Федеральное государственное образовательное учреждение

высшего профессионального образования

Санкт-Петербургский государственный университет

Институт «Высшая школа менеджмента»

**Разработка онтологии для развития системы поиска в компании**

**Газпром Нефть**

Выпускная квалификационная работа

студентки 4 курса бакалаврской программы,

профиль – Информационный менеджмент,

**СУТОРМИНОЙ Екатерины Павловны**

*(подпись)*

Научный руководитель:

Старший преподаватель кафедры

информационных технологий в менеджменте

**ЛЕЩЕВА Ирина Анатольевна**

*(подпись)*

Санкт-Петербург

2021

ЗАЯВЛЕНИЕ О САМОСТОЯТЕЛЬНОМ ХАРАКТЕРЕ КУРСОВОЙ РАБОТЫ

Я, Сутормина Екатерина Павловна, студентка 4 курса направления Менеджмент профиля подготовки – Информационный менеджмент, заявляю, что в моей̆ курсовой̆ работе на тему «Разработка онтологии для развития системы поиска в компании ПАО «Газпром Нефть»» представленной̆ в офис бакалаврской̆ программы для публичной̆ защиты, не содержится элементов плагиата.

Все прямые заимствования из печатных и электронных источников, а также из защищенных ранее курсовых и выпускных квалификационных работ, кандидатских и докторских диссертаций имеют соответствующие ссылки.

Я ознакомлена с пунктом 12.4.14 «Правил обучения на бакалаврской̆ программе ВШМ СПбГУ» «обнаружение в ВКР студента элементов плагиата (контекстуальное или прямое заимствование текста из печатных и электронных оригинальных источников, а также из защищенных ранее выпускных квалификационных работ, кандидатских и докторских диссертаций без соответствующих ссылок) является основанием для выставления за курсовую работу оценки «неудовлетворительно».

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

*(подпись)*

«13» апреля 2021 года

**ОГЛАВЛЕНИЕ**

[Введение 5](#_Toc73655997)

[Глава 1. Обзор теоретических материалов по теме онтологического инженеринга 6](#_Toc73655998)

[1.1. Определение онтологии 6](#_Toc73655999)

[1.1.1. Понятие онтологии 6](#_Toc73656000)

[1.2. Классификация онтологий 8](#_Toc73656001)

[1.2.1. По объекту концептуализации 10](#_Toc73656002)

[1.2.2. По степени формальности 11](#_Toc73656003)

[1.2.3. По уровню выразительности 14](#_Toc73656004)

[1.2.4. По языку представления онтологических знаний 15](#_Toc73656005)

[1.3. Методы разработки онтологий 15](#_Toc73656006)

[1.4. Тезаурусы 18](#_Toc73656007)

[Глава 2. Описание компании ПАО «Газпром нефть» и её дочерних компаний 19](#_Toc73656008)

[2.1. Общее описание компании «Газпром Нефть» 19](#_Toc73656009)

[2.1.1. Создание собственного ИТ-кластера 21](#_Toc73656010)

[2.2. Общее описание компании «Газпром нефть НТЦ» 22](#_Toc73656011)

[2.3. Общее описание компании «Газпромнефть-Цифровые решения» 22](#_Toc73656012)

[Глава 3. Развитие онтологии корпоративного поиска 24](#_Toc73656013)

[3.1 Обзор системы корпоративного поиска «Газпром Нефть» 24](#_Toc73656014)

[3.2. Описание и анализ текущей онтологии в корпоративном поиске 28](#_Toc73656015)

[3.3. Разработка тезауруса предметной области «Геологоразведка» 30](#_Toc73656016)

[3.3.1. Методика построения 32](#_Toc73656017)

[3.3.3. Подготовка неформальной онтологии 36](#_Toc73656018)

[3.3.4. Создание онтологии в Protégé 41](#_Toc73656019)

[Заключение 51](#_Toc73656020)

[Список использованной литературы 53](#_Toc73656021)

# Введение

Сегодня сотрудники организаций, особенно сотрудники крупных компаний, тратят огромное количество времени на поиск информации. По результатам исследований[[1]](#footnote-1), 36% времени уходит на поиск знаний в компаниях. Огромный рост технологий приводит к тому, что накапливается огромные массивы знаний. Проблема заключается в том, что зачастую эти знания так и остаются в необработанном виде, а инструменты для структурирования этой информации попросту отсутствует. Нужно понимать, что компаниям приходится расплачиваться за свое неумение организовать и структурировать свои данные упущенными возможностями. Если внутри компании плохо работает так называемый трансфер информации, компания непременно теряет часть данных, а из-за этого у нее не будет возможности налаживать и оптимизировать свои бизнес-процессы. Некоторые компании прибегают к локальному решению проблемы, они создают отдельные базы знаний, в которых эксперты вручную заносят знания. Но очевидно, это будет не совсем актуально для крупных компаний. Например, таких, как Газпром Нефть. В этой организации объем файлового хранилища уже превышает 350 ТБ.

Целью моей работы является разработка онтологии одной из предметных областей компании «Газпром нефть», которая поможет развитию системы корпоративного поиска.

В своей работе я поставила следующие задачи:

1.Рассмотрение теоретических основ разработки онтологии

2.Рассмотрение принципов работы корпоративного поиска и роль онтологии в не

3.Создание тезауруса выбранной области

4.Рассмотрение и адаптация текущей методологии разработки онтологий в компании

# Глава 1. Обзор теоретических материалов по теме онтологического инженеринга

## Определение онтологии

### Понятие онтологии

Термин «онтология» впервые пришёл из философии и означал учение о бытии и сущем, изучая категории, структуру и закономерности. Сейчас же термин относится и к информационным технологиям. Трактованные термина «онтология» разнится от авторов, времени, контекста и целей его использования.

Основоположником применения понятия «онтологии» в инженерии или информационных технологиях является Грубер, рассмотревший различные аспекты взаимоотношений интеллектуальных систем между собой и человеком и охарактеризовавший онтологию как описание (спецификацию) концептов и взаимосвязей (концептуализации)[[2]](#footnote-2). Онтология – это спецификация концептуализации по определению.

Также можно сказать, термин «онтология» обозначает систему понятий, их свойств и связей между нами. Гаврилова Т.А. определяет онтологию как «точную спецификацию некоторой области, которая включает в себя словарь терминов предметной области и множество связей (типа «элемент-класс», «часть –целое»), которые описывают, как эти термины соотносятся между собой»[[3]](#footnote-3).

Несмотря на то, что в самом определении понятия онтологии содержится объяснение элементов, которые входят в онтологию, я считаю нужным остановиться более подробно на каждом из них. Как уже было упомянуто выше, понятие онтологии зависит от в первую очередь от контекста её использования, то же самое происходит и с составом элементов внутри онтологии. Так, например, связи или отношения могут быть совершенно разного типа, от «является» до «состоит из» и другие. Онтология может состоять из различных типов сущностей или концептов.

Можно выделать следующие виды самых базовых типов сущностей присущих онтологии:[[4]](#footnote-4).

