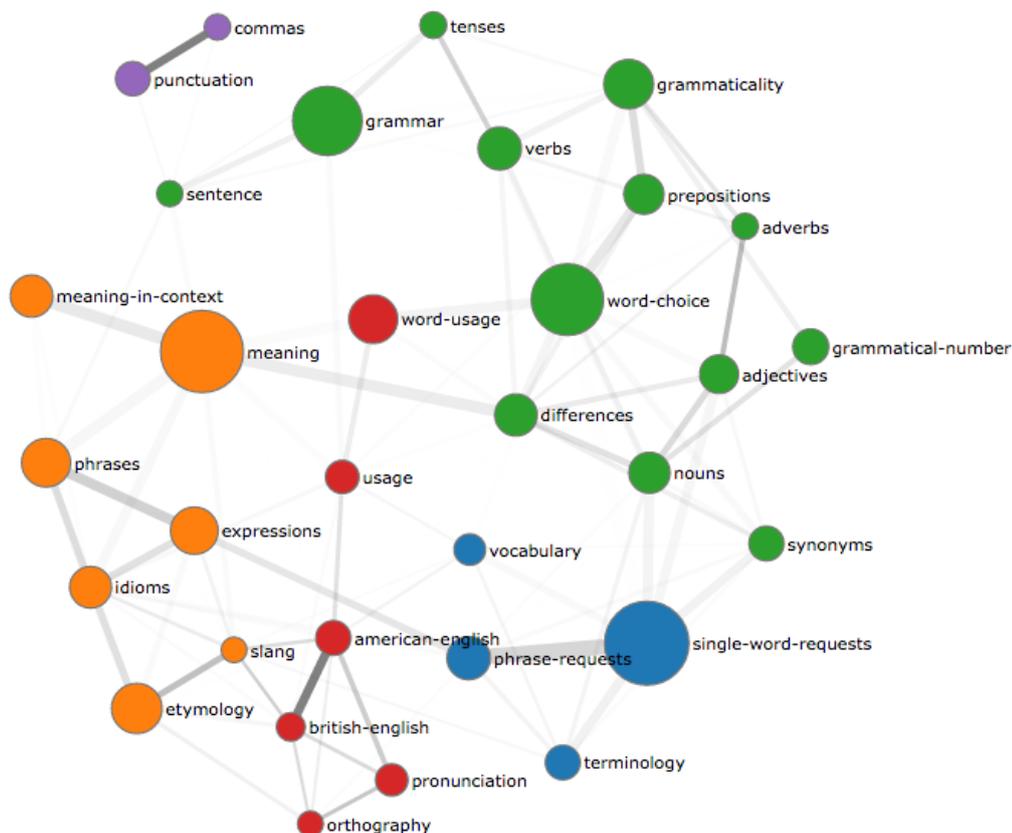


Рисунок 2.4 – Совместное употребление слов

## 2.5 Точечная взаимная информация и сжатие

Компрессоры – воспринимайте познание как сжатие. Сжатое восприятие, приближительная матричная факторизация - (n) Культуры машинного обучения.

В принципе, мы можем вычислить  $P(a / b)$  для каждой пары слов. Но даже с небольшим словарем на 100 000 слов (имейте в виду, что нам нужно сохранить все склонения, имена собственные и вещи, которые не входят в официальные словари, но которые используются), для отслеживания всех пар потребуется 8 гигабайт места.[3]

Часто вместо работы с условными вероятностями мы используем точечную взаимную информацию ( $PMI$ ), определяемую как (2.2):

$$\begin{aligned} \text{PMI}(a,b) &= \log[P(a,b)P(a)P(b)] = \log[P(a|b)P(a)], \\ \text{PMI}(a,b) &= \log_{f_0}[P(a,b)P(a)P(b)] = \log_{f_0}[P(a|b)P(a)], \quad (2.2) \end{aligned}$$

Его прямая интерпретация состоит в том, насколько более вероятно, что мы получим пару, чем если бы она была случайной. Логарифм упрощает работу со словами, встречающимися с частотой разного порядка. Мы можем аппроксимировать PMI как скалярное произведение:

$$\text{PMI}(a,b) = \vec{v}^T a \cdot \vec{v}^T b, \quad \text{PMI}(a,b) = \vec{v} \rightarrow a \cdot \vec{v} \rightarrow b, \quad (2.3)$$

где  $\vec{v}^T i \vec{v} \rightarrow i$  - векторы, обычно 50-300 размерностей (в нашем случае будет 300 по стандарту Google).

На первый взгляд может показаться странным, что все слова можно сжать до пространства гораздо меньшей размерности. Но есть слова, которые можно тривиально поменять местами (например, от Иоанна до Петра), и в целом имеется много структур.

Тот факт, что это сжатие с потерями, может дать ему преимущество, поскольку оно может обнаруживать шаблоны, а не только запоминать каждую пару. Например, в системах рекомендаций для рейтингов фильмов каждый рейтинг аппроксимируется скалярным произведением двух векторов - содержания фильма и предпочтений пользователя. Это используется для прогнозирования оценок для еще не просмотренных фильмов, см. Факторизация матрицы с помощью TensorFlow - Кэтрин Бейли [3].

## 2.6 Сходство слов и близость векторов

Покажем сходство слов в векторном пространстве. Условие, что  $P(w|a)=P(w|b)P(w|a)=P(w|b)$  эквивалентно.

$$\text{PMI}(w,a)=\text{PMI}(w,b), \text{PMI}(w,a)=\text{PMI}(w,b), \quad (2.4)$$

разделив обе части на  $P(w)$  и взяв их логарифм. Выразив PMI с помощью векторных произведений, мы получаем:

$$\begin{aligned} \vec{v} \cdot \vec{w} \cdot \vec{v} \cdot \vec{a} &= \vec{v} \cdot \vec{w} \cdot \vec{v} \cdot \vec{b} \quad \vec{v} \rightarrow \vec{w} \cdot \vec{v} \rightarrow \vec{a} = \vec{v} \rightarrow \vec{w} \cdot \vec{v} \rightarrow \vec{b} \\ \vec{v} \cdot \vec{w} \cdot (\vec{v} \cdot \vec{a} - \vec{v} \cdot \vec{b}) &= 0, \quad \vec{v} \rightarrow \vec{w} \cdot (\vec{v} \rightarrow \vec{a} - \vec{v} \rightarrow \vec{b}) = 0, \end{aligned} \quad (2.5)$$

Если это должно работать для каждого  $\vec{v} \quad \vec{w} \vec{v} \rightarrow \vec{w}$ , то:

$$\vec{v} \cdot \vec{a} = \vec{v} \cdot \vec{b}, \quad \vec{v} \rightarrow \vec{a} = \vec{v} \rightarrow \vec{b}, \quad (2.6)$$

Конечно, в каждом практическом случае мы не получим точного равенства, просто слова, близкие друг к другу. Слова, близкие к этому пространству, часто являются синонимами (например, счастлив и рад), антонимами (например, добро и зло) или другими легко взаимозаменяемыми словами (например, желтым и синим). В частности, у большинства противоположных идей (например, религии и атеизма) будет похожий контекст.

Что мне гораздо интереснее, так это то, что слова образуют линейное пространство. В частности, нулевой вектор представляет собой совершенно нехарактерное слово, встречающееся с любым другим словом на уровне случайной вероятности (поскольку его скалярное произведение на каждое слово равно нулю, равно как и его PMI) [5].

Это одна из причин, почему для векторного подобия люди часто используют косинусное расстояние, т.е.:

$$\vec{v} \cdot \vec{a} \cdot \vec{v} \cdot \vec{b} / |\vec{v} \cdot \vec{a}| |\vec{v} \cdot \vec{b}| \quad \vec{v} \rightarrow \vec{a} \cdot \vec{v} \rightarrow \vec{b} / |\vec{v} \rightarrow \vec{a}| |\vec{v} \rightarrow \vec{b}|, \quad (2.8)$$

То есть он делает акцент на направлении, в котором данное слово встречается вместе с другими словами, а не на силе этого эффекта.

## 2.7 Аналогии и линейное пространство

Если мы хотим провести аналогии со словами ( $a$  относится к  $b$ , как  $A$  относится к  $B$ ), можно утверждать, что  $in$  может быть выражено как равенство условных отношений вероятностей (2.9).

$$P(w|a) P(w|b) = P(w|A) P(w|B) \quad P(w|a) P(w|b) = P(w|A) P(w|B), \quad (2.9)$$

Для каждого слова  $w$  естественно, это выглядит (и остается!) сомнительным предположением, но все же это лучшее, что мы можем сделать с условной вероятностью. Я не буду защищать эту формулировку (2.9) - я покажу ее эквивалентные формулировки.

Например, если мы хотим сказать, что собака относится к щенку, как кошка к котенку, мы ожидаем, что если, например, слово красивый встречается как с собакой, так и с кошкой (вероятно, с разной частотой), затем оно одновременно встречается со словом щенок и котенок по одному и тому же фактору. Похоже, это правда, с коэффициентом два в пользу детеньшей - сравните пары со словами из Google Книг Ngram Viewer (в то время как n-граммы смотрят только на соседние слова, они могут быть своего рода приближением).

Предлагая соотношения для аналогий слов, мы неявно предполагаем, что вероятности слов могут быть факторизованы по разным размерам слова.

Для обсуждаемого случая это будет (2.10):

$$\begin{aligned} P(w|dog) &= f(w|species=dog) \times f(w|age=adult) \times P(w|is\_a\_pet) \quad P(w|puppy) = \\ & f(w|species=dog) \times f(w|age=cub) \times P(w|is\_a\_pet) \quad P(w|cat) = f(w|species=cat) \times \\ & f(w|age=adult) \times P(w|is\_a\_pet) \quad P(w|kitten) = f(w|species=cat) \times f(w|age=cub) \times \\ & P(w|is\_a\_pet) \quad P(w|dog) = f(w|species=dog) \times f(w|age=adult) \times P(w|is\_a\_pet) \\ & P(w|puppy) = f(w|species=dog) \times f(w|age=cub) \times P(w|is\_a\_pet) \quad P(w|cat) = \\ & f(w|species=cat) \times f(w|age=adult) \times P(w|is\_a\_pet) \quad P(w|kitten) = f(w|species=cat) \times \\ & f(w|age=cub) \times P(w|is\_a\_pet), \quad (2.10) \end{aligned}$$

В частности (2.11):

$$\begin{aligned} P(w|dog)P(w|puppy) &= f(w|age=adult) f(w|age=cub) = P(w|cat) P(w|kitten) \\ P(w|dog)P(w|puppy) &= f(w|age=adult) f(w|age=cub) = P(w|cat) P(w|kitten), \quad (2.11) \end{aligned}$$

Как равенство условных отношений вероятностей переводится на слова векторов? Если мы выразим это как взаимную информацию (опять же,  $P(w)$  и логарифмы), мы получим из 2.11:

$$\vec{v} \cdot \vec{w} \cdot \vec{v} \rightarrow \vec{a} - \vec{v} \cdot \vec{w} \cdot \vec{v} \rightarrow \vec{b} = \vec{v} \cdot \vec{w} \cdot \vec{v} \rightarrow \vec{A} - \vec{v} \cdot \vec{w} \cdot \vec{v} \rightarrow \vec{B}, \quad \vec{v} \rightarrow \vec{w} \cdot \vec{v} \rightarrow \vec{a} - \vec{v} \rightarrow \vec{w} \cdot \vec{v} \rightarrow \vec{b} = \vec{v} \rightarrow \vec{w} \cdot \vec{v} \rightarrow \vec{A} - \vec{v} \rightarrow \vec{w} \cdot \vec{v} \rightarrow \vec{B}, \quad (2.12)$$

который совпадает с формулой 2.5

Опять же, если мы хотим, чтобы оно сохранялось для любого слова  $w$ , эта векторная разность должна быть равна нулю.

Мы можем использовать аналогии для значения (например, изменение пола с помощью векторов), грамматики (например, изменение времен) или другие аналогии (например, города в их почтовых индексах). Кажется, что аналогии - это не только вычислительный трюк - мы действительно можем использовать их, чтобы постоянно думать, см.:

- Джордж Лакофф, Марк Джонсон, Метафоры, которыми мы живем (1980)

- и их список концептуальных метафор на английском языке (веб-архив), в частности поиск X is Up, приведенный ниже (рис. 2.5)

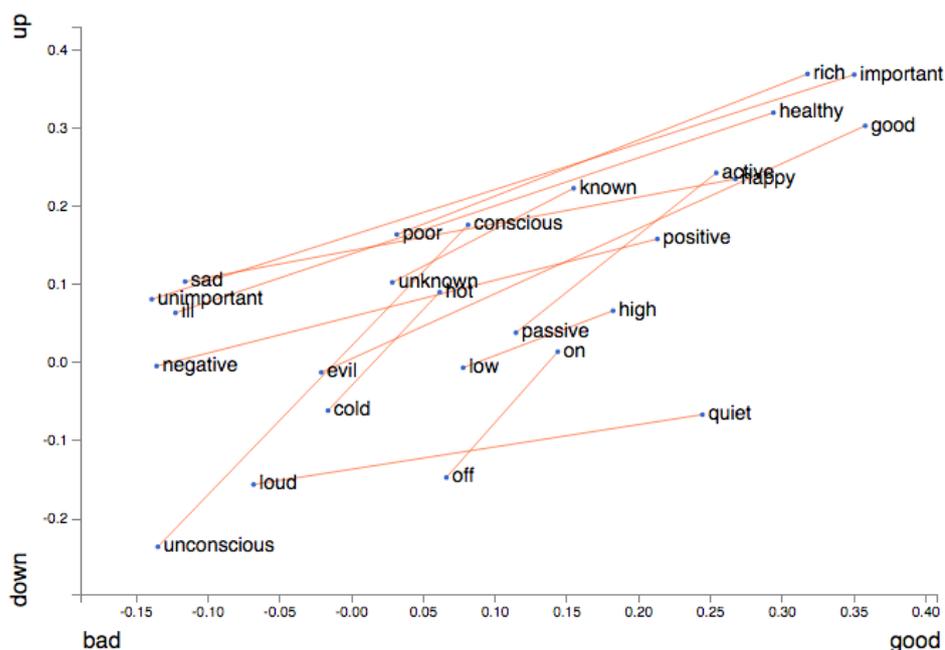


Рисунок 2.5 – Список концептуальных метафор.

## 2.8 Различия и прогнозы

$woman - man = female - male = she - he$ , so  $wo = fe = s$

Разность векторов слов, например:

$$\vec{v} \rightarrow she - \vec{v} \rightarrow he \rightarrow she - \vec{v} \rightarrow he$$

Не являются векторами слов сами по себе. Однако интересно спроецировать слово на эту ось. Мы видим, что проекция:

$$\vec{v} \rightarrow w \cdot (\vec{v} \rightarrow a - \vec{v} \rightarrow b) = \log[P(w|a)] - \log[P(w|b)] \vec{v} \rightarrow w \cdot (\vec{v} \rightarrow a - \vec{v} \rightarrow b) = \log_{f_0}[P(w|a)] - \log_{f_0}[P(w|b)], \quad (2.13)$$

в точности относительное вхождение слова в разных контекстах.

Когда мы хотим посмотреть на общие аспекты слова, более естественно усреднить два вектора, а не брать их сумму. Хотя люди используют его как синонимы, он работает только потому, что косинусное расстояние игнорирует абсолютную длину вектора. Итак, для гендерно-нейтрального местоимения используем  $(\vec{v} \rightarrow she + \vec{v} \rightarrow he) / 2$  ( $\vec{v} \rightarrow she + \vec{v} \rightarrow he$ ) / 2, а не их сумму.

Контурные и другие резкие перепады яркости на изображении (например, связанные с шумом) вносят значительный вклад в высокочастотную часть

## 2.9 Реализация алгоритма с помощью библиотеки gensim

В основе оценки данного алгоритма лежит коэффициент косинуса или еще называют коэффициент Отаи, более подробно про механизмы оценки векторов можете прочитать в статье Томаса Миколова “Efficient Estimation of Word Representations in Vector Space” а также в литературе данной работы, я больше ориентировался именно на решения проблемы [5].

Таким образом реализация данного алгоритма word2vec есть в open-source библиотеке gensim.

Теперь нам нужно много слов, примерно миллион комбинации и как минимум 500 тысяч. Меньше данной цифры алгоритм word2vec показывает себя не стабильно.

В рамках глобальной выборки у нас 1 миллион задач, но эти задачи очень часто совпадают и уникальных названий всего 65 тыс., соответственно комбинация этих предложений недостаточно чтобы построить нормальный модель, таким образом, было принято решения использовать комбинации Национального корпуса русского языка и дампы Википедии до 2018 года, таким образом методами NLP обучили модель понимать слова и подобрать по смыслу похожие значения.

К модели НКРЯ и Вики добавили данные по нашим задачам.

В итоге после генерации получаем модель word2vec векторов к каждому слову, размерность вектора было принято 300 по стандарту Google.