* **Классы**

Классы (classes) – совокупность объектов, обладающих общими свойствами или признаками. В онтологии образуют иерархические отношения. Важное решение при определении моделирования классов: наличие транзитивности (если А подкласс В, а В подкласс С, то А подкласс С). Иногда классы называют понятиями, концептами, сущностями или категориями)[[5]](#footnote-5)

* **Атрибуты**

Атрибуты определяют свойства или признаки объектов определённого класса. Атрибуты также являются классами (или экземплярами). Атрибутами могут быть: имя объекта, его определение, его составные части.

* **Отношения**

Отношения используют для определения того, как объекты связаны друг с другом. Есть отношения «класс-подкласс», определяющие иерархию или «часть-целое» для определения патронимии.

* **Аксиомы**

Под аксиомами понимают логические утверждения, описывающие априорное знание об области и позволяющие при этом делать логический вывод[[6]](#footnote-6). Наличие аксиом влияет на классификацию онтологии, если в онтологии не содержатся аксиомы, то она будет относится к категории легковесной онтологии, с аксиомами же онтология является тяжеловесной[[7]](#footnote-7). Более подробно классификация онтологий будет рассмотрена мною ниже.

* **Экземпляры**

Экземпляры находятся на самом нижнем уровне онтологии, а их добавление в неё называется наполнением. [[8]](#footnote-8)

Онтология может рассматриваться как способ взаимодействия интеллектуальных систем как между собой, так и с человеком [[9]](#footnote-9), ведь любая онтология позволяет правильно соотносить знаки, используемые людьми и компьютерами, с обозначаемыми объектами реального мира. (рис.1.).



1. Использование онтологий для коммуникаций между людьми и/или машинами

[Источник: Кудрявцев Д. В. Системы управления знаниями и применение онтологий: учебное пособие. – 2010. С-57.]

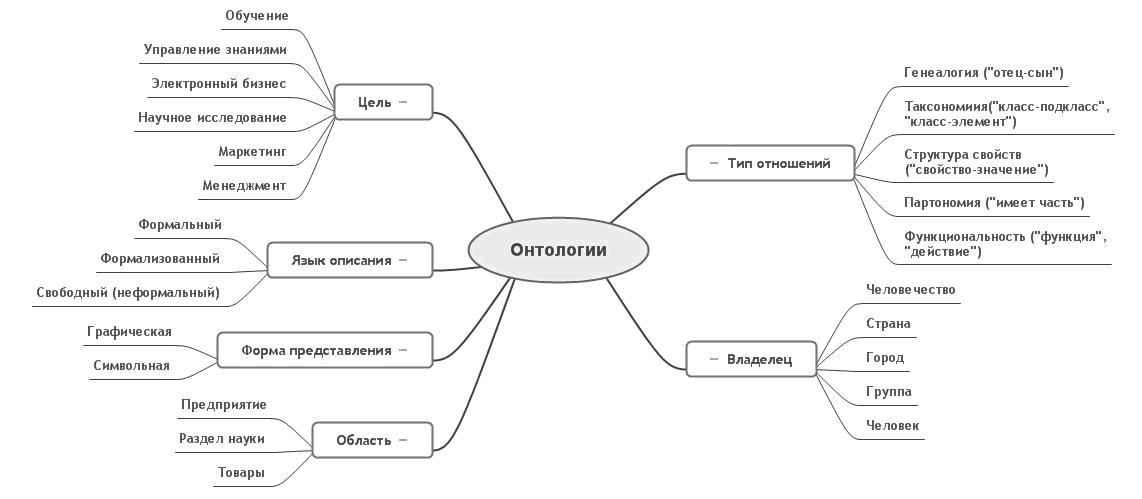
## Классификация онтологий

Считаю важным остановиться более подробно для рассмотрения классификаций онтологии. Мною были рассмотрены базовые определения онтологии, но, как было написано ранее, существует большая зависимость он контекста, в рамках которого онтология разрабатывается.

Изучая зарубежную и российскую литературу, я столкнулась с разными подходами в систематизации знаний об онтологиях. Разные авторы используют разные подходы в систематизации представлений и исследований в области онтологий.

Так, Т.А. Гаврилова предлагает следующие возможные классификации:

* по типу отношений
* по владельцу или пользователю
* по языку описания
* по области применения
* по цели разработки



1. Основные принципы возможных классификаций онтологий

[Источник: Гаврилова Т.А. От инженерии знаний к онтологическому инжинирингу [Электронный ресурс] / Сайт posp.raai.org, URL: <http://posp.raai.org/data/posp2005/gavrilova/gavrilova.html> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)]

Д. В. Кудрявцев позднее расширил и изменил классификацию онтологий, добавив разные типы онтологий по уровню обобщения, глубине проработки и предметной области.



1. Систематизация знаний в области онтологий

[Фрагмент отчета по научно-исследовательской работе «Эталонные модели организации деятельности в государственном секторе», выполненной сотрудниками АНО КМЦ «Бизнес-Инжиниринг» совместно с ИПГМУ ВШЭ, 2006 г: [Электронный ресурс] // Сайт: <http://bigc.ru/>. Режим доступа: <http://bigc.ru/theory/km/onto_technologies.php#1> 22.02.2021 ]

При изучении теоретического материала по онтологическому инжинирингу я отталкивалась от классификации, предложенной в учебнике «Инженерия знаний. Модели и методы: учебник для вузов». Тем не мене, в рамках моей работы считаю нужным остановиться только на некоторых разделах данной классификации, добавив и другие типы критерии выделения типов онтологий.

### 1.2.1. По объекту концептуализации[[10]](#footnote-10)

Построение любой онтологии имеет определенную цель и в зависимости от неё выбираются методы и инструменты построения онтологии.

В рамках классификации по цели создания или уровню обобщения выделяют четыре уровня:

1. **Онтология представления**: связана с концептуализацией формализмов представления знаний,
2. **Онтология верхнего уровня**: повторно используемая в разных предметных областях,

Этот тип онтологий включает в себя очень общие концепты, которые будут актуальны для абсолютно любой предметной области.

1. **Онтология предметной области:** повторно используемая внутри одной предметной области

Этот тип онтологий строится в рамках понятий, их отношения и объектов в модели определенной предметной области.[[11]](#footnote-11) Эта предметная область также называется доменом, одним из самых примеров могут являться предметные онтологии в области медицины, где созданы большие стандартные, структурированные словари, такие как UMLS[[12]](#footnote-12) (Unified Medical Language System), который объединяет и распространяет ключевую терминологию, стандарты классификации и кодирования, а также связанные ресурсы для содействия созданию более эффективных и функционально совместимых биомедицинских информационных систем и услуг, включая электронные медицинские записи. Такие стандартные онтологии домена активно разрабатываются совместными усилиями специалиста из одной области для совместного использования и дальнейшего развития[[13]](#footnote-13).

1. **Прикладная онтология**: нет возможности повторно использовать.

Этот тип онтологий ориентирован на то, чтобы описать концептуальную модель конкретной задачи или приложения. Прикладные онтологии описывают понятия, которые зависят как предметной области, так и от задач, стоящей перед ней. В такой онтологии могут содержаться как общие концепты, характерные для предметной области, так и узкие концепты, отражающие её специфику и которые и не позволяют использовать построенную онтологию повторно.

### 1.2.2. По степени формальности

Важной классификацией онтологий является классификация по степени формальности, предложенная Ласселом и Маггинессом.[[14]](#footnote-14) На рисунке 4 показан онтологический спектр по степени формальности, где точкой представлен определенный вид онтологии. Косая черта разделяет виды онтологий на две группы.

К первой группе относятся:

* каталоги

Или контролируемый словарь – конечный список терминов

* глоссарии

Является списком терминов с их значениями

* отношение «Выше-Ниже»

Представляют из себя тезаурусы, определяющие связи между терминами

* неформальные таксономии.

Которые характеризуются свойством «класс-подкласс»

Во вторую группу входят:

* формальные таксономии

Которые характеризуются не просто свойством «класс-подкласс», но и транзитивностью.

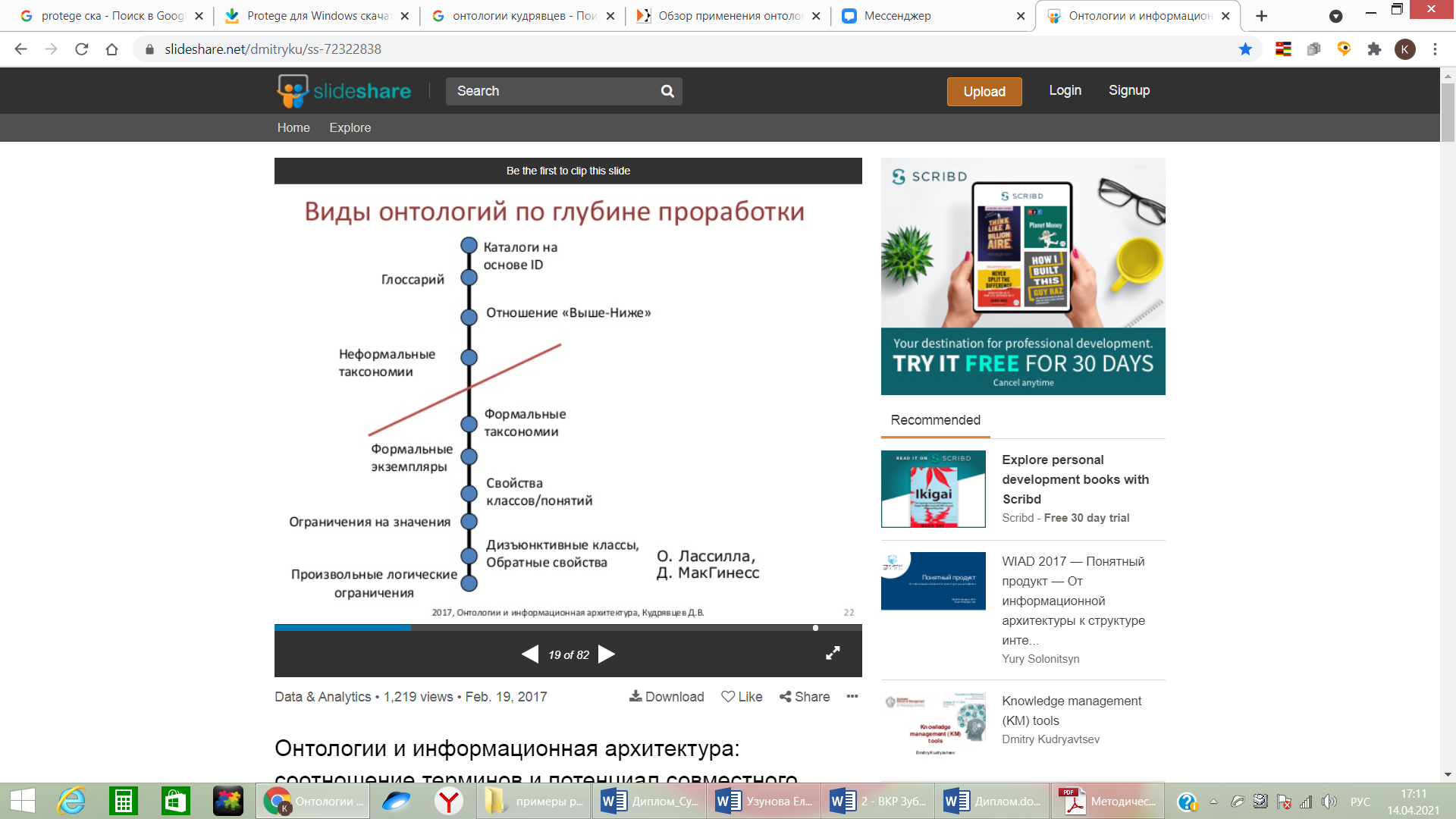
* формальные экземпляры

Включают в себя отношение «экземпляр-класс» с наследуемостью

* свойства классов\понятий

В этом случае онтология имеет классы с информацией о свойствах.

* ограничения на значения
* дизъюнктивные классы, обратные свойства
* Произвольные логические ограничения (аксиомы).



1. Виды онтологий по глубине проработки

[Источник: Lassila O, McGuiness D, The role of frame-based representation on the semantic web //Linköping Electronic Articles in Computer and Information Science. – 2001.]

Группы соответственно разделены по степени формальности, где первая группа является неформальной, то есть «человеко-понятной», а вторая – формальной, то есть «машино-понятной». Различные онтологии могут быть представлены как спектр онтологий, которые зависит детализации исполнения и разработки.

Так как передо мной стояла цель разработать онтологию, представленную в виде тезауруса, более подробно опишу только эту глубину проработки онтологии. При анализе литературы я столкнулась с множеством определение тезауруса, которые сильно отличаются друг от друга в концептуальном плане.

Некоторые источники определят тезаурус как «универсальный словарь, имеющий много входов»[[15]](#footnote-15).

Другие считают, что тезаурус – это «опирающаяся на онтологию организация пространства лексики и, в меру возможности, других компонентов языка.»[[16]](#footnote-16)

В учебно-методическом пособии Доброва и др. «Онтолгии и Тезаурусы»[[17]](#footnote-17) информационно-поисковые тезаурусы рассматриваются как **вид онтологических ресурсов.** Авторы этой книги также считают, что специальные тезаурусы, являясь ресурсом для автоматической обработки текстов, соединяют в себе три существующие традиции в области разработки ресурсов онтологического типа:

* формальные онтологии,
* лингвистические онтологии,
* традиционные информационно-поисковые тезаурусы.

Отношения свойственные для тезаурусов: синонимия, иерархическое отношение (отношения вида «выше-ниже» и ассоциация.

Важно отметить, что я также изучала зарубежную литературу и в оригинальных источниках связи отношения внутри тезауруса обозначены как «broader term-narrower term»[[18]](#footnote-18), что переводится как «шире-уже».