Добавляем модель Word2Vec и загружаем наши уникальные названия задач (рис. 2.6).

```
from gensim.models import Word2Vec, Doc2Vec, KeyedVectors
import nltk
import pandas as pd
import re

df = pd.read_csv('./data/tasks-optimaze.csv')
```

df.head(10)

	id	times	owner_id	name
0	00014a82-970c-11e9-865b-0242ac130007	14400	NaN	Оформление ПП на отм. +18,250. Блок 1, 2
1	0001c00a-07cb-11eb-b5eb-0242c0a87005	205200	2be0ccf2-3608-11ea-9dc1-0242c0a85009	Секции 6,7. Гидравлический расчет
2	000235a6-7eee-11ea-acd4-0242c0a87005	86400	NaN	Составление эскиза ТЗ на Инженерно-Геологическ...
3	000a739d-2919-4f85-8775-506041deebac	86400	NaN	Формирование ведомости монтажной пены
4	000b41b7-6a4f-11ea-acd4-0242c0a87005	32400	NaN	Расстановка НО и компенсаторов 1 блока

Рисунок 2.6 – Добавление данных

После этого удаляем не нужные символы и переводим все в нижний регистр (рис. 2.7).

```
7 newsVac = [nltk.word_tokenize( re.sub(r'\W+', # Search for all non-letters
    ' ', # Replace all non-letters with spaces
    str(title).lower())) for title in titles]

6 [print(re.sub(r'\W+', ' ',str(title).lower())) for title in titles]

оформление пп на отм 18 250 блок 1 2
секции 6 7 гидравлический расчет
составление эскиза тз на инженерно геологические данные
формирование ведомости монтажной пены
расстановка но и компенсаторов 1 блока
формирование рпз
исправление коллизий
моделирование и расчёт системы отопления
узлы и фрагменты по кровле
```

Рисунок 2.7 – Оптимизация данных

Похожие по смыслу на слово создать из 65 тыс. данных получаем следующий результат (рис. 2.8).

```
18 model = Word2Vec(newsVac, size=300, window=5, min_count=1, workers=4)

19 model.save("../initiator-models/initiator-word2vec.model")

25 model.most_similar('создать')

<ipython-input-25-98d70469710a>:1: DeprecationWarning: Call to deprecated `most_similar` (Method will be removed in 4.
model.most_similar('создать')

25 [('документация', 0.9921648502349854),
    ('генподрядчик', 0.9916430115699768),
    ('телефония', 0.989388644695282),
    ('лишнее', 0.9877297282218933),
    ('прошлого', 0.9874616861343384),
    ('сделать', 0.9870604872703552),
    ('главы', 0.9866387248039246),
    ('управление', 0.986609935760498),
    ('электронную', 0.9858256578445435),
    ('располагаться', 0.9857725501060486)]

26 word_vectors = model.wv
```

Рисунок 2.8 – Наибольшее сходство слов к слову ‘создать’

Исходя из рис. 2.8 можно смело сделать вывод что получили не точные значения, как отметили выше при 35 тыс. невозможно получить точные данные, соединяем данный модель с моделями НКРЯ и дампы Википедии 2018.

Также добавил наш проект в Docker и установил Flask, чтобы можно было через http получить доступ к модели. Попробуем сделать запрос и найти максимально схожие слова к слову ‘разум’ (рис. 2.9).

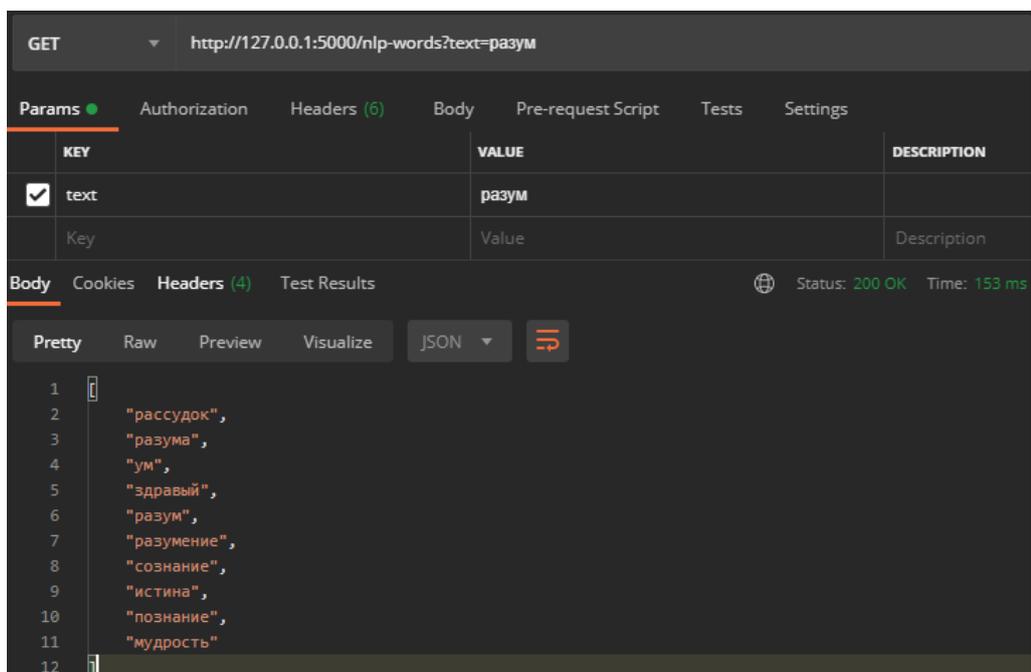


Рисунок 10 – Наибольшее сходство к слову ‘разум’

Исходя из рис.10 можно сказать что первый этап по поиску схожих слов пройден успешно, также в запросе можно обратить внимания на время выполнения 153 ms, на самом деле это очень важно при обработке и дальнейшего анализа.

Часть реализации кода данного алгоритма показано в приложение Б.

## 3 Функция оценки длительности задач

### 3.1 Статистический анализ

Теперь попробуем из комбинации каких-то слов, название задачи (рис. 3.1), найти похожие по смыслу из нашей основной выборки. Предположим, что в массиве NLP у нас только слово "узлы" и загрузим данные по задачам. Также добавим к примеру какого-то сотрудника. (рис. 3.2).

```
In [1]: import pandas as pd
In [2]: df = pd.read_csv('../data/tasks-from-mover.csv')
In [3]: user_id = '2be0ccf2-3608-11ea-9dc1-0242c0a85009'
        nlp_words = [
            "узлы",
        ]
In [4]: df.head(10)
Out[4]:
```

	d	name	times	owner_id
0	00014a82-970c-11e9-865b-0242ac130007	Оформление ПП на отм. +18,250. Блок 1, 2	14400	NaN
1	0001c00a-07cb-11eb-b5eb-0242c0a87005	Секции 6,7. Гидравлический расчет	205200	2be0ccf2-3608-11ea-9dc1-0242c0a85009
2	000235a6-7eee-11ea-acd4-0242c0a87005	Составление эскиза ТЗ на Инженерно-Геологическ...	86400	NaN
3	000a739d-2919-4f85-8775-506041deebac	Формирование ведомости монтажной пены	86400	NaN
4	000b41b7-6a4f-11ea-acd4-0242c0a87005	Расстановка НО и компенсаторов 1 блока	32400	NaN
5	00130f74-f335-11ea-a22c-0242c0a87005	Формирование РПЗ	205200	217
6	0013de98-8a97-11ea-acd4-0242c0a87005	Исправление коллизий	2419200	NaN
7	00153461-6251-11e8-8633-0242ac130004	Моделирование и расчёт системы отопления	6220800	NaN
8	0016de40-05a8-11e9-8b7e-0242ac130005	Узлы и фрагменты по кровле	432000	NaN
9	0019e0a0-ed2a-4494-931a-994b0fe35250	Проверка/сборка альбома	86400	NaN

```
In [5]: # only once!
        titles = df['name'].str.lower()
```

Рисунок 3.1 – Загрузка задач

После этого попробуем получить индексы этих задач и кол-во совпадений слов (т.е. это и есть рейтинг, скажем каждое совпадение это 1 балл, рис. 3.2)

```
In [7]: tasks = []
        tmp = []
In [8]: for idx, words in enumerate(titles):
        for word in str(words).split():
            if word in nlp_words:
                if idx in tmp:
                    for task_idx, [index, _] in enumerate(tasks):
                        if index == idx:
                            tasks[task_idx][1] += 1 # [task_idx][1] - it's a rating!
                else:
                    tmp.append(idx)
                    tasks.append([idx, 1])
In [11]: tasks[1:5]
Out[11]: [[24, 1], [267, 1], [423, 1], [511, 1]]
In [12]: data = []
```

### Рисунок 3.2 – Индексы и рейтинги задач

Теперь получаем из основной выборки эти задачи по индексам и, если к примеру сотрудник, который будет делать данную задачу, уже делал такую задачу получает два балла, т.к. есть большой вероятности что он сделает эту задачу за такой промежуток времени, возможно даже меньше и также можно по другим критериям присваивать баллы. (рис 3.3).

```
In [12]: data = []

In [13]: for idx, rating in tasks:
          owner_id, times = df.loc[idx, ['owner_id', 'times']]
          # space for adding params
          if owner_id == user_id:
              rating += 2

          # END space for adding params

          data.append([rating, times])

In [14]: data[1:5]
Out[14]: [[1, 61200], [1, 18000], [1, 14400], [1, 90000]]
```

Рисунок 3.3 – Рейтинги и время за который можно сделать эту задачу

Выборка - количество проанализированных задач. Выборка формируется из всех аналогичных задач, имеющих в базе программы Инициатор как массив наиболее совпадающих названий. Минимальное количество задач (объем выборки), необходимых для выполнения расчетов - 10. Чем больше объем выборки и чем выше ее качество, тем точнее прогноз [32].

При расчетах прогнозируемой длительности задачи учитывается дисперсия (разброс значений величины от среднего значения). Чем больше разница между минимальным и максимальным значением, т.е. чем больше дисперсия, тем меньше точность прогноза.

Математическое ожидание случайной величины  $X$  (обозначается  $M(X)$  или реже  $E(X)$ ) характеризует среднее значение случайной величины (дискретной или непрерывной). Мат. ожидание - это первый начальный момент заданной СВ [32].

Математическое ожидание относят к так называемым характеристикам положения распределения (к которым также принадлежат мода и медиана). Эта характеристика описывает некое усредненное положение случайной величины на числовой оси. Скажем, если математическое ожидание случайной величины - срока службы лампы, равно 100 часов, то считается, что значения срока службы сосредоточены (с обеих сторон) от этого значения (с тем или иным разбросом, о котором уже говорит дисперсия) [32].

Математическое ожидание дискретной случайной величины  $X$  вычисляется как сумма произведений значений  $x_i$ , которые принимает СВ  $X$ , на соответствующие вероятности  $p_i$ :

$$M(X) = \sum_{i=1}^n x_i \cdot p_i. \quad (3.1)$$

Дисперсия - это мера разброса значений случайной величины  $X$  относительно ее математического ожидания  $M(X)$  (см. как найти математическое ожидание случайной величины). Дисперсия показывает, насколько в среднем значения сосредоточены, сгруппированы около  $M(X)$ : если дисперсия маленькая - значения сравнительно близки друг к другу, если большая - далеки друг от друга (см. примеры нахождения дисперсии ниже) [32].

Если случайная величина описывает физические объекты с некоторой размерностью (метры, секунды, килограммы и т.п.), то дисперсия будет выражаться в квадратных единицах (метры в квадрате, секунды в квадрате и т.п.). Ясно, что это не совсем удобно для анализа, поэтому часто вычисляют также корень из дисперсии - среднеквадратическое отклонение  $\sigma(X) = \sqrt{D(X)}$ , которое имеет ту же размерность, что и исходная величина и также описывает разброс [32].

Еще одно формальное определение дисперсии звучит так: "Дисперсия - это второй центральный момент случайной величины" (напомним, что первый начальный момент - это как раз математическое ожидание) [32].

Дисперсия случайной величины  $X$  вычисляется по следующей формуле (3.2).

$$D(X) = M(X - M(X))^2, \quad (3.2)$$

которую также часто записывают в более удобном для расчетов виде (3.3).

$$D(X) = M(X^2) - (M(X))^2. \quad (3.3)$$

Числовые характеристики дискретной случайной величины  $X$ , которые обычно требуется находить в учебных задачах по теории вероятностей, это математическое ожидание  $M(X)$ , дисперсия  $D(X)$  и среднее квадратическое отклонение  $\sigma(X)$  [32].

### **3.2 Реализация оценки длительности задач**

Нормализуем данные для статистики т.е. рейтинги переводим в доли и находим необходимые величины (рис. 3.4).

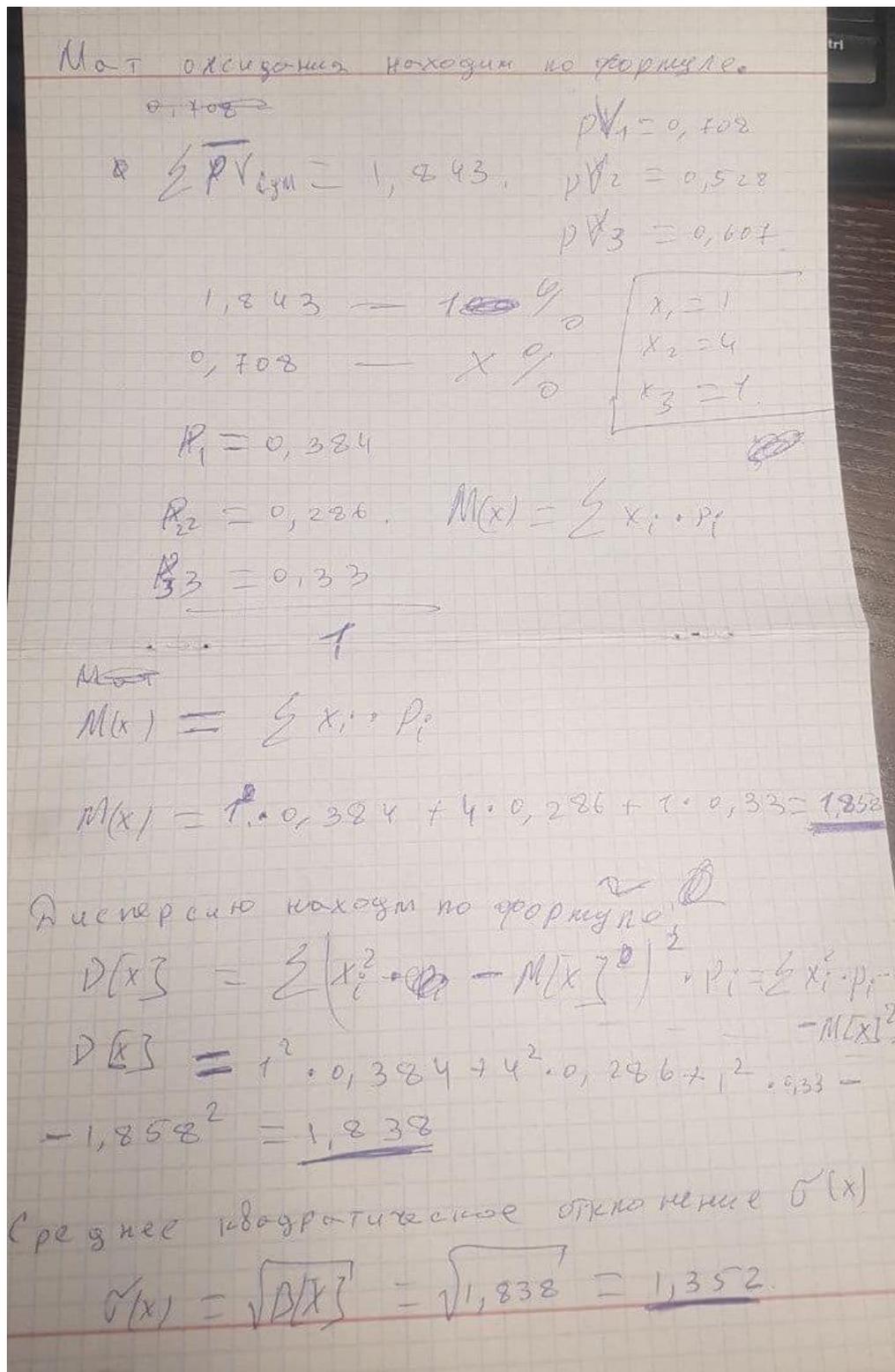


Рисунок 3.4 – Мат ожидание, дисперсия и среднеквадратическое отклонение случайной величины

Реализуем тоже самое в Python (рис. 3.5):

```

42 # Нормализация данных для мат. ожидания
43 data = stat.pv_normalize_tasks(data, total_ratings)
44
45 # Математическое ожидание случайной величины
46 mx = int(stat.math_expectation_x(data))
47
48 # Дисперсия случайной величины
49 dx = int(stat.variance_x(data, mx))
50
51 # Среднеквадратическое отклонение случайной величины для задачи
52 gx = int(math.sqrt(dx))
53
54 return { 'mx': display_time(mx), 'dx': display_time(dx), 'gx': display_time(gx) }
55 # return { 'mx': mx, 'dx': dx, 'gx': gx }

```

Рисунок 3.5 – Реализация в python

### 3.3 Интерфейс функции оценка длительности задачи

Оценка длительности. При наведении курсора появляется всплывающее окно с информацией (рис. 3.6).

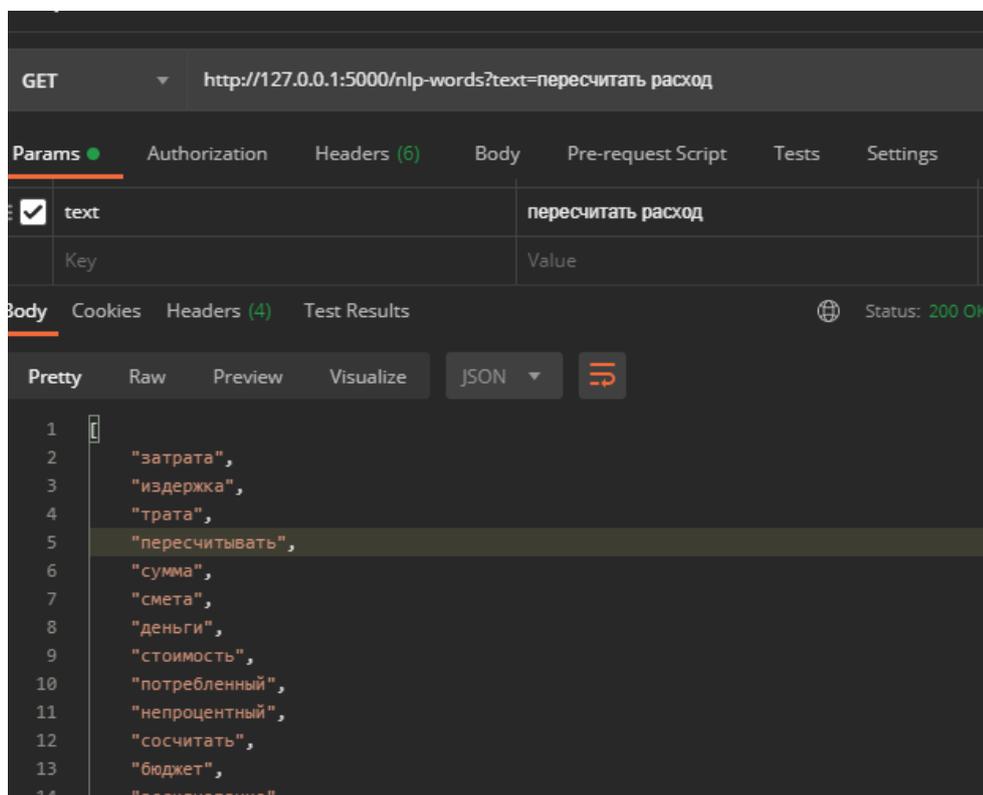


Рисунок 3.6 – Поиск похожих слов к “Пересчитать расходы”

Теперь попробуем найти время за который можно решить такую задачу исходя из дедлайны других задач (рис. 3.7):

The screenshot shows a task management interface for a task titled "Оформление раздела ПБ (часть AP)". The task is marked as "Просрочена" (Overdue) and has a duration of 3d 6h and a cost of 2.50. A popup window titled "Оценка длительности" (Duration Estimation) is displayed over the task details. The popup provides the following information:

- Рекомендуемая длительность: 3д 6ч
- Наиболее схожая: 3д 4ч
- Название: Разработка ТЧ раздела ПБ (часть общие данные, ПЗУ и AP)
- Максимальная длительность: 2н 1д
- Название: Проверка раздела по части AP
- Минимальная длительность: 3ч
- Название: Оформление плана кровли, схема декоративного покрытия кровли
- Объем выборки: 30

The popup also includes a note: "Оценка длительности вычисляется методами NLP исходя из других задач Инициатора." and a list of NLP words: "#оформление, #дизайн, #оформлять, #отделка, #декорирование, #интерьер, #декор, #декора, #оформить, #части, #часть, #половина, #ар, #раздела, #пб".

Рисунок 3.7 – Информация для оценки длительности задач

В карточке указана следующая информация:

Рекомендуемая длительность - оптимальное время для выполнения задачи на основании выполненных расчетов

Наиболее схожая - информация о длительности задачи, название которой максимально совпадает с оцениваемой задачей

Минимальная длительность - наименьшая длительность задачи в выборке. также указано название этой задачи.

Максимальная длительность - наименьшая длительность задачи в выборке. также указано название этой задачи.

Объем выборки - количество задач, использованных для расчетов.

Таким образом если нужно пересчитать расходы алгоритм будет предлагать для данной задачи выделить в среднем 10 часов исходя из всех успешных задач.

### 3.4 Интерфейс самые рискованные задачи

Самые рискованные задачи — это те задачи где параметры риска максимальны (рис. 3.8)

**Риски задач**

Требуется экспертная оценка ситуации и ежедневный контроль

- Вероятность: 5
- Влияние: 5
- Текущее влияние (в стадии): 5
- Рейтинг доверия исполнителя: 5

Данные по рискам обновляется каждый час.

Рисунок 3.8 – Самая рискованная задача

Также в проекте Инициатор сделали вкладку Отчет самых рискованных задач компании (рис. 3.9).

#	Задача	Проект	З/В/П	Просрочено	Ведущий	Исполнитель	Риски
1	еест copy task	Т001	1.0 / 0 / 0 тр.д.	13 д. / 1.0 тр.д.	Реннит А. А.	Реннит А. А.	13.00
2	Petr Ivanov	Т001	0.1 / 0 / 0 тр.д.	132 д. / 0.1 тр.д.	Реннит А. А.	Реннит А. А.	13.00
3	test 1	1	1.0 / 0 / 0 тр.д.	126 д. / 1.0 тр.д.	Триш Д. Р.	Триш Д. Р.	13.00
4	СОСИ 2	1	1.0 / 0 / 0 тр.д.	126 д. / 1.0 тр.д.	Триш Д. Р.	Реннит А. А.	13.00
5	asdsad	1	1.0 / 0 / 0 тр.д.	139 д. / 1.0 тр.д.	Не установлен	Вандам Ж. К.	13.00
6	asdsad	1	1.0 / 0 / 0 тр.д.	49 д. / 1.0 тр.д.	Не установлен	Вандам Ж. К.	13.00
7	test service task	1	1.0 / 0 / 0 тр.д.	48 д. / 1.0 тр.д.	Не установлен	Триш Д. Р.	13.00
8	сервисная задача	1	1.0 / 0 / 0 тр.д.	124 д. / 1.0 тр.д.	Не установлен	Реннит А. А.	13.00
9	test 1	1	1.0 / 0 / 0 тр.д.	5 д. / 1.0 тр.д.	Не установлен	Триш Д. Р.	13.00
10	Копия СОСИ 2	1	1.0 / 0 / 0 тр.д.	0 д. / 1.0 тр.д.	Не установлен	Реннит А. А.	13.00

Рисунок 3.9 – Отчет самые рискованные задачи в компании

## 4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

### 4.1 Предпроектный анализ

#### 4.1.1 Потенциальные потребители результатов исследования

Данная работа посвящена разработке сервиса для анализа рисков управление проектным предприятием веб-приложения Инициатор.

Исходя из особенностей веб-приложения Инициатор, можно судить о круге лиц, которые потенциально будут заинтересованы в разработке. Целевым рынком нынешней разработки являются компании, которые занимаются строительством многоэтажных домов и зданий, также в качестве целевого рынка относятся BIM-проектировщики [31].

Сегментировать рынок услуг можно по степени потребности использования данных расчетов. Результат сегментирования представлены на таблице 4.1.

Таблица 4.1 – Карта сегментирования рынка

		Расчет трудозатрат для проектов	
		Профсоюзная ГП1 Тюмень	Разработка ТХПР
Точность анализа рисков для разных задач проекта Инициатор	Длинные	90%	85%
	Средние	85%	70%
	Мелкие	65%	45%

Исходя из проведенной карты можно заключить, что чем больше область применения задач встречается в проекте тем более точно можно предсказать трудозатраты.

## 4.1.2 Анализ конкурентных решений

Данная разработанное решение для анализа рисков является уникальной в своём роде, так как она содержит особенности и факторы проекта Инициатор которые влияют на прогнозируемый модель, также можно её модернизировать под каждого конкретного заказчика, но по стандарту анализируемы данные взяты из всех компаний. В качестве конкурирующих решения были выбраны следующие продукты:

1. App Annie — это платформа номер один для анализа данных о мобильных приложениях. Она объединяет в себе всё необходимое для издателей и разработчиков, помогая им понять требования рынка для создания и продажи мобильных приложений, а также для вложения средств в их разработку.

Компания, главный офис которой находится в Сан-Франциско, насчитывает более 425 сотрудников в 15 подразделениях по всему миру. Компания привлекла \$156 млн от 11 инвесторов, включая Sequoia Capital.

Клиенты App Annie — 90% из топ-100 публических компаний и свыше 700 тысяч зарегистрированных пользователей, контролирующих один миллион мобильных приложений.

Среди клиентов App Annie такие компании, как Electronic Arts, Google, LinkedIn, Microsoft, Nestle, Samsung, Tencent и многие другие. Со всеми ведется аналитическая работа в онлайн. Благодаря этому владельцы могут не только получить данные об использовании своих приложений, но и отслеживать деятельность конкурентов и происходящее в индустрии в целом. [18]

2. Программа Cloudera . Калифорнийская компания Cloudera создала собственный инструментарий для работы с большими данными на основе Apache Hadoop (открытое программное обеспечение). Этот сервис предлагает услуги по управлению данными и аналитике на основе технологии, позволяющей в одном месте хранить, обрабатывать и анализировать данные. В компании работает 1400 человек.

Компания привлекла чуть больше \$1 млрд от нескольких десятков инвесторов и собралась поднять еще \$200 млн на IPO в ближайшие месяцы с

ожидаемой оценкой в \$ 4 млрд. Интересно, что Accel Partners инвестировал \$5 млн в 2009 году, а Intel Capital — \$740 млн в 2014 году.

Выручка компании за прошлый год \$261 млн, что на 57% больше, чем годом ранее. Как и любой растущий и перспективный стартап они предпочитают не накапливать прибыль, а тратить деньги на развитие, поэтому компания получила убыток в \$187 млн.

Большинство клиентов — крупные компании из топ-500 мировых брендов. С «бумом» на создание технологий для Big Data, Cloudera неоднократно отмечена как одна из самых многообещающих компаний, способных решать задачи соответствующего класса. [19]

3. PricewaterhouseCoopers — международная сеть компаний, предлагающих услуги в области консалтинга и аудита. Под «PricewaterhouseCoopers» понимаются компании, входящие в сеть компаний PricewaterhouseCoopers International Limited, каждая из которых является самостоятельным юридическим лицом. Бренд существует на протяжении более 160 лет и входит в так называемую большую четвёрку аудиторских компаний. Штаб-квартира сети — в Лондоне.

Управляйте рисками и уверенно используйте возможности для роста и создания конкурентного преимущества

#### 1. Технические критерии оценки ресурсоэффективности

##### а. Удобство эксплуатации

Для каждой системы нужно настроить экосистему передачи данных, чтобы легче через разные протоколы сервера общались, наилучший способ эксплуатации — это завернуть всю инфраструктуру наших микросервисов в docker образ который позволит легко разворачивать на любой машине.

##### б. Надёжность

Надёжность системы от DoS-атаки и передачи данных между другими сервисами проекта.

##### с. Требуемые ресурсы

Ресурсы, которые необходимы для быстрого и успешного анализа рисков

и проход через сложные нейросетевые модели.

d. Функциональность

Способность, обновлять и добавлять новые возможности и изменять поведения алгоритма в зависимости от выбранной области компании (к примеру, если область выбранной компании не совпадает с другими проектами Инициатор, чтобы легко и функционально, мы могли построить и прогнозировать риски для этой выбранной компании).

e. Простота эксплуатации

На разных серверах Инициатора возможность легко, быстрого развертывания и эксплуатации.

2. Экономические критерии оценки эффективности

a. Обслуживание

Минимальные затраты на обслуживание системы.

b. Срок разработки

Возможность разбиение работ и областей по частям, чтобы уменьшить срок.

Экспертная оценка основных технических характеристик данных продуктов представлена в таблице 4.2.

Таблица 4.2. Оценочная карта сравнения технических решений

№	Критерии оценки	Вес	Баллы				Конкурентоспособность			
			Бф	Бк1	Бк2	Бк3	Кф	К1	К2	К3
	1	2	3	4	5	6	7	8	9	10
<b>Технические критерии оценки ресурсоэффективности</b>										
1	Удобство эксплуатации	0,18	5	1	2	4	0,93	0,19	0,37	0,37
2	Надёжность	0,12	5	5	4	5	0,44	0,56	0,44	0,44
3	Требуемые ресурсы	0,13	4	2	3	4	0,56	0,22	0,33	0,44
4	Функциональность	0,17	5	2	3	3	0,93	0,19	0,56	0,37
5	Простота эксплуатации	0,15	5	3	2	2	0,74	0,44	0,30	0,15
<b>Экономические критерии оценки эффективности</b>										
6	Обслуживание	0,19	5	2	3	4	0,74	0,37	0,14	0,74
7	Срок разработки	0,07	4	5	3	3	0,30	0,37	0,22	0,22
	<b>Итого</b>	1	34	20	20	25	4,63	2,33	2,36	2,74

Исходя из проведенного анализа можно заключить, что уязвимость конкурентных технологических решений связана, прежде всего с удобством эксплуатации и функциональную мощность.

Однако стоит ценить наличие проникновения на рынок, хотя в данном случае это вообще не играет никакой роли.

Наиболее сильным конкурентом можно считать программа PricewaterhouseCoopers, однако она рассчитана в области консалтинга и аудита, спектр, анализ и скорость его будет в десятки раз ниже нашего сервиса. В связи с этим было принято решения не использовать какие-то сторонние сервисы, разрабатывать собственную модель учитывая все факторы внутри компании и приватность данных проекта.

#### 4.1.3 SWOT-анализ

SWOT-анализ применяют для исследования внешней и внутренней среды проекта. Матрица составляется на основе анализа рынка и конкурентных технических решений, и показывает сильные и слабые стороны проекта, возможности и угрозы для разработки. Первый этап заключается в описании сильных и слабых сторон проекта, в выявлении возможностей и угроз для реализации проекта, которые проявились или могут появиться в его внешней среде. Матрица SWOT представлена в таблице 4.3.

Таблица 4.3 – SWOT-анализ

	<b>Сильные стороны</b>	<b>Слабые стороны</b>
	Си1. Широкий функционал	Сл1. Необходимость расширения списка задач или сущностей для анализа и более точного предсказания
	Си2. Удобный интерфейс, наличие использования на разных устройствах	Сл2. Необходимость изменения системы для каждого заказчика
	Си3. Система написана самостоятельно	Сл3. Необходимость локального размещения
<b>Возможности</b>		

В1. Доработка в связи с пожеланиями	В1Си3. Система написана нами, соответственно расширить её легко В1Си1. Система уже имеет широкий функционал, поэтому доработать просто	В1Сл2. Заказчик может попросить сложную модернизацию
В2. Нет необходимости в серверном хранении данных	В2Си2. Заранее хорошо проработанный интерфейс позволит не заниматься постоянным обновлением систем у клиентов	В2Сл3. Могут возникнуть проблемы в начале размещения в связи с большими данными для анализа
В3. Упаковка системы в docker	В3Си1. Docker и Kubernetes позволяет удобно компоновать составляющие системы и подстраиваться под разные нагрузки серверов	В3Сл3. Docker и Kubernetes позволяет проще развёртывать систему не затрагивая основные настройки
<b>Угрозы</b>		
У1. Сбои, потери пакетов	У1Си3. Система самописная, соответственно ошибки будет легко исправить и также легко разворачивать	У1Сл3. Так как система расположена на сервере проекта Инициатор, то могут возникнуть проблемы с утерей памяти на сервере, для избегания необходимо проводить тщательное тестирование.
У2. Проблемы при разворачивании	У2Си3. Так как мы хорошо разбираемся в системе и используем для этого Docker то в целом риск возникновения проблем минимальные	У2Сл3. Серверные проблемы намного сложнее отлаживать, поэтому можно записывать логи по всем операциям на сервере и попросить в случае проблемы эти данные для исследования.
У3. Внесение новых модели	У3У4Си3. Это будет легко сделать, так как мы хорошо понимаем ПО	У3Сл1. Возможна долгая работа над расширением, что увеличит сроки, поэтому их необходимо закладывать на этапе составления технического задания
У4. Медленная работа системы		У4Сл3. Скорость работы может зависеть от мощностей сервера. Для избежания угрозы нужно тестировать на слабых мощностях.