Таким образом, тезаурус – это определенный вид онтологии, имееющий три характеристики:

* Иерархичность (более широкий термин/более узкий термин)
* Ассоциативность (связан с)
* Эквивалентность

### 1.2.3. По уровню выразительности

Вышеупомянутые классификации можно дополнить еще одной – классификацией по уровню выразительности, согласно которой все онтологии делятся на тяжеловесные и легковесные в зависимости от наличия в них вышеупомянутых аксиом или произвольных логических ограничений.

Тяжеловесные онтологии содержат аксиомы, сто позволяет осуществлять онтологическое связывание явно и избежать неоднозначности из-за неправильной интерпретации. Важно отметить, что каждая тяжеловесная онтология может иметь легковесную версию.

Легковесные же онтологии слабо аксиоматизированы, как правило по причине однозначности значение терминов.

### 1.2.4. По языку представления онтологических знаний

Помимо прочего, есть целый ряд языков, с помощью которых может быть описана онтология. Любая онтология может быть написана на формальном и не формальном языке.

Неформальный язык – это естественный язык, будь то русский язык, английский или любой другой. На самом деле, принято также выделять промежуточные варианты: неформально-структурированные онтологии и полуформальные. Я же хочу остановиться на формальном языке, то есть языке, ан котором онтология будет кодироваться для того, чтобы её понимал не только человек, но и машина. Таких языков несколько:

* RDF
* DAML+OIL
* OWL
* KIF
* CycL и другие

В своей работе я кратко опишу два из них.

Первый из них это RDF - Resource Description Framework. Этот язык был разработан консорциумом Всемирной паутины в рамках проекта Семантик-веб (Semantic Web). Язык разрабатывался для описания метаданных документов, размещаемых в Интернете.

В его основе лежит триплет, представляющий из себя субъект, который связан с объектом через предикат. При использовании этого языка утверждения представляют из себя ориентированный граф, где субъекты и объекты – вершины графа, а рёбра отображают отношения.

Язык OWL – Web Ontology Language – является языком следщего поколения и используется для более сложных отношений внутри онтологий.

В основе это языка лежит модель данных «объект — свойство» благодаря чему возможен более продвинутый механизм семантического анализа.

## 1.3. Методы разработки онтологий

На тему онтологического инжиниринга или методов разработки онтологий существует множество рекомендаций и руководств от разных авторов, написанные в разные годы. Однако до сих пор не выделено какой-либо единой общепринятой методологии построения онтологий.

Главная причина этого – большая вариативность видов онтологии. Очевидно, что этапы разработки глоссария и формальной тяжеловесной таксономии будут отличаться по своей сложности. Помимо прочего, необходимо учитывать специфики области использования онтологии и задач стоящими перед ней.

Одной их самых известных методологий построения онтологии является руководство, предложенное Н. Ной и Д. Макгиннес[[19]](#footnote-19). Авторы в своём руководстве описывают общие моменты, необходимые для разработки онтологии и называют свою методологию простой методологией инженерии знаний, использую итеративный подход в разработке онтологий. Сначала происходит черновой просмотр онтологии, а уже потом уточнение в деталях вместе с проверкой и дальнейшей доработкой.

Таким образом, выделяются следующие этапы:

1. Определение области и масштаба онтологии

Для этого необходимо ответить на ряд вопросов

* Какую область будет охватывать онтология?
* Для чего мы собираемся использовать онтологию?
* На какие типы вопросов должна давать ответы информация в онтологии?
* Кто будет использовать и поддерживать онтологию?

1. Рассмотрение вариантов повторного использования существующих онтологий

Всегда стоит проверять открытые различные источники, которые могут как содержать подходящую онтологию, которую можно расширить, дополнить или изменить под те задачи, которые стоят перед онтологией. Инженерия знаний активно развивается, поэтому не стоит недооценивать возможность использования онтологий, созданных другими специалистами. Помимо прочего, нередко в крупных компаниях специалисты разных отделов не знают о текущих наработках внутри организации.

1. Перечисление важных терминов в онтологии

Авторы рекомендуют составить список обширный список всех терминов, которые могут быть использованы в онтологии.

1. Определение классов и иерархии классов

Здесь важно отменить, что существует три подхода для выделения иерархии классов[[20]](#footnote-20):

* Нисходящий подход или «сверху-вниз»

В этом случае разработка начинается с общих понятий, которые потом детализируются более узкими или низкими понятиями.

* Восходящей подход или «снизу—вверх»

При выборе этого подхода всё происходит наоборот, разработка начинается с определения конкретных понятий, которые затем группируются в классы более общих понятий.

* Комбинированный подход

При этом подходе разработки онтологии выбираются главные или основные определения, которые затем можно обобщать и ограничивать. Этот способ сочетает в себе два предыдущих и является наиболее универсальным в том случае, когда явно выделены наиболее заметные термины, которые в дальнейшем будет включать в себя онтология.

1. Определение свойств классов – слотов

Так как классы не дают достаточную информацию, то необходимо определить их свойства для того, чтобы описать свойства классов и экземпляров и

1. Определение фацетов слотов

Следующим этапом является выбор ограничений для слотов. Эти ограничения представляют из себя значения, которые может принимать слот.

1. Создание экземпляров

Этот этап является последним в разработки онтологии. Создаётся конкретный экземпляр выбранного класса вместе со значением слота.

Необходимо отметить, что руководство было написано на основании работы в программе построении и поддержке онтологий в ряде онтологических сред, включая Protege, с которой я в дальнейшем столкнулась, поэтому изучение статьи авторов в дальнейшем помогло быстрее понять принципы работы программы.

## 1.4. Тезаурусы

Мною уже были коротко описаны тезаурусы и их отличие от других уровней онтологического представления знаний, тем не менее, для более глубокого понимания устройства тезаурусов мне было необходимо изучить больше литературы на эту тему, поскольку при написании дипломной работы мне приходилось общаться с техническими специалистами, которые обладают обширными знаниями в этой области. Я не считаю нужным останавливаться более подробно на описании языков кодирования тезаурусов или проводить сравнительный анализ инструментов для работы с ними, при это важно коротко рассказать о самых распространённых моделях тезаурусов:

* SKOS/SKOS-XL
* WordNet/EuroWordNet
* и РуТез

Мною будет описана моедль SKOS, потому что именно она используетя в тезаурусе корпоративного поиска.

Тезаурус организован таким образом,что концепты объединены связями выше/ниже.У каждого концепта существует обозначение имени,другими словами-у них имеется ярлык, который непосредственно может быть написан на разных языках. Ярлыки в модели SKOS это свойства концептов,следовательно у них нет их собственного свойства.Вышеописанное ограничение устраняет специальное расширение для модели SKOS.Что касается SKOS XL , там ярлыки представляют из себя отдельные объекты, и обладают свойствами.

# Глава 2. Описание компании ПАО «Газпром нефть» и её дочерних компаний

## 2.1. Общее описание компании «Газпром Нефть»

ПАО «Газпром нефть»[[21]](#footnote-21) - вертикально- интегрированная нефтяная компания, занимающаяся разведкой, добычей и переработкой нефти и газа, производством и реализацией нефтепродуктов. «Газпром нефть» делает акцент на развитие технологий, внедрение новых решений для достижения поставленных стратегических целей.