В рамках SWOT анализа были рассмотрены по три слабых и сильных сторон проекта, а также три возможности и четыре угрозы. Указаны стратегии, которые будут ликвидировать слабые стороны разработки и бороться с угрозами.

Второй этап состоит в выявлении соответствия сильных и слабых сторон научно-исследовательского проекта внешним условиям окружающей среды. Это соответствие или несоответствие должны помочь выявить степень необходимости проведения стратегических изменений.

Соотношения параметров представлены в таблице 4.4.

Таблица 4 – Интерактивная матрица проекта

Сильные стороны проекта				
Возможности		Си1	Си2	Си3
	В1	+		+
	В2		+	
	В3	+		
Слабые стороны проекта				
Возможности		Сл1	Сл2	Сл3
	В1		+	
	В2			+
	В3			+
Сильные стороны проекта				
Угрозы		Си1	Си2	Си3
	У1			+
	У2			+
	У3			+
	У4			+
Слабые стороны проекта				
Угрозы		Сл1	Сл2	Сл3
	У1			+
	У2			+
	У3	+		
	У4			+

Самым большим преимуществом данной системы является то, что она написана самостоятельно, а недостатком то, что необходимо оптимизировать модели под разные компании и проекты.

#### 4.1.4 Оценка готовности проекта к коммерциализации

На какой бы стадии жизненного цикла не находилась научная разработка полезно оценить степень ее готовности к коммерциализации и выяснить уровень собственных знаний для ее проведения (или завершения).

Для этого заполним форму в таблице 5, содержащую показатели степени проработанности проекта с позиции коммерциализации и компетенциям разработчика научного проекта.

Таблица 4.5. Оценка степени готовности проекта к коммерциализации

№ п/п	Наименование	Ком-мерция	Компе-тенции
1	Определен имеющийся научно-технический задел	3	3
2	Определены перспективные направления коммерциализации научно-технического задела	5	4
3	Определены отрасли и технологии для предложения	4	3
4	Определена форма для представления на рынок	4	4
5	Определены авторы и осуществлена охрана их прав	1	1
6	Оценена стоимость интеллектуальной собственности	1	1
7	Проведены маркетинговые исследования рынков сбыта	1	1
8	Разработан бизнес-план коммерциализации разработки	1	1
9	Определены пути продвижения разработки на рынок	3	4
10	Разработана стратегия (форма) реализации разработки	4	4
11	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	1	1
12	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	2	2
13	Проработаны вопросы финансирования коммерциализации научной разработки	5	5
14	Имеется команда для коммерциализации разработки	4	4
15	Проработан механизм реализации научного проекта	4	5
	<b>ИТОГО БАЛЛОВ</b>	43	43

Итоговые значения проработанности научного проекта и знания у разработчика 43, что говорит о том, что некоторые аспекты проекта практически не были проработаны, а именно охрана прав, а также анализ рынков сбыта. Так как работа выполняется для компании и поставляется в рамках аналитической системы для анализа рисков компаний, то это и не требуется.

#### 4.1.5 Цели и результаты проекта

Перед определением целей необходимо перечислить заинтересованные стороны проекта. Информация по заинтересованным сторонам представлена в таблице 4.6:

Таблица 4.6. Заинтересованные стороны проекта

<b>Заинтересованные стороны</b>	<b>Ожидания</b>
Пользователь	Простота в использовании программного продукта
Компания пользователя	Облегчен процесс контроля за проектом компании
Разработчик	Получение заработной платы
Компания разработчика	Получение прибыли с полноценного комплекса продукта Инициатор
Научный руководитель, студент	Выполненная выпускная квалификационная работа

Цели и результат проекта представлены в таблице 7.

Таблица 4.7. Цели и результат проекта

Цели проекта	<ul style="list-style-type: none"><li>• Изучить систему</li><li>• Настроить сервер</li><li>• Выбрать средства разработки</li><li>• Развернуть и настроить инфраструктуру</li><li>• Установить и настроить необходимые сервисы</li><li>• Создать макет страниц сайта</li><li>• Осуществить верстку клиентской части (frontend)</li><li>• Реализовать код для серверной части (backend)</li></ul>
Ожидаемые результаты	<ul style="list-style-type: none"><li>• Успешное внедрение разработки в комплекс</li><li>• Сдана выпускная квалификационная работа</li></ul>
Критерии приёмки	<ul style="list-style-type: none"><li>• Успешное тестирование функционала в соответствии с функциональным требованием</li></ul>
Требования к результату проекта	<ul style="list-style-type: none"><li>• Выполнены все пункты функционального требования и требования к пользовательскому интерфейсу.</li><li>• Разработанный функционал полностью соответствует проектным решениям.</li></ul>

## 4.2 Планирование научно-исследовательских работ

### 4.2.1 План проекта

Диаграмма Ганта – это тип столбчатых диаграмм (гистограмм), который используется для иллюстрации календарного плана проекта, на котором работы по теме представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ.

График строится в виде таблицы 9 с разбивкой по декадам (10 дней) за период времени выполнения научного проекта. При этом работы на графике следует выделить различной штриховкой в зависимости от исполнителей, ответственных за ту или иную работу.

Таблица 4.8. Календарный план-график проведения НИОКР по теме

К о д	Вид	Исп	Тк, ч. / Т раб	Продолжительность																
				Янв.			Февр.			Март			Апр.			Май				
				1	2	3	1	2	3	1	2	3	1	2	3	1	2	3		
1	Выбор направления исследования	Р, С	20 / 2.5	■																
2	Описание требований	С	15 / 2		■															
3	Патентный поиск	С	5 / 1		■															
4	Составление технического задания	С	15 / 2			■														
5	Изучение литературы	С	20 / 2.5				■													
6	Изучить систему	С	50 / 6					■	■											
7	Настроить сервер	С	12 / 1							■										
8	Выбрать средства разработки	С	15 / 1							■										
9	Развернуть и настроить инфраструктуру	С	25 / 3								■									



## 4.2.2 Бюджет научного исследования

При планировании бюджета научного исследования должно быть обеспечено полное и достоверное отражение всех видов планируемых расходов, необходимых для его выполнения. В процессе формирования бюджета, планируемые затраты группируются по статьям, представленным в таблице.

### 4.2.2.1 Специальное оборудование для научных работ

В данную статью включают все затраты, связанные с приобретением специального оборудования (приборов, контрольно-измерительной аппаратуры, стендов, устройств и механизмов), необходимого для проведения работ по конкретной теме. Определение стоимости спецоборудования производится по действующим прейскурантам, а в ряде случаев по договорной цене.

Для разработки нынешней системы требуется обычное оборудование в виде персонального компьютера и купленных серверов на время жизненного цикла программного продукта. Среда и средство разработки, программный софт и другие комплектующие, нужные для разработки, распространяются бесплатно и не требуют дополнительных затрат (таблица 4.9).

Таблица 4.9 – Расчет затрат на «Спецоборудование для научных работ»

№	Наименование оборудования	Кол-во	Цена, руб.	Стоимость, руб.
1.	Персональный компьютер	1	-	-
2.	Среда разработки WebStorm +PyCharm	1	5800	5800
3.	AWS EC2 – полная комплектация	1	24 458	24 458
Итого				30258

#### 4.2.2.2 Основная заработная плата

В настоящую статью включается основная заработная плата научных и инженерно-технических работников, рабочих макетных мастерских и опытных производств, непосредственно участвующих в выполнении работ по данной теме.

Должность руководителя – доцент, к.т.н. – 35 111.5 рублей в месяц

Должность инженера – студент – 22 695,68 рублей в месяц

Расчет основной заработной платы сводится в таблице 4.10.

Таблица 4.10 – Расчет основной заработной платы

№ п/п	Наименование этапов	Исп	Трудоемкость, чел.-дн.	Зарплата, чел.-дн., руб	Всего руб.
1		Р	4	35 111.5	35 111.5
2		С	39	22 695.68	22 695.68

$$C_{зп} = Z_{осн} + Z_{доп},$$

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

$Z_{доп}$  – дополнительная заработная плата.

Основная заработная плата  $Z_{осн}$  руководителя рассчитывается по следующей формуле::

$$Z_m = Z_{дн} \cdot T_{раб}$$

где  $T_{раб}$  – продолжительность работ, выполняемых научно-техническим работником, рабочие дни. (таблица 4.13);

$Z_{дн}$  – среднедневная заработная плата работника, руб.

Значит, для руководителя:

$$Z_m = 35\,111.5 \cdot 1,3 = 45\,645 \text{ рублей}$$

Для инженера:

$$Z_m = 22\,695.68 \cdot 1,3 = 29\,504 \text{ рублей}$$

Среднедневная заработная плата рассчитывается по формуле:

$$Z_{дн} = (Z_m \cdot M) / F_d$$

где  $Z_m$  – месячный должностной оклад работника, руб

$M$  – количество месяцев работы без отпуска в течение года:

при отпуске в 45 раб. дней  $M=10,4$  месяца, 6-дневная неделя;

Фд – действительный годовой фонд рабочего времени научно-технического персонала (в рабочих днях) (таблица 14). Тогда,

Для руководителя:

$$З_{\text{дн}} = \frac{45\,645 * 10,4}{254} = 1\,869 \text{ рублей}$$

Для инженера:

$$З_{\text{дн}} = \frac{22\,695.68 * 11,2}{217} = 1\,171 \text{ рублей}$$

Баланс рабочего времени представлен в таблице 4.11.

Таблица 4.11 – Баланс рабочего времени

<b>Показатели рабочего времени</b>	<b>Руководитель</b>	<b>Инженер</b>
Календарное число дней	365	365
Количество нерабочих дней		
- выходные дни	52	82
- праздничные дни	11	14
Потери рабочего времени		
- отпуск	56	24
- невыходы по болезни	–	–
<b>Действительный годовой фонд рабочего времени</b>	<b>254</b>	<b>217</b>

Таблица 4.12 – Результаты расчета основной заработной платы

Исполнители	З <sub>б</sub> , руб.	k <sub>р</sub>	З <sub>м</sub> , руб	З <sub>дн</sub> , руб.	T <sub>р</sub> , дн.	З <sub>осн</sub> , руб.
Руководитель	35 111.5	1.3	45 645	1 869	4	7 476
Инженер	22 695.68	1.3	29 504	1 171	39	45 669
Итого по статье З <sub>осн</sub> :			53 145			

### 4.2.2.3 Дополнительная заработная плата научно-производственного персонала

Дополнительная заработная плата включает оплату за непроработанное время (очередной и учебный отпуск, выполнение государственных обязанностей, выплата вознаграждений за выслугу лет и т.п.) и рассчитывается исходя из 10-15% от основной заработной платы, работников, непосредственно участвующих в выполнении темы:

$$Z_{\text{доп}} = k_{\text{доп}} * Z_{\text{осн}} \quad (4.9)$$

где  $Z_{\text{доп}}$  – дополнительная заработная плата, руб.;

$k_{\text{доп}}$  – коэффициент дополнительной зарплаты ( $k_{\text{доп}} = 0,1$ );

$Z_{\text{осн}}$  – основная заработная плата, руб.

Для руководителя:

$$Z_{\text{доп}} = 7\,476 * 0,1 = 747,6 \text{ рублей}$$

Для инженера:

$$Z_{\text{доп}} = 45\,669 * 0,1 = 4\,566,9 \text{ рублей}$$

В таблице 14 приведен расчёт основной и дополнительной заработной платы.

Таблица 4.13 – Заработная плата исполнителей ВКР, руб

Заработная плата	Руководитель	Инженер
Основная зарплата	7 476	45 669
Дополнительная зарплата	747.6	4 566.9
Зарплата исполнителя	8 223.6	50 235.9
Итого	58 459	

#### 4.2.2.4 Отчисления на социальные нужды

Статья включает в себя отчисления во внебюджетные фонды.

$$C_{\text{внеб}} = k_{\text{внеб}} \cdot (Z_{\text{осн}} + Z_{\text{доп}}) = 0,3 \cdot (7\,476 + 747,6) = 2\,467,1 \text{ руб.}$$

Для инженера:

$$C_{\text{внеб}} = k_{\text{внеб}} \cdot (Z_{\text{осн}} + Z_{\text{доп}}) = 0,3 \cdot (45\,669 + 4\,566,9) = 15\,070,7 \text{ руб.}$$

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

Для руководителя – 2 467,1 руб.

Для инженера – 15 070,7 руб.

Итого 17 538 рублей

#### 4.2.2.5 Накладные расходы

В эту статью относятся расходы по содержанию, эксплуатации и ремонту оборудования, производственного инструмента и инвентаря, зданий, сооружений и др. В расчетах эти расходы принимаются в размере 16 % от суммы основной заработной платы научно-производственного персонала данной научно-технической организации.

Накладные расходы составляют 16% от суммы основной и дополнительной заработной платы, работников, непосредственно участвующих в выполнение темы.

Расчет накладных расходов ведется по следующей формуле:

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

где  $k_{\text{накл}}$  – коэффициент накладных расходов.

$$C_{\text{накл}} = 0,16 * 58\,459 = 9\,353 \text{ руб.}$$

#### 4.2.2.6 Формирование бюджета затрат научно-исследовательского проекта

Рассчитанная величина затрат научно-исследовательской работы является основой для формирования бюджета затрат проекта, который при формировании договора с заказчиком защищается научной организацией в качестве нижнего предела затрат на разработку научно-технической продукции.

Таблица 4.14 – Бюджет затрат НИИ

Сырье	Оборудован	Осн зп	Доп зп	Накл рас	Отчисл на соц	Итого
-	30 258	53 145	5 314	9 353	17 538	<b>115 608</b>

В результате было получено, что бюджет на разработку НИИ составит 115 608 руб.

### 4.3 Оценка сравнительной эффективности исследования

#### 4.3.1 Интегральный показатель финансовой эффективности

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:

$$I_{\text{финр}}^{\text{исп}i} = \frac{\Phi_{pi}}{\Phi_{\text{max}}},$$

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).

$$\begin{aligned} I_{\text{ф}}^{\text{p}} &= \frac{\Phi_i^{\text{p}}}{\Phi_{\text{max}}} = \frac{115\,608}{150\,000} = 0.77 \\ I_{\text{ф}}^{\text{a}} &= \frac{\Phi_i^{\text{a}}}{\Phi_{\text{max}}} = \frac{130\,565}{150\,000} = 0.87 \\ I_{\text{ф}}^{\text{b}} &= \frac{\Phi_i^{\text{a}}}{\Phi_{\text{max}}} = \frac{118\,500}{150\,000} = 0.79 \end{aligned}$$

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).

### 4.3.2 Интегральный показатель ресурсоэффективности

Можно определить следующим образом:

$$I_{pi} = \sum a_i \cdot b_i,$$

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

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

$b_i$  – бальная оценка  $i$ -го варианта исполнения разработки;

$n$  – число параметров сравнения.

Расчет приведен в форме таблице 4.15.

Таблица 4.15. Сравнение характеристик вариантов исполнения проекта

ПО Критерии	Вес	Текущий	Аналог 1	Аналог 2
Удобство эксплуатации	0.18	5	5	3
Надёжность	0.12	5	5	5
Требуемые ресурсы	0.13	5	5	4
Функциональность	0.17	5	3	5
Простота эксплуатации	0.14	5	4	5
Обслуживание	0.18	4	3	4
Срок разработки	0.09	4	5	5
ИТОГО	1	33	30	31
$I_T^x$		4.78	4.21	4.64
$I_\Phi^x$		0.77	0.87	0.79
$I_{финр}^x$		6.3	4.83	5.87

$$\mathcal{E}_{cp} = \frac{I_\Phi^p}{I_\Phi^a} = \frac{6.3}{5.87} = 1.07$$

Таблица 4.16. Сравнительная эффективность разработки

№	Показатели	Разработка	Аналог 2
1	Интегральный финансовый показатель разработки	0.77	0.79
2	Интегральный показатель ресурсоэффективности разработки	4.78	4.64
3	Интегральный показатель эффективности	0.77	0.79
4	Сравнительная эффективность вариантов исполнения	1.07	

Сравнение значений интегральных показателей эффективности позволяет судить о приемлемости существующего варианта решения поставленной в магистерской диссертации технической задачи с позиции финансовой и ресурсной эффективности.

Таким образом, мы определили сравнительную эффективность проекта равную 1.07. Сравнение значений интегральных показателей эффективности позволяет судить о приемлемости существующего варианта решения поставленной в магистерской диссертации технической задачи с позиции финансовой и ресурсной эффективности.

#### **4.4 Вывод по разделу**

В ходе данной работе были рассмотрены потенциальные потребители результатов исследования, так же для анализа конкурентных технических решений. С позиции ресурсоэффективности и ресурсосбережения была составлена оценочная карта сравнения конкурентных технических решений, по результату которой разрабатываемая система имеет лучшие качества.

Далее был сформирован SWOT-анализ, в котором балы описаны сильные и слабые стороны проекта, в выявлении возможностей и угроз для реализации проекта, для выявления соответствия и несоответствия была составлена интерактивная матрица проекта.

Проведена оценка готовности проекта к коммерциализации, которая показала, что перспективность разработки средняя.

В рамках процессов инициации определены внутренние и внешние заинтересованные стороны проекта с их ожиданиями от проекта, цели и результат проекта.

План проекта представлен на диаграмме Ганта, из которого видно какой исполнитель (студент или руководитель) какой вид работ осуществлял и в течении какого количества дней.

В бюджет инженерно-технического проекта занесена стоимость оборудования, которая составила 30 258 рублей. Была рассчитана основная и дополнительная заработная плата исполнителей проекта, сумма которых составила 58 459 рублей. Вместе с отчислениями на социальные нужды и накладными расходами бюджет проекта получился равным 115 608 рублей.

Был рассчитан интегральный финансовый показатель в ходе оценки бюджета затрат трех вариантов исполнения. Наибольший интегральный показатель реализации задачи равен 0.77. Значение меньше единицы, а значит полученная величина показывает численное удешевление стоимости в размах. Рассчитан интегральный показатель ресурсоэффективности для трех вариантов исполнения 4.78, 4.21. и 4.64. Расчет интегрального показателя эффективности для разработки и аналога позволил сравнить их и рассчитать сравнительную

эффективность разработки, которая говорит нам о приемлемости существующего варианта решения поставленной в магистерской диссертации технической задачи с позиции финансовой и ресурсной эффективности.

## **5 Социальная ответственность**

Разработанный проект предназначен для предсказания рисков компаний исходя из факторов, которые влияют на проекты и задачи проекта.

В данном разделе будут рассмотрены вредные и опасные факторы, которые оказывают влияние на рабочую деятельность персонала, рассмотрены воздействия разрабатываемой программы на окружающую среду, правовые и организационные вопросы, мероприятия в чрезвычайных ситуациях.

Работа выполнялась в десятом корпусе ТПУ (четвертый этаж) в отделении информационных технологий. Аудитория 402 была помещением выполнения исследования. Дисплей - размещен в зоне в центре, клавиатура - перед дисплеем; системный блок размещен слева, принтер находится справа.

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

Согласно 91 статье ТК РФ нормальная продолжительность рабочего времени не может превышать 40 часов в неделю. Работодатель обязан вести учет времени, фактически отработанного каждым работником. Каждый работник имеет право на выходные дни, то есть периоды еженедельного непрерывного отдыха. Согласно 110 статье ТК РФ продолжительность еженедельного непрерывного отдыха не может быть менее 42 часов.

Нормативными документами, регламентирующими деятельность работника с использованием ПЭВМ, являются СП 2.2.3670-20. Гигиенические требования к ПЭВМ и организации труда и Инструкция по охране труда при работе с ПК. Нормативным документом установлено, что при работе с компьютером, пользователь обязан делать перерыв в размере 15 минут каждый час непрерывной работы.

Во время регламентированных перерывов целесообразно выполнять комплексы упражнений и осуществлять проветривание помещения. Не рекомендуется работать за компьютером более 6 часов за смену. Для того чтобы ПЭВМ соответствовали нормам, осуществляется производственный контроль и надзор внутри предприятия-производителя. Эксплуатирующие предприятия

также следят за характеристиками используемой аппаратуры.

### **5.1.1 Специальные правовые нормы трудового законодательства**

В данном разделе рассмотрены основные нормативные документы, выполнение которых необходимо для безопасного исполнения рабочих обязанностей диспетчером, который проводит большую часть времени за компьютером.

Прядок организации рабочего процесса для лиц, труд которых связан с компьютерной техникой регулируют:

- «Трудовой кодекс Российской Федерации» от 30.12.2001 № 197-ФЗ (ред. от 30.04.2021) (с изм. и доп., вступ. в силу с 01.05.2021).

- СП 2.1.3678-20 «Санитарно-эпидемиологические требования к эксплуатации помещений, зданий, сооружений, оборудования и транспорта, а также условиям деятельности хозяйствующих субъектов, осуществляющих продажу товаров, выполнение работ или оказание услуг» (постановление Главного государственного санитарного врача РФ от 24.12.2020 г. №44);

- «ТОИ Р-45-084-01. Типовая инструкция по охране труда при работе на персональном компьютере» (утв. Приказом Минсвязи РФ от 02.07.2001 N 162) регламентирует вопрос работы с компьютером. Согласно данному документу, без перерыва работник может находиться за монитором компьютера не более двух часов.

## **5.1.2 Профессиональная социальная безопасность.**

### **5.1.2.1 Анализ вредных и опасных факторов, которые могут возникнуть на рабочем месте при разработке**

Требования к организации рабочих мест пользователей:

Конструкция рабочей мебели (рабочий стол, кресло, подставка для ног) должна обеспечивать возможность индивидуальной регулировки соответственно росту пользователя и создавать удобную позу для работы.

Основное внимание необходимо уделить делению на рабочие зоны так, как указано на рисунках 5.1 и 5.2, чтобы каждый сотрудник мог работать максимально эффективно, как отдельно так при необходимости работать в команде. Эргономика рабочего места напрямую влияет на производительность труда [35, 36].

Согласно требованиям:

- расстояние между рабочими столами с видеомониторами должно быть не менее 2,0 м, а расстояние между боковыми поверхностями видеомониторов – не менее 1,2 м;
- экран видеомонитора должен находиться от глаз пользователя на расстоянии от 600 до 700 мм, но не ближе 500 мм с учетом размеров алфавитно-цифровых знаков и символов;
- конструкция рабочего стола должна обеспечивать оптимальное размещение используемого оборудования с учетом характера выполняемой работы;
- поверхность рабочего стола должна иметь коэффициент отражения от 0,5 до 0,7.

На рисунках 5.1 и 5.2 схематично представлены требования к рабочему месту и пространству вокруг работника.



Рисунок 5.1 – Организация рабочего места

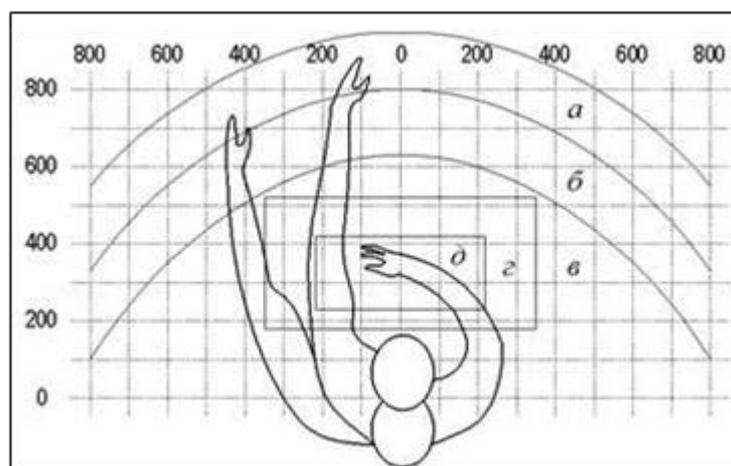


Рисунок 5.2 – Эргономические требования: а – зона максимальной досягаемости; б – зона досягаемости пальцев при вытянутой руке; в – зона легкой досягаемости ладони; г – оптимальной пространство для грубой ручной работы; д – оптимальное пространство для точной ручной работы

Оптимальное размещение предметов труда и документации в зонах досягаемости согласно [11]:

- дисплей размещается в зоне «а» (в центре);
- системный блок размещается в предусмотренной нише стола;
- клавиатура – в зоне «г/д»;

- «мышь» – в зоне «в» справа;
- документация, необходимая при работе – в зоне легкой досягаемости ладони – «б», а в выдвижных ящиках стола – редко используемая литература.

## 5.2 Производственная безопасность

Для обеспечения безопасности во время эксплуатации и разработки программы, необходимо провести анализ вредных и опасных воздействий на человека, которые могут возникать при разработке или эксплуатации проекта.

В таблице 5.1 приведены возможные и опасные вредные факторы:

Таблица 5.1 – Возможные и опасные вредные факторы.

Факторы (ГОСТ-12.0.003-2015)	Этапы работы	Нормативные документы
	Разработка	
1. Повышенная или пониженная температура и влажность воздуха рабочей	+	СанПиН 1.2.3685-21 «Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания»
2. Недостаточная освещенность рабочей зоны	+	СанПиН 1.2.3685-21 «Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания»
3. Повышенный уровень шума на рабочем месте	+	СанПиН 1.2.3685-21 «Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания»; ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности.

4. Опасные и вредные производственные факторы, связанные с электромагнитными полями	+	СанПиН 1.2.3685-21 «Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания» ГОСТ 12.1.006-84 Система стандартов безопасности труда. Электромагнитные поля радиочастот. Допустимые уровни на рабочих местах и требования к проведению контроля
5. Психофизиологические факторы	+	ГОСТ 12.0.003-2015 Система стандартов по безопасности труда. Опасные и вредные производственные факторы. Классификация

## 5.2.1 Вредные и опасные производственные факторы

### 5.2.1.1 Повышенная или пониженная температура и влажность воздуха рабочей зоны

Пониженная или же повышенная температура воздуха рабочей зоны считается вредным производственным фактором и является фактором микроклимата рабочей среды, параметры которого регулируются СанПиН 1.2.3685-21 [1]. Он более характерен для рабочей среды программиста-разработчика.

Характеристиками микроклимата в производственных помещениях, являются:

- Температура воздуха ( $t$ , °C);
- Температура поверхностей ( $t$ , °C);
- Относительная влажность воздуха ( $\phi$ , %);
- Скорость движения воздуха ( $v$ , м/с);
- Интенсивность теплового облучения ( $I$ , Вт/м<sup>2</sup>).

Во время работы с ПЭВМ в производственных помещениях происходит систематическое выделение тепла вычислительной техникой, а также вспомогательными устройствами и средствами освещения. По причине того, что оператор находится поблизости от источников выделения тепла, данный фактор является одним из вредных факторов производственной среды оператора ПЭВМ, а высокая температура воздуха способствует быстрому перегреву организма и повышению утомляемости.

Влажность оказывает сильное влияние на терморегуляцию организма. Высокие показатели относительной влажности (более 85 %) затрудняют терморегуляцию, показатели влажности менее 20 % вызывают пересыхание слизистых оболочек человека.

Санитарные нормы устанавливают оптимальные и допустимые значения величин показателей микроклимата рабочих мест для различных категорий работ в теплый и холодный периоды года.

Для создания благоприятных условий труда и повышения производительности необходимо поддерживать оптимальные параметры микроклимата производственных помещений. Для этого должны быть предусмотрены следующие средства: центральное отопление, вентиляция (искусственная и естественная), искусственное кондиционирование.

Исходя из требований, рассматриваемых в данном разделе, в аудитории, в которой проводилось исследование поддерживается температура равная 20 – 21°C, при относительной влажности в 55–58%. В помещении функционирует система принудительной вентиляции. Проводится ежедневная влажная уборка и систематическое проветривание помещения. В зимнее время в помещении предусмотрена система водяного отопления со встроенными нагревательными элементами и терморегуляторами.

В таблице 5.2 представлены оптимальные и допустимые величины показателей микроклимата на рабочих местах производственных помещений. Величины соответствуют Ia категории интенсивности энергозатрат, так как вся

работа производится сидя и сопровождающиеся незначительным физическим напряжением.

Таблица 5.2 – Оптимальные величины показателей микроклимата на рабочих местах производственных помещений (СанПиН 1.2.3685-21 [1]).

Период года	Температура воздуха, °С	Температура поверхности, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	22-24	21-25	40-60	0,1
Теплый	23-25	22-26	40-60	0,1

### 5.2.1.2 Недостаточная освещенность рабочей зоны

Недостаточная освещенность рабочей зоны является вредным производственным фактором, возникающим при работе с ПЭВМ. Причиной недостаточной освещенности являются недостаточность естественного освещения, недостаточность искусственного освещения, пониженная контрастность.

Работа с компьютером подразумевает постоянный зрительный контакт с дисплеем ПЭВМ и занимает от 80 % рабочего времени. Недостаточность освещения снижает производительность труда, увеличивает утомляемость и может привести к появлению профессиональных болезней зрения.

Существуют общие требования и рекомендации к организации освещения на рабочем месте:

- Искусственное освещение в помещениях для эксплуатации ПЭВМ должно осуществляться системой общего равномерного освещения.
- Рабочие места следует размещать таким образом, чтобы естественный свет падал преимущественно слева, а дисплеи монитора были ориентированы боковой стороной к световым проемам.

Требования к освещению приведены в таблице 5.3:

Таблица 5.3 – Требования к освещению на рабочих местах с ПК (СанПиН

1.2.3685-21 [1]).

Вид	Требование
Освещенность на рабочем столе	200-400 лк
Освещенность на экране ПК	Не выше 200 лк
Блик на экране	Не выше 40 кд/м <sup>2</sup>
Прямая блеклость источника света	200 кд/м <sup>2</sup>
Показатель ослепленности	Не более 20
Показатель дискомфорта	Не более 15
Отношение яркости	
Между рабочими поверхностями	3:1-5:1
Между поверхностями стен и оборудования	10:1
Коэффициент пульсации	Не более 10%

### 5.2.1.3 Расчет имеющегося освещения для операторов ПЭВМ

Из схемы, которая представлена на рисунке 1 видно, что помещение имеет 2 двери и 2 окна. Площадь помещения составляет 45 м<sup>2</sup> (длина 9000, ширина 5000 мм, высота от уровня пола 3000 мм). Первой задачей размещения светильников является определение расчетной высоты подвеса  $H_D$ :

$$H_D = H - h_n - h_{\partial} \quad (5.1)$$

где  $H$  – высота помещения, м;

$h_n$  – расстояние светильников от перекрытия, как правило, принимается в пределах 0–1,5 м;

$h_{\partial}$  – высота рабочей поверхности над полом, м.

Соблюдение данных мер позволит сохранить зрение работника или избежать пагубного воздействия на глаза. Так как высота потолка данного помещения 3,0 м, то оптимальное значение размещения 3,0 м.

Светильники с линейными излучателями (люминесцентными лампами) обычно располагают рядами параллельно длинной стороне помещения или стене с окнами. В зависимости от уровня нормированной освещенности и возможности

размещения светильники располагают непрерывными рядами или рядами с разрывами. Нежелательно размещать светильники в один ряд, это допускается только для узких помещений.

Коэффициенты отражения был получен исходя из источника [10].

Формулы для расчетов основаны на источнике [11] страница 154.

Высота рабочей поверхности  $h_p = 0,7$  м;

Размеры помещения:  $A = 9$  м;  $B = 5$  м;  $H = 2,9$  м;  $S = 45$  м<sup>2</sup>; Коэффициент отражения потолка  $\rho_p = 70\%$  (для побеленного потолка). Коэффициент отражения стен  $\rho_{ст} = 50\%$ .

Коэффициент отражения пола  $\rho_r = 30\%$ .

Коэффициент запаса для помещения с малым выделением пыли  $K_z = 1,5$

Коэффициент неравномерности освещения  $Z=1,1$  (для люминесцентных ламп).

Параметр для светильника типа ОДР с защитной решеткой  $\lambda = 1,1- 1,3$ .  
Световой поток одной лампы  $\Phi = 2500$  лм.

Индекс помещения  $i$  рассчитывается по формуле 5.2:

$$i = \frac{S}{H \cdot (A+B)} \quad (7.2)$$

где  $S$  – площадь помещения;  $H$  – высота помещения;  $A$  – длина помещения;  $B$  – ширина помещения.

Вычислим индекс помещения  $i$  по формуле 5.2, смотреть формулу 5.3:

$$i = \frac{45}{3 \cdot (9+5)} = 1,07 \quad (7.3)$$

Получили в результате значение  $i = 1,07$ , округлим:  $i=1,1$ . В связи с учётом индекса помещения и коэффициентов отражения стен, согласно таблице

5.3 из источника [10] коэффициент использования светового потока равен  $\eta =$

46 %. Преобразовав формулу (7.4) из источника [10] получаем формулу (7.5) для вычисления требуемого количества люминесцентных ламп.

$$\Phi = \frac{E_n \cdot S \cdot K \cdot Z}{N \cdot \eta}, \quad (7.4)$$

$$N = \frac{E_n \cdot S \cdot K \cdot Z}{\Phi \cdot \eta}, \quad (7.5)$$

где  $\Phi$  – световой поток;  $E_n$  – освещенность рабочей поверхности от систем общего освещения;  $K$  – коэффициент пульсации освещенности;  $Z$  – коэффициент неравномерности освещения;  $N$  – число светильников.

Согласно СанПиН 1.2.3685-21 [1] нормированная степень освещенности для рассматриваемого помещения составляет 300 лк.