**Миссия ПАО «Газпром нефть»2:**

«Развиваться, чтобы развивать мир. Созидать, чтобы гордиться созданным.

Мы создаем ресурсы для будущего, обогащая мир энергией, знаниями и технологиями для уверенного движения к лучшему.»

**Цель компании:3**

В 2020 году «Газпром нефть» работала в сложных условиях снижения цен и спроса на нефть и нефтепродукты на мировом и внутреннем рынках, вызванного ограничительными мерами в условиях пандемии COVID-19. Несмотря на это компания сумела обеспечить непрерывность производственных процессов, сохранила финансовую стабильность и продолжила реализацию стратегических проектов.

***Стратегическая цель «Газпром нефти»****—* стать образцом безопасности, технологичности и эффективности в масштабе всей мировой нефтегазовой отрасли.

1. Основные показатели ПАО «Газпром нефть» в 2020 году

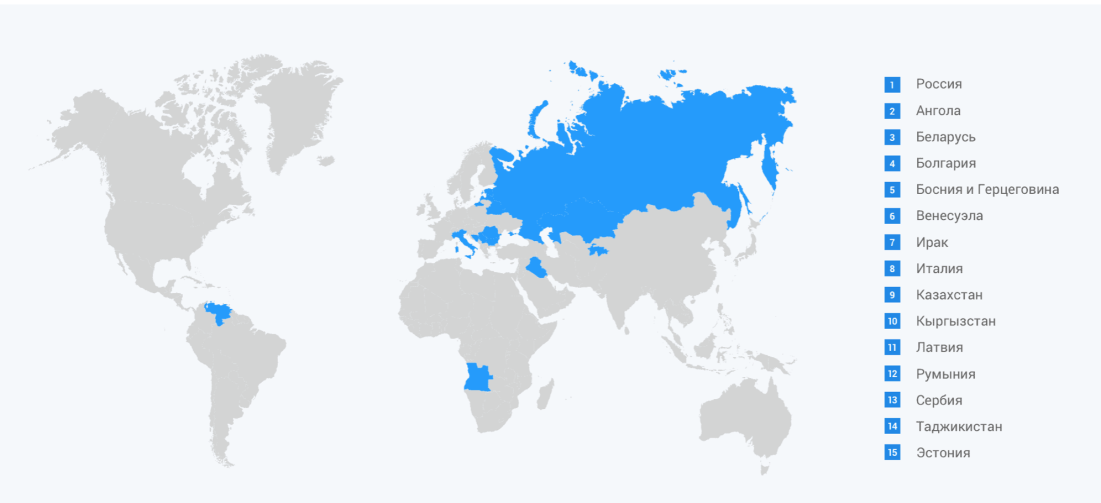
|  |  |
| --- | --- |
| Показатели | Ключевые цифры |
| Чистая прибыль | 117,7 млрд. руб. |
| Объем добычи углеводородов | 96,1 млн. тонн н.э. |
| Объем переработки нефти | 40,4 млн .тонн |

[Источник: [Электронный ресурс], Официальный сайт компании «Газпром нефть». Режим доступа <https://www.gazprom-neft.ru/company/about/at-a-glance/> 15.02.2021)]

Стоит отметить, что компания «Газпром нефть» входит в тройку лидеров российской нефтяной отрасли. Компания владеет более 70 нефтедобывающих, нефтеперерабатывающих и сбытовых предприятий не только на территории России, но и за ее пределами.

Например, первое место по объемам нефтепереработки в России занимает Омский нефтеперерабатывающий завод[[22]](#footnote-22). Также он является технологическим лидером отрасли. Московский нефтеперерабатывающий завод[[23]](#footnote-23) это один из российских лидеров в сфере реализации экологических проектов и ведущий поставщик нефтепродуктов в московском регионе. Нефтеперерабатывающий завод Панчево[[24]](#footnote-24)(Сербия) один из самых современных заводов в Юго-Восточной Европе. Поставщик нефтепродуктов не только на рынок Сербии, но и экспортер топлива в страны Балканского региона.

Продукция компании «Газпром нефть» экспортируется более чем в 50 стран мира. Также существует и достаточно развитая собственная розничная сеть автозаправочных станций на территории Российской Федерации, стран СНГ и Юго -Восточной Европы. Общее их количество достигает более 2000 станций. Количество сотрудников компании превышает 70 тысяч человек.



1. Географическое расположение активов компании

[Источник: [Электронный ресурс], Официальный сайт компании «Газпром нефть». Режим доступа <https://www.gazprom-neft.ru/company/about/map/> 16.02.2021]

«Газпром Нефть» активно развивает ИТ компетенции в своей и дочерних компаниях. Компания занимается развитием способностей к быстрой трансформации, гибкости и адаптивность ИТ-направления компании под конкретные задачи. По мнению департамента информационных технологий, это является важным условием для достижения успеха на инновационном пути развития.[[25]](#footnote-25)

### 2.1.1. Создание собственного ИТ-кластера

«Газпром нефть» создает собственный ИТ-кластер для ускорения цифровой трансформации бизнеса.[[26]](#footnote-26) Деятельность планируемого ИТ-кластера «Газпром нефти» должна будет быть направлена на:

* реализацию корпоративной стратегии развития ИТ-инфраструктуры компании,
* внедрение продуктового подхода
* разработку новых сервисов как для клиентов, так и партнеров компании.

ИТ-кластер формируется на базе двух дочерних организаций компании «Газпром нефти», а именно: «Газпромнефть-Цифровые решения» и «Газпромнефть Информационно-Технологический оператор». «Газпром Нефть» ставит перед собой цель успешного формирования ИТ-кластер. Для этого компания выбрает переход на продуктовую модель, которая будет способствовать не только быстрому запуску ИТ-продуктов, но и долгосрочному развитию ИТ-продуктов в кросс-функциональных командах разработчиков и многочисленных бизнес-подразделений.

Ожидается, что такие внедрения сделают возможными реализацию высококачественных решений в области цифровизации максимально эффективным образом благодаря высоким компетенциям сотрудников в области технологий вместе с понимаем специфики каждого из многочисленных направлений бизнеса компании «Газпром нефть».

Стратегия цифровой трансформации была утверждена в сентябре 2019 года в соответствие с единой корпоративной стратегией развития бизнеса компании «Газпром нефть» вплоть до 2030 года.

## 2.2. Общее описание компании «Газпром нефть НТЦ»

Теперь остановимся на научно-техническом центре ООО «Газпром нефть НТЦ»[[27]](#footnote-27) – одном из дочерних предприятий « Газпром нефти». Научно-Технический Центр «Газпром нефти» был создан в октябре 2007 года, в общей сложности численность сотрудников «Газпром нефть НТЦ» составляет около 1100 человек.

Именно Научно-Технический Центр «Газпром нефти» отвечает за реализацию Технологической стратегии компании. При этом организация совмещает разработку технологий добычи нефти и дистанционное управление высокотехнологичными производственными процессами с научными исследованиями. Так за 2020 году в «Газпром нефть НТЦ» было подано 73 заявки в Роспатент.