Вычислим требуемое количество люминесцентных ламп формуле 7.5, смотреть формулу 7.6:

$$N = \frac{300 \cdot 45 \cdot 1,5 \cdot 1,1}{2500 \cdot 0,46} = 19,3, \quad (7.6)$$

Получили 19,3 люминесцентных ламп. Округлив до ближайшего целого, получим 19 ламп. Такое количество ламп можно получить, применяя 2 ряда светильников по 5 светильников в ряд по 2 лампы. Итого 20 ламп.

На основании полученного значения можно вычислить фактическое значение освещения в помещении по формуле:

$$E_{\text{факт}} = \frac{\Phi \cdot N \cdot \eta}{S \cdot K \cdot Z}, \quad (7.7)$$

где  $\Phi$  – световой поток;  $E_n$  – освещенность рабочей поверхности от систем общего освещения;  $K$  – коэффициент пульсации освещенности;  $Z$

– коэффициент неравномерности освещения; N – количество рядов светильников. Вычислим фактическое значение освещения в помещении формуле 7.7,

смотреть формулу 7.8:

$$E_{\text{факт}} = \frac{2500 \cdot 20 \cdot 0,46}{45 \cdot 1,5 \cdot 1,1} = 309,76 \quad (7.8)$$

В результате получили значение  $E_{\text{факт}} = 309,76$  лк. Согласно СанПиН 1.2.3685-21 [1] освещённость для офисных помещений при общем освещении составляет 300 лк. С учётом того, что рассчитанная освещённость не должна отличаться от нормы более чем на 10% можно сделать вывод, что найденный световой поток, падающий на участок поверхности единичной площади, обеспечивает требуемую освещённость рабочего помещения.

#### **5.2.1.4 Повышенный уровень шума на рабочем месте**

Ненормированный показатель шума на рабочих местах оказывает негативное воздействие на психологическое состояние сотрудника. У работника на поставленной ему задаче понижаются концентрация и сосредоточенность, а увеличивается утомляемость и стресс. Повышенный уровень шума приводит к нарушению слуха или являться помехой для коммуникаций между сотрудниками [4].

Разработка модели соответствует виду трудовой деятельности, входящему в перечень «творческая деятельность, руководящая работа с повышенными требованиями, научная деятельность, конструирование и проектирование, программирование, преподавание и обучение, врачебная деятельность»; а рабочее место соответствует рабочему месту, входящему в перечень «рабочие места в помещениях дирекции, проектно-конструкторских бюро, расчетчиков, программистов вычислительных машин, в лабораториях для теоретических работ и обработки данных, приема больных в здравпунктах». В таблице 5.4

приведены допустимые значения уровней звукового давления в октавных полосах частот и уровня звука, создаваемого ПЭВМ.

Измерение уровня звука и уровней звукового давления производится на расстоянии 50 см от поверхности оборудования и на уровне расположения источника(ков) звука. Уровень шума исправного компьютера находится в пределах 35-50 дБА, что значительно ниже, чем допустимый уровень шума для данного рабочего места. На рабочем месте уровень шума считается допустимым, так как не вызывает беспокойства, и не воздействует на психологическое состояние.

При выполнении основной работы на мониторах и ПЭВМ, где работают инженерно-технические работники, уровень шума не должен превышать 60 дБА, в помещениях операторов ЭВМ (без дисплеев) –65 дБА, на рабочих местах в помещениях, где размещаются шумные агрегаты вычислительных машин –75 дБА. При выполнении магистерской работы воздействовал уровень шума не более 60 дБА [22].

Таблице 5.4 – Уровни звукового давления в октавных полосах со среднегеометрическими частотами.

Уровни звукового давления, дБ, в октавных полосах со среднегеометрическими частотами, Гц									Уровни звука и эквивалентные уровни звука (дБА)
31,5	63	125	250	500	1000	2000	4000	8000	50
86	71	61	54	49	45	42	40	38	

Наиболее типичным источником шума при данном виде деятельности будет являться компьютерная техника, в частности работающая система охлаждения ПЭВМ. Для снижения негативного влияния шума на рабочем месте могут быть применены следующие средства:

- использование охлаждающих систем с меньшим уровнем шума;
- использование звукопоглощающих материалов для системного блока компьютера для смягчения вибрации.

При отсутствии возможности избегания шума во время работы

необходимо организовать места отдыха на некотором расстоянии от источников шума.

#### **5.2.1.5 Производственные факторы, связанные с электромагнитными полями**

Источниками повышенной напряженности электромагнитного поля в данном случае является персональный компьютер. Допустимым считается 8-часовой рабочий день для сотрудника на своем рабочем месте, с предельно допустимым уровнем напряженности должен составлять не более 8 кА/м, а уровень магнитной индукции – 10 мТл. Соблюдение данных норм дает возможность избежать негативного воздействия электромагнитных излучений.

Для уменьшения уровня электромагнитного поля от персонального компьютера рекомендуется подключать к одной розетке не более двух компьютеров, сделать защитное заземление, подключать компьютер к розетке через нейтрализатор электрического поля [11].

Спектральные компьютерные очки рекомендуется использовать всем операторам при работе больше 2 часов в день, а при нарушении зрения на 2 диоптрии и более – независимо от продолжительности работы. Рекомендуемая интенсивность окраски линз 25-30%. Источниками электромагнитного излучения на рабочем месте выступают системные блоки и мониторы включённых компьютеров. Чтобы снизить воздействие таких видов излучения, рекомендуется использовать такие мониторы, уровень излучения которых понижен, а также установить защитные экраны и соблюдать режимы труда и отдыха.

Допустимые уровни ЭМП, создаваемых ПК приведены в таблице 5.5.

Таблица 5.5 – Допустимые уровни ЭМП, создаваемых ПК (СанПиН 1.2.3685-21 [1]).

Наименование параметров		ВДУ ЭМП
Напряженность	в диапазоне частот 5 Гц - 2 кГц	25 В/м

электрического поля	в диапазоне частот 2 кГц - 400 кГц	2,5 В/м
Плотность магнитного потока	в диапазоне частот 5 Гц - 2 кГц	250 нТл
	в диапазоне частот 2 кГц - 400 кГц	25 нТл
Электростатический потенциал экрана видеомонитора		500

### 5.2.2 Мероприятия и рекомендации по устранению и минимизации

Для поддержания нормальных значений параметров микроклимата на рабочих местах рекомендуется оснащать их системами отопления, вентиляции и кондиционирования воздуха. Также, в некоторых случаях, целесообразно обеспечить питьевое водоснабжение. В помещениях для работы с ПЭВМ должна производиться ежедневная влажная уборка, а также систематическое проветривание после каждого часа работы. Для защиты операторов ПЭВМ от негативного воздействия электромагнитных полей в первую очередь необходимо, чтобы используемая техника удовлетворяла нормам и правилам сертификации.

Для создания и поддержания благоприятных условий освещения для операторов ПЭВМ, их рабочие места должны соответствовать санитарно-эпидемиологическим правилам СанПиН 1.2.3685-21 [1].

Рабочее помещение должно иметь естественное и искусственное освещение, соответствующее показателям. В качестве источников искусственного освещения должны быть использованы люминесцентные лампы, лампы накаливания – для местного освещения.

В случаях, когда характер работы требует постоянного взаимодействия с компьютером (работа программиста разработчика) с напряжением внимания и сосредоточенности, при исключении возможности периодического переключения на другие виды трудовой деятельности, не связанные с ПЭВМ, рекомендуется организация перерывов на 10–15 мин. через каждые 45–60 мин. работы.

К мероприятиям по предотвращению возможности поражения электрическим током относятся:

- При производстве монтажных работ необходимо использовать только исправный инструмент, аттестованный службой КИПиА;
- С целью защиты от поражения электрическим током, возникающим между корпусом приборов и инструментом при пробое сетевого напряжения на корпус, корпуса приборов и инструментов должны быть заземлены;
- При включенном сетевом напряжении работы на задней панели должны быть запрещены;
- Все работы по устранению неисправностей должен производить квалифицированный персонал;
- Необходимо постоянно следить за исправностью электропроводки.

### 5.3 Экологическая безопасность

В данном разделе рассматривается влияние на окружающую среду деятельности по разработке проекта, а также самого продукта в результате его реализации на производстве. Непосредственно программный продукт, разработанный в ходе выполнения магистерской диссертации, не наносит вреда окружающей среде ни на стадиях его разработки, ни на стадиях эксплуатации. Однако, средства, необходимые для его разработки и эксплуатации могут наносить вред окружающей среде.

В лаборатории не ведется никакого производства. К отходам, производимым в помещении, можно отнести, в первую очередь, это бумажные отходы – макулатура, пластиковые отходы, неисправные детали персональных компьютеров и других видов ЭВМ. Бумажные отходы рекомендуется накапливать и передавать их в пункты приема макулатуры для дальнейшей переработки. Пластиковые бутылки складывать в специально предназначенные контейнеры.

Современные ПЭВМ производят практически без использования вредных веществ, опасных для человека и окружающей среды. Исключением являются аккумуляторные батареи компьютеров и мобильных устройств. В аккумуляторах содержатся тяжелые металлы, кислоты и щелочи, которые могут наносить ущерб окружающей среде, попадая в гидросферу и литосферу, если они были неправильно утилизированы. Для утилизации аккумуляторов необходимо обращаться в специальные организации, специализировано занимающиеся приемом, утилизацией и переработкой аккумуляторных батарей. Люминесцентные лампы, применяющиеся для искусственного освещения рабочих мест, также требуют особой утилизации, т.к. в них присутствует от 10 до 70 мг ртути, которая относится к чрезвычайно -опасным химическим веществам и может стать причиной отравления живых существ, а также загрязнения атмосферы, гидросферы и литосферы. Согласно ГОСТу Р 52105-

2003 отходы люминесцентных ламп, именуемые по Федеральному классификационному каталогу отходов (ФККО), как «Ртутные лампы, люминесцентные ртутьсодержащие трубки отработанные и брак», экспериментально установлен 1 класс опасности по степени токсичности и степени влияния на здоровье человека и окружающую его среду. Сроки службы таких ламп составляют около 5-ти лет, после чего их необходимо сдавать на переработку в специальных пунктах приема. Юридические лица обязаны сдавать лампы на переработку и вести паспорт для данного вида отходов. Дополнительным методом снижения отходов является увеличение доли электронного документооборота [17].

## 5.4 Безопасность в чрезвычайных ситуациях

Наиболее характерной для объекта, где размещаются рабочие помещения, оборудованные ПЭВМ, чрезвычайной ситуацией является пожар. Помещение для работы операторов ПЭВМ по системе классификации категорий помещений по взрывопожарной и пожарной опасности относится к категории Д (из 5-ти категорий А, Б, В1-В4, Г, Д), т.к. относится к помещениям с негорючими веществами и материалами в холодном состоянии [18].

Все сотрудники организации обязаны быть ознакомлены с инструкцией по пожарной безопасности, пройти инструктаж по технике безопасности и строго соблюдать его. Запрещается использовать электроприборы в условиях, не соответствующих требованиям инструкций изготовителей, или имеющие различного рода неисправности, которые в соответствии с инструкцией по эксплуатации могут привести к пожару, а также использовать электропровода и кабели с поврежденной или потерявшей защитные свойства изоляцией.

Перед уходом из служебного помещения требуется провести его осмотр, закрыть окна, убедиться в том, что в помещении отсутствуют источники возможного возгорания, все электроприборы отключены и выключено освещение.

С периодичностью не менее одного раза в три года необходимо проводить замеры сопротивления изоляции токоведущих частей силового и осветительного оборудования. Увеличение устойчивости достигается за счет проведения соответствующих организационно-технических мероприятий, подготовки персонала к работе в ЧС.

Сотрудник при обнаружении пожара или признаков горения (задымление, запах гари, увеличение температуры и т.п.) должен:

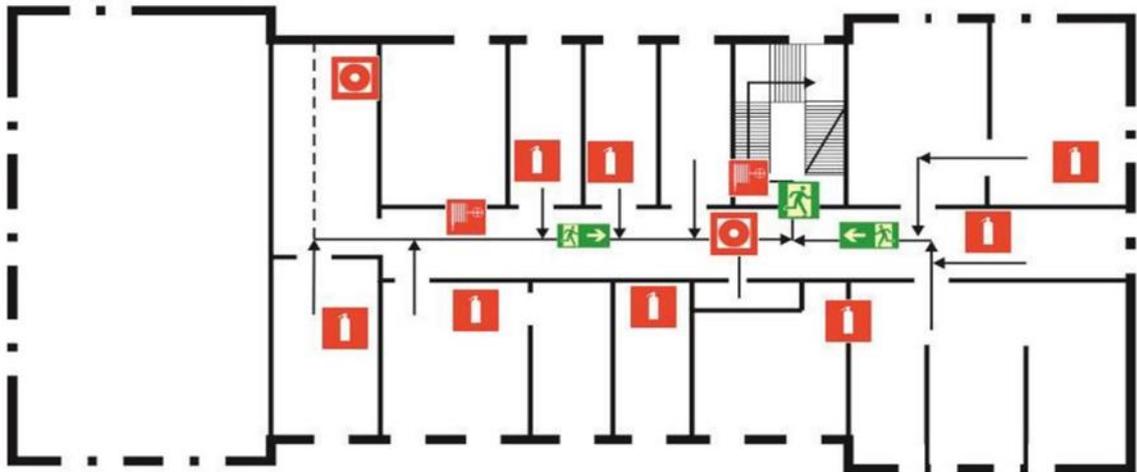
- Требуется прекратить работу, вызвать пожарную охрану по телефону «01»;
- По возможности принять меры по эвакуации людей и материальных ценностей;

- Отключить от сети электрооборудование;
- Приступить к тушению пожара имеющимися средствами пожаротушения;
- Сообщить непосредственному или вышестоящему начальнику и оповестить окружающих сотрудников;
- При общем сигнале опасности покинуть здание согласно «Плану эвакуации людей при пожаре и других ЧС».

Для тушения пожара применять ручные углекислотные огнетушители (типа ОУ-2, ОУ-5), находящиеся в помещениях офиса, и пожарный кран внутреннего противопожарного водопровода. Они предназначены для тушения начальных возгораний различных веществ и материалов, за исключением веществ, горение которых происходит без доступа воздуха. Огнетушители должны постоянно содержаться в исправном состоянии и быть готовыми к действию. Категорически запрещается тушить возгорания в помещениях офиса при помощи химических пенных огнетушителей (типа ОХП-10).

Рабочее место располагается в 10 корпусе ТПУ 402 аудитория. На рисунке 4 представлен план эвакуации четвертого этажа 10 корпуса.

## ПЛАН ЭВАКУАЦИИ ПРИ ПОЖАРЕ И ДРУГИХ ЧС



Действия при пожаре Сохранять спокойствие!	
<b>1</b> Сообщить по телефону: 01, сот 010	<ul style="list-style-type: none"> <li>- адрес объекта</li> <li>- место возникновения пожара</li> <li>- свою фамилию</li> </ul>
<b>2</b> Эвакуировать людей	<ul style="list-style-type: none"> <li>- ориентироваться по знакам направления движения</li> <li>- взять с собой пострадавших</li> </ul>
<b>3</b> По возможности принять меры по тушению пожара	<ul style="list-style-type: none"> <li>- использовать средства противопожарной защиты</li> <li>- при необходимости обесточить помещение</li> </ul>



Рисунок 5.3 – План эвакуации.

### **5.5 Выводы по разделу**

Каждый работник должен проводить профессиональную деятельность с учетом социальных, правовых, экологических и культурных аспектов, вопросов здоровья и безопасности, нести социальную ответственность за принимаемые решения, осознавать необходимость устойчивого развития.

В данном разделе были рассмотрены основные вопросы соблюдения прав работника на труд, выполнения правил к безопасности труда, промышленной безопасности, экологии и ресурсосбережения.

Установлено, что рабочее место исследователя удовлетворяет требованиям безопасности и гигиены труда во время реализации проекта, а вредное воздействие объекта исследования на окружающую среду не превышает норму.

## **Заключение**

В ходе выполнения выпускной квалификационной работы был проведен анализ предметной области, включающий в себя предсказание трудозатрат по задачам с учетом разных факторов, которые влияют на длительность.

Нейронная модель помогает принять решения при оценке длительности задачи.

Наиболее вероятная длительность задачи, рассчитанная программой, служит ориентиром при выставлении трудозатрат задачи. Функция помогает более точно назначить срок задачи, ориентируясь среднюю длительность выполнения задачи, а также на минимальный и максимальный срок выполнения.

Значительное отклонение от предлагаемой программой длительности может стать поводом для того, чтобы скорректировать установленные сроки. Однако последнее слово при установлении длительности задачи остается за ведущим проектировщиком, поскольку программа не может учесть все многообразие факторов.

В том случае, если срок, установленный для выполнения задачи, критично превышает рекомендуемое значение, ведущий проектировщик может и должен обратить на это внимание и выяснить причину увеличения или сокращения сроков. Это необходимо для объективного планирования - слишком короткие сроки выполнения задачи без обоснования причин (допустим, сотрудник нашел способ выполнять задачу быстрее) могут привести к нарушению ТХПР или увеличению сроков задачи на этапе выполнения проекта, что увеличит общие трудозатраты. Слишком долгие сроки задач - повод помочь проектировщику освоить новые способы работы, которые позволят выполнять эту задачу быстрее.

## Список использованных источников

1. Word2vec в картинках. [Электронный ресурс] - <https://habr.com/ru/post/446530/> (дата обращения 20.05.2021)
2. Distributed Representations of Words and Phrases and their Compositionality. [Электронный ресурс] – <https://papers.nips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf> (дата обращения 20.05.2021)
3. Distributed Representations of Words and Phrases and their Compositionality. [Электронный ресурс] – <https://arxiv.org/pdf/1301.3781.pdf>
4. Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. A neural probabilistic language model. *The Journal of Machine Learning Research*, 3:1137–1155, 2003.
5. Ronan Collobert and Jason Weston. A unified architecture for natural language processing: deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167. ACM, 2008.
6. Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *ICML*, 513–520, 2011.
7. Классификация рисков [Электронный ресурс] // URL: <https://fireman.club/inseklodepia/klassifikatsiya-riskov/> Y. Zhong (дата обращения 20.05.2021)
8. Text line segmentation in historical document images using an adaptive U-Net architecture / O. Mechi, M. Mehri, R. Ingold, and N. Essoukri Ben Amara // 2019 International Conference on Document Analysis and Recognition (ICDAR). – 2019. – P.369-374.
9. Robust wide-baseline stereo from maximally stable extremal regions / J. Matas, O. Chum, M. Urban, and T. Pajdla // *Image and Vision Computing*. – 2004. – Vol. 22, №10. – P.761-767.
10. Zhang, Z. Multi-oriented text detection with fully convolutional networks.

/ Z. Zhang, C. Zhang, W. Shen, C. Yao, W. Liu, X. Bai // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – P.4159-4167.

11. СанПиН 1.2.3685-21 «Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания»

12. ГОСТ 12.2.032-78 ССБТ «Рабочее место при выполнении работ сидя. Общие эргономические требования»

13. СанПиН 2.2.4.548-96 Гигиенические требования к микроклимату производственных помещений

14. СанПиН 2.2.1/2.1.1.1278–03. Гигиенические требования к естественному, искусственному и совмещённому освещению жилых и общественных зданий. М.: Минздрав России, 2003.

15. СН 2.2.4/2.1.8.562-96 Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки. Санитарные нормы.

16. ГОСТ 12.1.006-84 Система стандартов безопасности труда (ССБТ). Электромагнитные поля радиочастот. Допустимые уровни на рабочих местах и требования к проведению контроля.

17. ТОИ Р-45-084-01 Типовая инструкция по охране труда при работе на персональном компьютере

18. ГОСТ 12.1.004-91 Система стандартов безопасности труда (ССБТ). Пожарная безопасность. Общие требования

19. ГОСТ 12.1.038-82 Система стандартов безопасности труда (ССБТ). Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов.

20. Метод коэффициента использования [Электронный ресурс]. – Режим доступа: <https://www.websor.ru/metodkoeffi.html> (Дата обращения: 02.06.2020).

21. Кнорринг, Г. М. Осветительные установки / Г. М. Кнорринг. – Л.: Энергоиздат. Ленингр. отделение, 1981. – 288 с.

22. СНиП II-4-79 Естественное и искусственное освещение

23. СанПиН 2.21/2.1.1/1278-03 Гигиенические требования к

естественному, искусственному и совмещенному освещению жилых и общественных зданий

24. Электробезопасность [Электронный ресурс]. Режим доступа: <https://ru.wikipedia.org/wiki/Электробезопасность>.

25. СНиП 21-01-97\* Пожарная безопасность зданий и сооружений.

26. Электромагнитные поля на рабочем месте [Электронный ресурс] URL: <https://habr.com/ru/post/140431/> (Дата обращения: 30.05.2021).

27. ГОСТ Р 52105-2003 Ресурсосбережение. Обращение с отходами. Классификация и методы переработки ртутьсодержащих отходов. Основные положения.

28. СП 12.13130.2009 Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности.

29. Development of a software framework for real-time management of intelligent devices: Статья / Naumovic Tamara, Baljak Luka, Zivojinović Lazar, Filipovic Filip. – Belgrade, 2019. С. 35-46.

30. Web приложение реального времени для простых устройств [Электронный ресурс] // URL: <https://habr.com/ru/post/325942/> (дата обращения: 20.05.2021).

31. База знаний Инициатора [Электронный ресурс] // URL: <https://help.mover.run/?category=ed4a2255-9efe-11e9-a8b0-0242ac170002/> (дата обращения: 20.05.2021).

32. Годин, А. М. Статистика: учебник / А. М. Годин. – Москва: Дашков и К, 2016. – 451 с.

## Приложение А

### Раздел 1

#### Development of models for determining the most probable duration of with NLP

Студент:

Группа	ФИО	Подпись	Дата
8BM92	Курбонов Шамил Алиевич		

Руководитель ВКР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Фадеев А.С.	к.т.н.		

Консультант-лингвист отделения иностранных языков ШБИП:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент	Аксёнова Н. В.	к.ф.н.		

## 1 Input data and normalization

As a global sample, based on the project at the moment, we have almost a million tasks, and within these tasks there are tasks with the following statuses (Pic. 5).

```
const
  STATUS_NEW = 1,      // Запланирована
  STATUS_CANCELLED = 2, // Отменена
  STATUS_NEED_EDIT = 3, // Требуются изменения
  STATUS_ISSUED = 4,   // Возникла проблема
  STATUS_DONE = 5,     // Выполнена
  STATUS_VERIFIED = 6, // Принята
```

Picture 1. Task statuses

Based on the task statuses, we see that tasks that have a status from 1 to 4 do not have any exact data on labor costs, of course we can proceed from some factors, for example, if the status is cancelled, then you can navigate and give less rating, if names of the problem coincide in the meaning with what we are trying to predict, but for the first stage such insignificant factors are not taken into account, since there are more significant factors and you can gradually improve the algorithm and add these factors.

The situation with the status completed is very interesting, on the one hand, these are tasks where we have all the necessary values, labor costs, performers, etc., on the other hand, within the framework of the project, there may be improvements on this status, for this reason we will leave this status.

As the main sample, we receive tasks with the Accepted status, and there are almost 40 thousand of them within the project, so the main factors to be guided by in the first release are the task deadlines from which we will get the total execution period, owner and name of the task.

Thus, we got the main sample for analyzing tasks in csv format. To automatically update this list, and how we make requests to the server, I will tell you at the end, in the 6th section.

### 1.1 Sample quality

Many current NLP systems and techniques treat words as atomic units - there is no notion of similarity between words, as these are represented as indices in a

vocabulary. This choice has several good reasons - simplicity, robustness and the observation that simple models trained on huge amounts of data outperform complex systems trained on less data. An example is the popular N-gram model used for statistical language modeling - today, it is possible to train N-grams on virtually all available data (trillions of words [3]).

However, the simple techniques are at their limits in many tasks. For example, the amount of relevant in-domain data for automatic speech recognition is limited - the performance is usually dominated by the size of high quality transcribed speech data (often just millions of words). In machine translation, the existing corpora for many languages contain only a few billions of words or less. Thus, there are situations where simple scaling up of the basic techniques will not result in any significant progress, and we have to focus on more advanced techniques.

With progress of machine learning techniques in recent years, it has become possible to train more complex models on much larger data set, and they typically outperform the simple models. Probably the most successful concept is to use distributed representations of words [10]. For example, neural network based language models significantly outperform N-gram models [17].

The quality of the sample depends on the accuracy of the task names. The more accurately the task is named, the more accurately its expected duration is calculated, since the calculations will analyze the most similar tasks. The title of the task should make it clear its content.

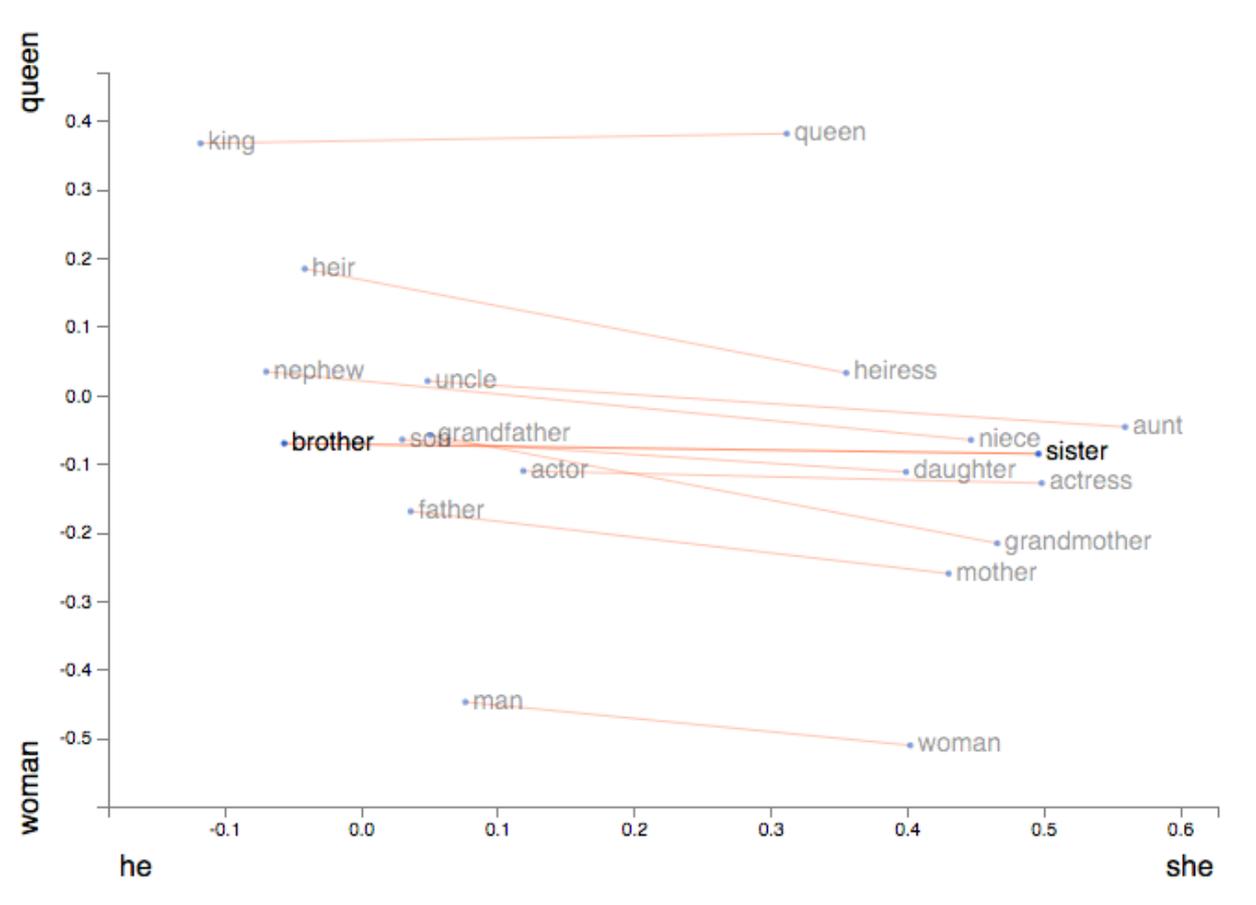
Tasks whose title consists of one word (for example, “Calculations”, “Printing”), without specifying the type of work, are not included in the sample. The standard uniform naming of tasks allows you to make the selection more accurate and determine the expected duration for the most specific task.

The main requirement that I would note in this section is that you need to take into account the load of various factors that affect the deadlines of tasks, and so that at any time you can add other input parameters for analysis, such as employee loyalty, employee failures when solving a task from this stage, etc.

### 1.2 Analysis of the word2vec algorithm

Word2vec is an algorithm that transforms words into vectors, so that words with similar meaning end up laying close to each other, developed in 2013. Moreover, it allows us to use vector arithmetics to work with analogies, for example the famous king - man + woman = queen. In addition to working with words, some of his concepts have proven effective in developing recommendation mechanisms and giving meaning to data, even in commercial, non-linguistic tasks. Companies such as Airbnb, Alibaba, Spotify and Anghami have used this technology in their recommendation engines.

I will try to explain how it works, with special emphasis on the meaning of vector differences, at the same time omitting as many technicalities as possible.



Picture 2. Word2Vec with special emphasis

### 1.3 Counts, coincidences and meaning

*I love letter co-occurrence in the word co-occurrence.*

Sometimes a seemingly naive technique gives powerful results. It turns out that merely looking at word coincidences, while ignoring all grammar and context, can

provide us insight into the meaning of a word. Consider this sentence:

*A small, fluffy roosety climbed a tree.*

What's a roosety? I would say that something like a squirrel, since the two words can be easily interchanged. Such reasoning is called the distributional hypothesis and can be summarized as:

*a word is characterized by the company it keeps - John Rupert Firth*

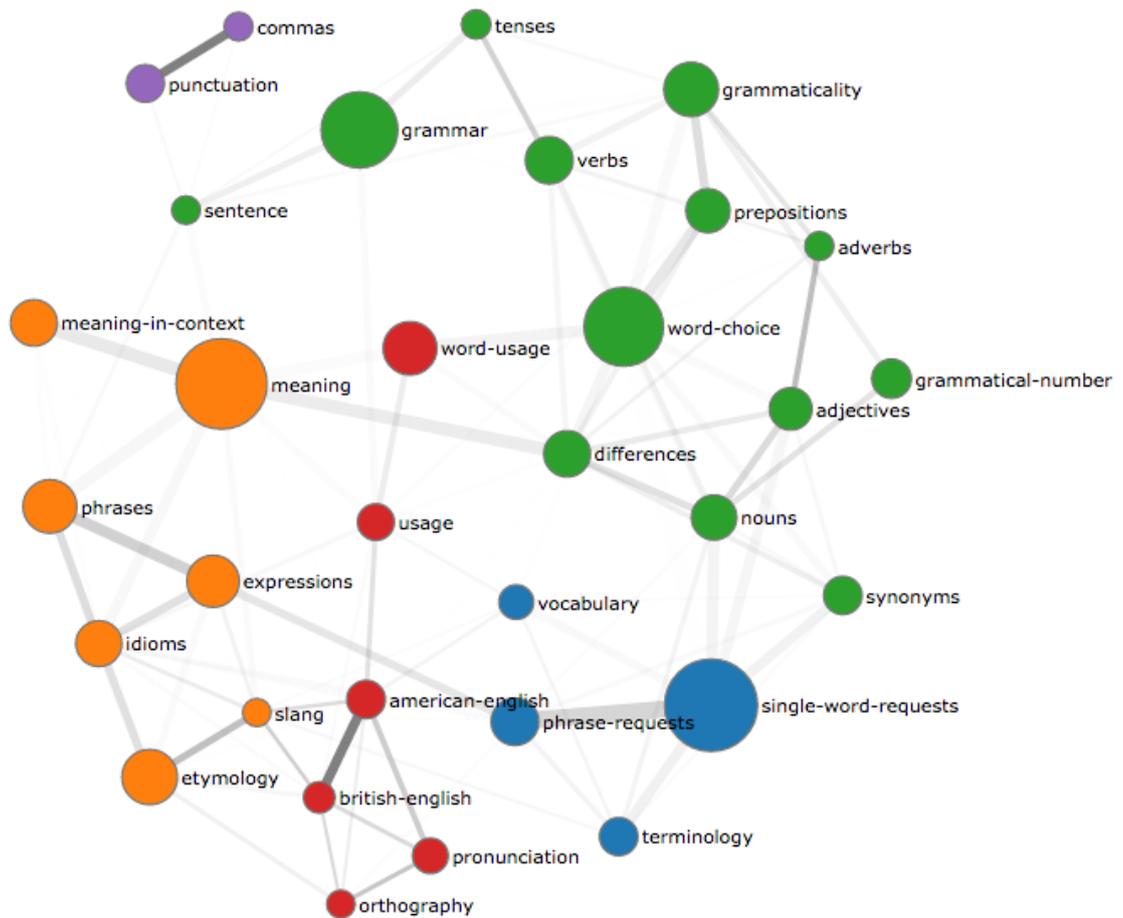
If we want to teach it to a computer, the simplest, approximated approach is making it look only at word pairs. Let  $P(a|b)$  be the conditional probability that given a word  $b$  there is a word  $a$  within a short distance (let's say - being spaced by no more than 2 words). Then we claim that two words  $a$  and  $b$  are similar if

$$P(w|a)=P(w|b)P(w|a)=P(w|b)$$

for every word  $w$ . In other words, if we have this equality, no matter if there is word  $a$  or  $b$ , all other words occur with the same frequency.