Своей целью Научно-Технический Центр «Газпром нефти считает «повышение нефтедобычи и ее эффективности за счет внедрения новых технологий и проектных решений на месторождения «Газпром нефти»»[[28]](#footnote-28).

## 2.3. Общее описание компании «Газпромнефть-Цифровые решения»

Компания ООО «Газпромнефть-Цифровые решения» является корпоративным интегратором компетенций по цифровому развитию. Компания занимается работой в трёх основных направлениях:

* Корпоративные системы управления,
* Разработка программного обеспечения,
* Информационные технологии.

Компания «Газпромнефть-Цифровые решения» была создана в мае 2020 года путём трансформирования из компании «Информационно-технологическая сервисная компания», которая также была дочерней компанией «Газпром Нефти». История компании начинается в 2002 году, когда произошло объединения трудовых коллективов подразделений ИТ, связи и автоматизации двух крупнейших групп предприятий российской экономики: ОАО «Газпром нефть» и ОАО «СИБУР Холдинг».

Благодаря своим финансовым возможностям и высококвалифицированным специалистам с компетенциями в различных областях компания «Газпромнефть-Цифровые решения» реализовывает различные виды проектов – от внедрения программных продуктов до сервисного обслуживания техники и обеспечения каналов связи.

Проекты цифровой трансформации, проводимые благодаря усилиям специалистов «Газпромнефть-Цифровые решения» в таких областях, как:

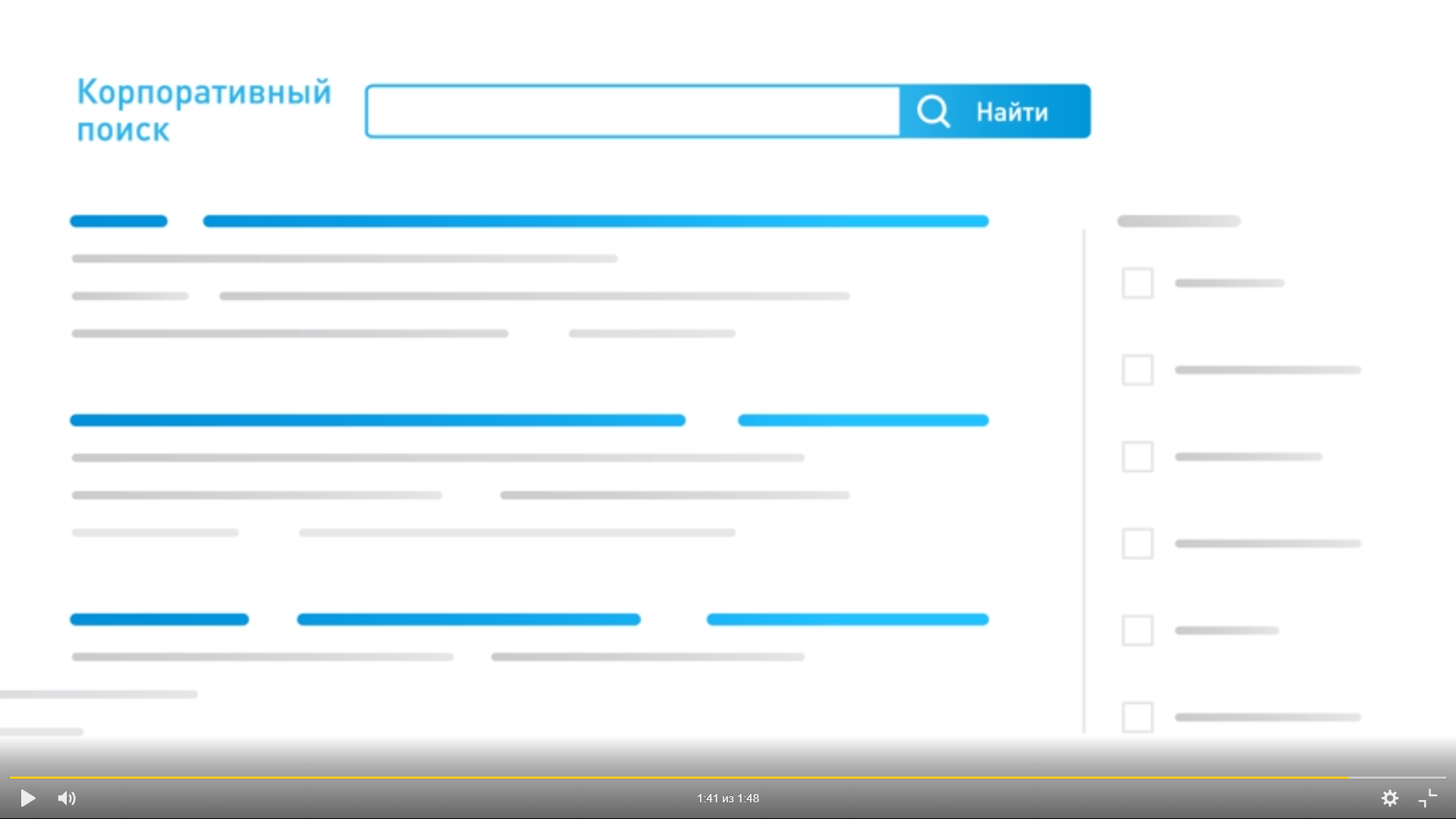
* блокчейн
* искусственный интеллект
* средства коллаборации
* носимые технологии
* искусственный интеллект
* роботизация и аддитивные технологии.

Разработки компании высоко ценятся профессиональным сообществом лидеров цифровой трансформации России. Компания часто получает награды за реализацию вы ИТ проектов. Так цифровые инструменты компании, позволившие обеспечить эффективную удаленную работу сотрудников компании в связи с пандемией, управление электронного документооборота, признаны лучшими проектами 2020 года.

# Глава 3. Развитие онтологии корпоративного поиска

## 3.1 Обзор системы корпоративного поиска «Газпром Нефть»

Умение эффективно работать с данными и извлекать их них нужную информацию становится всё более актуально с каждым днём. Без учета дочерних компаний в «Газпром нефти» объем хранилища корпоративных файлов давно превзошёл размер в 350 ТБ.[[29]](#footnote-29) Найти нужную информацию в таком объеме сложно. Поэтому в 2017 году компания «Газпром нефть» активно занялась решением проблемы быстрого и удобного доступа к накопленным знаниям.



1. Шаблон, иллюстрирующий корпоративный поиск в компании

[Источник: https://disk.yandex.ru/i/yjq6vL2ejZCHlA]

Корпоративный поиск представляет из себя точка доступа к накопленным в «Газпром нефти» корпоративным знаниям. Проект корпоративного или как его ещё называют когнитивного поиска основан на семантической обработке и кластеризации данных. Благодаря чего формируемая при индексировании текстов сематическая модель позволяют выдавать результаты, относящиеся к нужному пользователю тематике. Система позволяет производить поиск по различным источникам от внутренних приказов, до определений и данных сотрудников компании «Газпром-нефть».