Even simple word counts, compared by source, can give interesting results, e.g. that in lyrics of metal songs words (cries, eternity or ashes are popular, while words particularly or approximately are not, well, particularly common), see *Heavy Metal and Natural Language Processing*. See also *Gender Roles with Text Mining and N-grams* by Julia Silge.

Looking at co-occurrences can provide much more information (Pic. 2). For example, one of my projects, *TagOverflow*, gives insight into structure of programming, based only on the usage of tags on Stack Overflow. It also shows that I am in love with pointwise mutual information, which brings us to the next point.



Picture 3. Words co-occurrences

### 1.4 Pointwise mutual information and compression

*The Compressors: View cognition as compression. Compressed sensing, approximate matrix factorization - The (n) Cultures of Machine Learning - HN discussion*

In principle, we can compute  $P(a|b)$  for every word pair. But even with a small dictionary of 100 000 words (bear in mind that we need to keep all declinations, proper names and things which are not in official dictionaries, yet are in use) keeping track of all pairs would require 8 gigabytes of space.

Often instead of working with conditional probabilities, we use the pointwise mutual information (PMI), defined as:

$$PMI(a,b) = \log \left[ \frac{P(a,b)}{P(a)P(b)} \right] = \log \left[ \frac{P(a|b)P(b)}{P(a)} \right].$$

Its direct interpretation is how much more likely we get a pair than if it were at random. The logarithm makes it easier to work with words appearing at frequencies of different orders of magnitude. We can approximate PMI as a scalar product:

$$\text{PMI}(a,b)=\vec{v} \cdot \vec{a} \cdot \vec{v} \cdot \vec{b}, \text{PMI}(a,b)=\vec{v} \rightarrow \vec{a} \cdot \vec{v} \rightarrow \vec{b},$$

where  $\vec{v} \rightarrow i$  are vectors, typically of 50-300 dimensions.

At the first glance it may be strange that all words can be compressed to a space of much smaller dimensionality. But there are words that can be trivially interchanged (e.g. John to Peter) and there is a lot of structure in general.

The fact that this compression is lossy may give it an advantage, as it can discover patterns rather than only memorize each pair. For example, in recommendation systems for movie ratings, each rating is approximated by a scalar product of two vectors - a movie's content and a user's preference. This is used to predict scores for as yet unseen movies, see Matrix Factorization with TensorFlow - Katherine Bailey.

### 1.5 Word similarity and vector closeness

*Stock Market ≈ Thermometer - Five crazy abstractions my word2vec model just did*

Let us start with the simple stuff - showing word similarity in a vector space. The condition that  $P(w|a)=P(w|b)$  is equivalent to

$$\text{PMI}(w,a)=\text{PMI}(w,b),$$

by dividing both sides by  $P(w)$  and taking their logarithms. After expressing PMI with vector products, we get

$$\vec{v} \cdot \vec{w} \cdot \vec{a} = \vec{v} \cdot \vec{w} \cdot \vec{b}$$

$$\vec{v} \cdot \vec{w} \cdot (\vec{a} - \vec{b}) = 0$$

If it needs to work for every  $\vec{v}$ , then

$$\vec{a} = \vec{b}.$$

Of course, in every practical case we won't get an exact equality, just words being close to each other. Words close in this space are often synonyms (e.g. happy and delighted), antonyms (e.g. good and evil) or other easily interchangeable words (e.g. yellow and blue). In particular most of opposing ideas (e.g. religion and atheism) will have similar context.

What I find much more interesting is that words form a linear space. In particular, a zero vector represent a totally uncharacteristic word, occurring with every other word at the random chance level (as its scalar product with every word is zero,

so is its PMI).

It is one of the reasons why for vector similarity people often use cosine distance, i.e.

$$\frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} = \frac{a \cdot b}{\sqrt{a^2} \sqrt{b^2}} = \frac{a \cdot b}{|a| |b|}$$

That is, it puts emphasis on the direction in which a given word co-occurs with other words, rather than the strength of this effect.

### 1.6 Analogies and linear space

If we want to make word analogies (a is to b is as A is to B), one may argue that it can be expressed as an equality of conditional probability ratios

$$\frac{P(w|a)P(w|b)}{P(w|A)P(w|B)} = \frac{P(w|a)P(w|b)}{P(w|A)P(w|B)}$$

for every word w. Sure, it looks (and is!) a questionable assumption, but it is still the best thing we can do with conditional probability. I will not try defending this formulation - I will show its equivalent formulations.

For example, if we want to say dog is to puppy as cat is to kitten, we expect that if e.g. word nice co-occurs with both dog and cat (likely with different frequencies), then it co-occurs with puppy and kitten by the same factor. It appears it is true, with the factor of two favoring the cubs - compare pairs to words from Google Books Ngram Viewer (while n-grams look only at adjacent words, they can be some sort of approximation).

By proposing ratios for word analogies we implicitly assume that the probabilities of words can be factorized with respect to different dimensions of a word. For the discussed case it would be:

$$\begin{aligned} \frac{P(w|\text{dog})P(w|\text{puppy})}{P(w|\text{cat})P(w|\text{kitten})} &= \frac{f(w|\text{species}=\text{dog}) \times f(w|\text{age}=\text{adult}) \times P(w|\text{is\_a\_pet})}{f(w|\text{species}=\text{cat}) \times f(w|\text{age}=\text{adult}) \times P(w|\text{is\_a\_pet})} \\ &= \frac{f(w|\text{species}=\text{dog}) \times f(w|\text{age}=\text{cub}) \times P(w|\text{is\_a\_pet})}{f(w|\text{species}=\text{cat}) \times f(w|\text{age}=\text{cub}) \times P(w|\text{is\_a\_pet})} \end{aligned}$$

So, in particular:

$$\frac{P(w|\text{dog})P(w|\text{puppy})}{P(w|\text{cat})P(w|\text{kitten})} = \frac{f(w|\text{age}=\text{adult})f(w|\text{age}=\text{cub})}{f(w|\text{age}=\text{adult})f(w|\text{age}=\text{cub})}$$

$$f(w|age=adult)f(w|age=cub)=P(w|cat)P(w|kitten).$$

How does the equality of conditional probability ratios translate to the word vectors? If we express it as mutual information (again,  $P(w)$  and logarithms) we get

$$\vec{v} \cdot \vec{w} \cdot \vec{a} - \vec{v} \cdot \vec{w} \cdot \vec{b} = \vec{v} \cdot \vec{w} \cdot \vec{A} - \vec{v} \cdot \vec{w} \cdot \vec{B}$$

$$\vec{B}, \vec{v} \rightarrow \vec{w} \cdot \vec{v} \rightarrow \vec{a} - \vec{v} \rightarrow \vec{w} \cdot \vec{v} \rightarrow \vec{b} = \vec{v} \rightarrow \vec{w} \cdot \vec{v} \rightarrow \vec{A} - \vec{v} \rightarrow \vec{w} \cdot \vec{v} \rightarrow \vec{B},$$

which is the same as

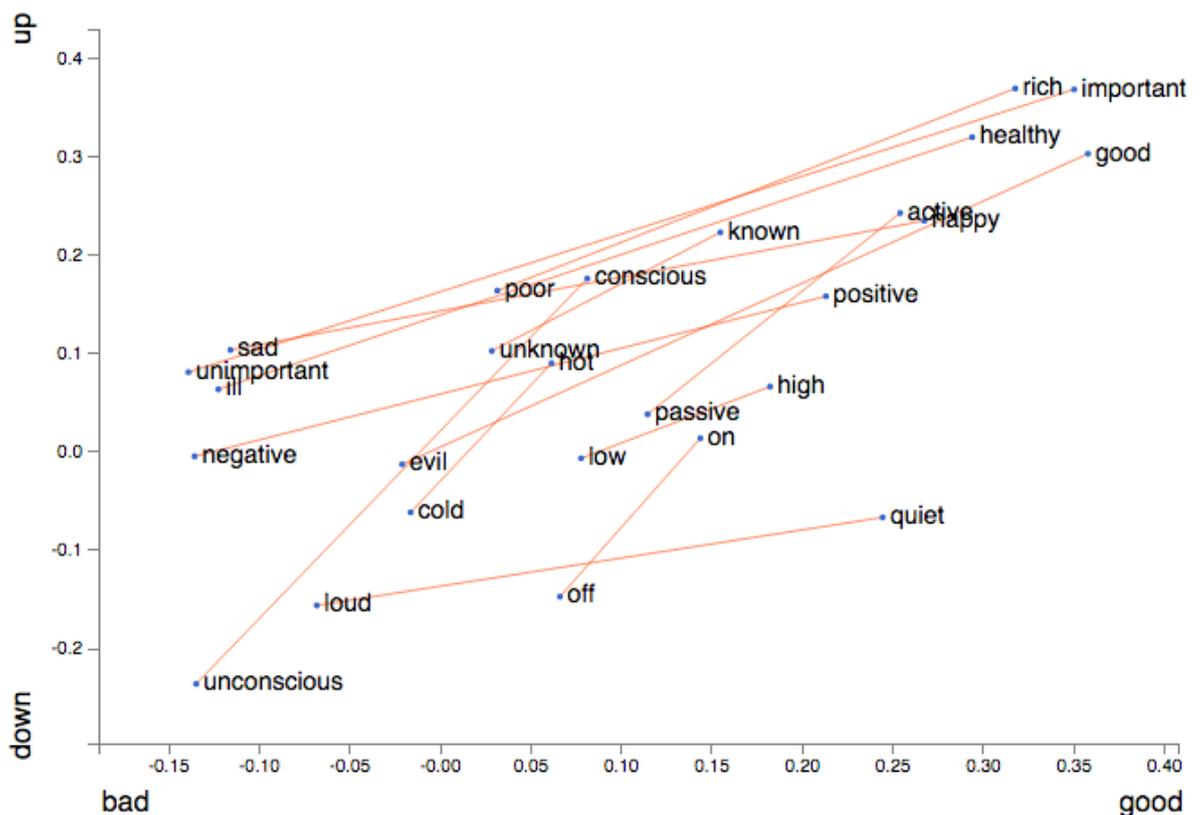
$$\vec{v} \cdot \vec{w} \cdot (\vec{v} \cdot \vec{a} - \vec{v} \cdot \vec{b} - \vec{A} + \vec{B}) = 0. \vec{v} \rightarrow \vec{w} \cdot (\vec{v} \rightarrow \vec{a} - \vec{v} \rightarrow \vec{b} - \vec{A} + \vec{B}) = 0.$$

Again, if we want it to hold for any word  $w$ , this vector difference needs to be zero.

We can use analogies for meaning (e.g. changing gender with vectors), grammar (e.g. changing tenses) or other analogies (e.g. cities into their zip codes). It seems that analogies are not only a computational trick - we may actually use them to think all the time, see:

George Lakoff, Mark Johnson, *Metaphors We Live By* (1980)

and their list of conceptual metaphors in English ([webarchive](#)), in particular look for X is Up, plotted below (Pic. 3)



## 1.7 Differences and projections

$woman - man = female - male = she - he$ , so  $wo = fe = s$

Difference of words vectors, like

$$\vec{v} \cdot she - \vec{v} \cdot he \rightarrow she - v \rightarrow he$$

are not words vectors by themselves. However, it is interesting to project a word on this axis. We can see that the projection

$$\vec{v} \cdot w \cdot (\vec{v} \cdot a - \vec{v} \cdot b) = \log[P(w|a)] - \log[P(w|b)] \rightarrow w \cdot (\vec{v} \rightarrow a - \vec{v} \rightarrow b) = \log_{f_0}^{[...]}[P(w|a)] - \log_{f_0}^{[...]}[P(w|b)]$$

is exactly a relative occurrence of a word within different contexts.

When we want to look at common aspects of a word it is more natural to average two vectors rather than take their sum. While people use it interchangeably, it only works because cosine distance ignores the absolute vector length. So, for a gender neutral pronoun we use  $(\vec{v} \cdot she + \vec{v} \cdot he) / 2$  rather than their sum.

## 1.8 Implementation of the algorithm using the gensim library

The evaluation of this algorithm is based on the cosine coefficient, or also called the Otai coefficient, you can read more about the vector estimation mechanisms in the article by Thomas Mikolov "Efficient Estimation of Word Representations in Vector Space" and also in the literature of this work, I was more focused on solving the problem.

Thus, the implementation of this algorithm word2vec is in the open-source gensim library.

Now we need a lot of words, about a million combinations and at least 500 thousand. Less than this figure, the word2vec algorithm is not stable.

Within the global sample, we have 1 million tasks, but these tasks very often coincide and there are only 65 thousand unique names, respectively, the combination of these sentences is not enough to build a normal model, thus, it was decided to use combinations of the National Corpus of the Russian language and the Wikipedia dump

to 2018, thus, using NLP methods, the model was trained to understand words and find similar meanings in meaning.

We added data on our tasks to the RNC and Wiki model.

As a result, after generation, we get a word2vec model of vectors for each word, the vector dimension was taken as 300 according to the Google standard.

Add the Word2Vec model and load our unique task names:

```
from gensim.models import Word2Vec, Doc2Vec, KeyedVectors
import nltk
import pandas as pd
import re

df = pd.read_csv('./data/tasks-optimaze.csv')
```

df.head(10)

	id	times	owner_id	name
0	00014a82-970c-11e9-865b-0242ac130007	14400	NaN	Оформление ПП на отм. +18,250. Блок 1, 2
1	0001c00a-07cb-11eb-b5eb-0242c0a87005	205200	2be0ccf2-3608-11ea-9dc1-0242c0a85009	Секции 6,7. Гидравлический расчет
2	000235a6-7eee-11ea-acd4-0242c0a87005	86400	NaN	Составление эскиза ТЗ на Инженерно-Геологическ...
3	000a739d-2919-4f85-8775-506041deebac	86400	NaN	Формирование ведомости монтажной пены
4	000b41b7-6a4f-11ea-acd4-0242c0a87005	32400	NaN	Расстановка НО и компенсаторов 1 блока

Picture 5. Adding data

## Приложение Б

### Поиск схожих слов

```
from gensim.models import KeyedVectors
import nltk
```

In [2]:

```
# Model with tags
# Russian National Corpus and Russian Wikipedia dump of December 2018
rncAndWikiWV = KeyedVectors.load("./initiator-models/rnc-and-wiki.wordvectors")
```

In [3]:

```
# Model without tags
# Initiator texts of task
initiatorWV = KeyedVectors.load("./initiator-models/initiator-word2vec.wordvectors")
```

In [4]:

```
nltk.tagset_mapping('ru-rnc', 'universal') == {'!': '.', 'A': 'ADJ', 'C': 'CONJ', 'AD': 'ADV', 'NN': 'NOUN', 'VG': 'VERB', 'COMP': 'CONJ', 'NC': 'NUM', 'VP': 'VERB', 'P': 'ADP', 'IJ': 'X', 'V': 'VERB', 'Z': 'X', 'VI': 'VERB', 'YES_NO_SENT': 'X', 'PTCL': 'PART'}
```

Out[4]:

```
True
```

In [5]:

```
words = nltk.word_tokenize('пересчитать расход')
```

In [6]:

```
wordsWithTag = []
```

```
for word in words:
    wordWithTag = nltk.pos_tag([word], tagset='universal', lang='rus')[0]
    wordsWithTag.append(wordWithTag[0] + '_' + wordWithTag[1])
```

In [7]:

```
def getMostSimilar(text, pv = 0.5):
    return [(word[0][0:word[0].find('_')]) for word in rncAndWikiWV.most_similar(text) if word[1] >= pv]
```

In [8]:

```
result = getMostSimilar(wordsWithTag)
```

In [9]:

```
result
```

Out[9]:

```
['затрата',
 'издержка',
 'трата',
 'пересчитывать',
 'сумма',
 'смета',
 'деньги',
 'стоимость',
 'потребленный',
 'непроцентный']
```

```
In [10]:
# check each word and find most similarity
for word in wordsWithTag:
    [result.append(text) for text in getMostSimilar(word) if text not in result]
```

```
In [11]:
result
```

```
Out[11]:
['затрата',
 'издержка',
 'трата',
 'пересчитывать',
 'сумма',
 'смета',
 'деньги',
 'стоимость',
 'потребленный',
 'непроцентный',
 'сосчитать',
 'бюджет',
 'ассигнование',
 'доход',
 'себестоимость',
 'расходный']
```

```
In [12]:
# Try to find most similarity words in initiator-model

initiatorWV.most_similar(words)
```

```
Out[12]:
[('авк', 0.9816160202026367),
 ('вентшахты', 0.9801523089408875),
 ('конвектор', 0.9799370765686035),
 ('nvd', 0.9795823693275452),
 ('расчётный', 0.9793539643287659),
 ('вдобра', 0.9787059426307678),
 ('арене', 0.9782119989395142),
 ('месторасположение', 0.9765300154685974),
 ('удельного', 0.9762309789657593),
 ('пожароопасности', 0.9757450819015503)]
```

```
In [13]:
# initiator-model with rnc-wiki-model!!!
words

Out[13]:
['пересчитать', 'расход']
```