Важную роль в решении проблемы поиска документов сыграл Научно-Технический Центр компании, который сначала создал прототип корпоративного поиска. Этот прототип основывался на когнитивном анализе и проект по его реализации имел название «когнитивный поиск».

Когнитивная система анализирует не просто не буквы в словах, а смысл слов, предложений и даже целых текстов, учитывает морфологию слов в запросе, понимает сокращения и специализированные термины. Например, если пользователь введёт запрос «легкие фракции нефти», то он получит результаты не только с фразой «легкие фракции нефти», но со словами «бензин» и «керосин». Потому что бензины и керосин непосредственно связаны с лёгкими фракциями нефти и это отражено системе.

Процесс создания корпоративного поиска по словам самих специалистов, работающих над этим проектом, был достаточно сложным. Все сложности они разделили на несколько блоков: [[30]](#footnote-30)

1. Процесс обработки данных. За 25-летнюю историю компании накопились колоссальные массивы документов, а это более 350 ТБ данных. Данные очень разрозненные, поэтому требуется провести большую работу по аналитике и гармонизации входных данных, чтобы представлять в поиске в едином формате поисковой выдачи совершенно различные типы объектов и документов.
2. Интеграционное взаимодействие. Система выступает агрегатором информации, что подразумевает большое количество интеграционных взаимодействий — важно проработать коннекторы к различным системам, минимизировать время на подключение нового источника и возможность широкого охвата внутренних систем. Требуется формирование оптимального плана обновления данных для балансировки нагрузки на сервера систем, а также разработка универсального интерфейса для взаимодействия с источниками.
3. Сложность в организационных процессах. У проекта очень много участников и заинтересованных сторон, поэтому необходимо очень четко выстраивать коммуникации, синхронизировать совместные релизы с командами-участниками и своевременно проводить демо. Эти задачи требуют высокой мотивации, слаженной работы команды и постоянной координации с внешними участниками. Также очень важно поддерживать репутацию — для формирования положительного имиджа инновационного продукта необходимо обеспечивать постоянно высокий уровень доверия к поиску и работать над релевантностью выдачи.
4. Технологические сложности. Обработка текстов — это очень активно развивающаяся отрасль, где постоянно появляются новые ресурсоемкие технологии и новое ПО, что требует постоянной актуализации и применения этих технологий в проекте с учетом его архитектуры. Объем данных в системе постоянно растет, поэтому необходимо оптимизировать выделенные мощности, решая инфраструктурные и архитектурные задачи. Важным правилом в разработке является также соответствие требованиям ИБ — в проекте разрабатываются внутренние технологии разграничения прав доступа к документам и объектам системы на основании ролевой модели систем источников данных.

Более подробно с техническими характеристиками можно в таблице ниже..

1. Использованное программное обеспечение

|  |  |
| --- | --- |
| **Разработка** | **Программное обеспечение** |
| ***Бэкенд*** | Scala, Java, Python |
| ***Фронтенд*** | ReactJS |
| ***База данных*** | PostgreSQL, MongoDB |
| ***Поиск*** | ElasticSearch |
| ***Анализ данных и языка*** | Apache Spark, Tika, Tesseract OCR, CRF, rule-based over LSA |
| ***Семантическая обработка*** | LSH, LSA, Word2vec, doc2vec/paragraph2vec |
| ***Инфраструктура*** | 4 сервера приложение и БД; 3 сервера обсчет данных, вычислительный кластер. Мощность каждого сервера: 16 ядер, 32ГБ оперативной памяти, 1ТБ HDD. |

[Составлено автором по проведенному анализу с использованием корпоративных данных компании «Газпром нефть»]

Несмотря на все сложности, с которыми столкнулась команда Научно-Технического Центра «Газпром нефти» в 2018 году проект получил награду конкурса «Проекта года» сообщества ИТ-директоров России Global CIO в специальной номинации «Выбор Global CIO». [[31]](#footnote-31) А в январе 2019 года проект получил награду в специальной номинации «Выбор Global CIO» конкурса «Проект года».[[32]](#footnote-32) В среднем время поиска сократилось на 35%.

Спустя 3 года корпоративный поиск доступен каждому сотруднику компании, он постоянно развивается, добавляю новые системы и источники. Когнитивная система поиска обеспечивает релевантную выдачу информации при поиске. Корпоративный поиск сделал процесс доступа к накопленным знаниям быстрее и удобнее засечёт того, что система не только учитывает морфологию вводимого запроса, но и понимает ряд сокращений и специализированных терминов.

В ноябре 2020 года команда развития корпоративного поиска анонсировала важное обновление в системе корпоративного поиска – концепты. Долгие месяцы команда вместе с экспертами активно собирала тезаурусы предметных областей, чтобы обучить корпоративный поиск понимать корпоративный язык.

Концепты – это механизм, позволяющий учитывать корпоративный язык, в частности тезаурусы предметных областей, при разборе пользовательского запроса и формировании выдачи релевантных документов.[[33]](#footnote-33)

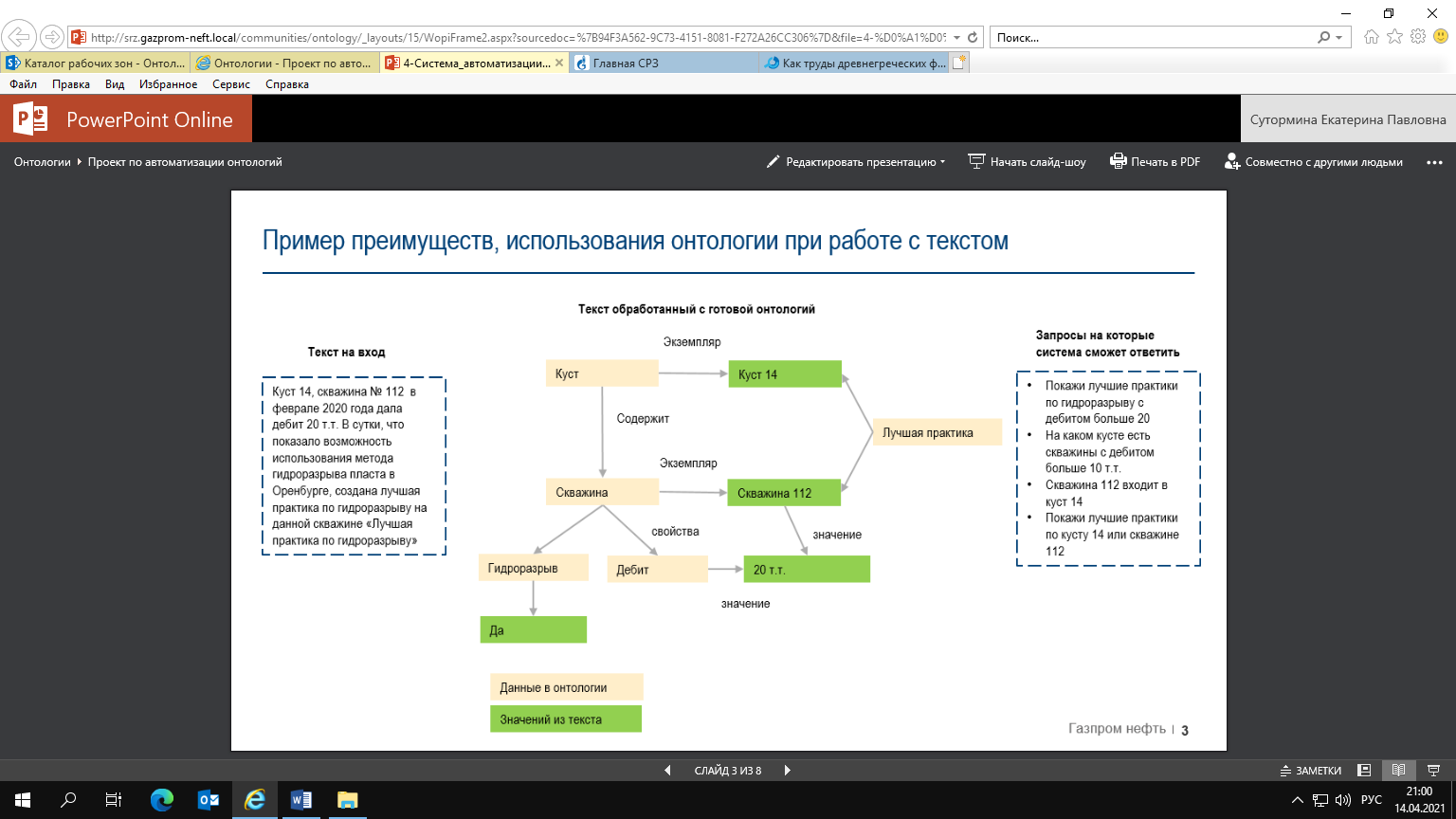
Если в запросе пользователя корпоративной поиск находит наличие терминов, определений или других контекстных слов, он предлагает расширить результаты поиска концептами, то есть смежными понятиями предметной области. Например, если пользователь ищет понятие «ПК», то корпоративный поиск предлагает сделать выбров с помощью концептов, помогая поисковой системе понять, что он имеет в виду – проектную команда или персональный компьютер.

Сейчас блок с выделенными концептами располагается в двух местах поиска:

1. в блоке с фильтрами после ввода запроса (позволяет изменять поисковую выдачу за счет выбора связанных концептов);
2. на карточке документа (для расширения контекста документов и лучшего понимания самих концептов).

## 3.2. Описание и анализ текущей онтологии в корпоративном поиске

Как я писала выше, одним из последних обновлений корпоративного поиска является сформированный командой тезаурус, который нужен системе для обучения терминологии нефтегазовой отрасли и особенностям корпоративного языка в компании «Гапром Нефть». Основой этого тезауруса и должны быт концепты, позволяющие выдавать пользователю релевантные результаты.



1. Пример преимуществ использования онтологии при работе с текстом

[Источник: Внутренняя документация компании «Газпром Нефть»]

Предметную онтологию можно рассматривать как фильтр, который позволяет семантическому ядру поиска видеть текст под определенным углом зрения, а пользователю — получать только релевантные данные, очищенные от информации, не представляющей для него интереса.

Создание онтологии заняло много несколько месяцев. Представители компании выделяют следующие этапы его создания:

* Построение базового "каркаса" для онтологии, в результате которой ыбла сформирована базовая онтологию предметной области.
* Для дальнейшего управления инструментом был создан механизм для ручного ввода новых слоев онтологии и редактирования связей и API[[34]](#footnote-34), который позволяюет получать информацию из онтологии и добавлять информацию в неё.
* После этого был настроен автоматический поиск связей. Благодаря этому этапу, опираясь на контексты и уже сформированные списки объектов, стало возможным строить гипотезы по отношению к новым объектам и фактам в тексте, а далее верифицировать или отклонять эти гипотезы.
* Было организовано триплетное хранилище, чтобы данные в онтологии хранились в виде структуры, совместимой с одной из разновидностей формата OWL.
* Затем было необходимо связать разработки с рабочей инфраструктурой корпоративного поиска и реализовали выполнение запросов оттуда к онтологии на специализированном языке SPARQL с учетом данных в запросе.

Мною было проведено интервью с ведущим аналитиком отдела развития систем корпоративного поиска, а также изучение внутренних и внешних ресурсов на тему корпоративного поиска. Таким образом, была составлена таблица с анализом онтологии в корпоративном поиске.

1. Краткое резюме с информациоей об онтологии

|  |  |
| --- | --- |
| Область применения: | Обучения системы корпоративного поиска особенностям корпоративного языка и терминологии нефтегазовой отрасли |
| Использование: | Концепты – это механизм, позволяющий учитывать корпоративный язык, в частности тезаурусы предметных областей, при разборе пользовательского запроса и формировании выдачи релевантных документов. там при поиске как бы автоматически выбирается какой-то один концепт, но также есть связанные концепты, и эти связанные концепты предлагаются на основе связанных экземпляров в тезаурусею  Запрос пользователя разбирается на концепты, если в его запросе есть концерты, они сопоставляются с тезаурусом и подтаскиваются в альтернативные названия, связанные концепты, |
| Среда разработки | Внутренний административный раздел корпоративного поиска, позволяющие вручную редактировать тезаурус. Возможность визуализации тезауруса внутри корпоративного поиска не предусмотрена. |
| Оценка готовности тезауруса и стадии разработки | Тезаурус требует большой работы над ним ещё очень много нужно поработать, чтобы он работал на 100%. Сейчас уже собрана большая основа всей терминологии нефтегазовой промышленности. Собрано достаточно много связей, которые помигают сейчас искать многие термины. Около 30% это функционала сейчас используется. |
| Уровень обобщения: | Онтология предметной области |
| Глубина проработки: | Тезаурус |
| Объём: | Несколько тысяч концептов, но при этом связь определена только между 500 из них. |
| Иерархичность: | Тип связи “уже, чем - шире” |
| Ассоциативность: | Тип связи “Связан с” |
| Эквивалентность: | Альтернативное название |
| Оценка тезауруса | Не происходит |

## 3.3. Разработка тезауруса предметной области «Геологоразведка»

Текущий тезаурус включает в себя понятия и определения, относящиеся к нескольким предметным областям: от административно-хозяйственного обеспечения и управления проектами до экономики и финансов. Однако тезаурус плохо развит и значительная часть понятий, хранящихся в словаре в виде концептов не имеют никакой связи между собой, при том что связи – один из атрибутов тезауруса и именно они помогают системе корпоративного поиска выдать пользователю наиболее точный результат поиска среди сотен документов и ресурсов. Поэтому перед мной стояла задать сделать поиск в системе корпоративного более эффективным благодаря концептам, содержащимся в тезаурусе.

Проведя интервью с ведущим аналитиком отдела развития систем корпоративного поиска и анализа контента, стало ясно, что представители компании не видят смысл перегружать тезаурус новым типами связей. Перед отделом стояла другая проблема – нехватка ресурсов для наполнения тезауруса новыми терминами как более широких, так и узких, и добавлений связей среди существующих. Помимо прочего, развитие тезауруса происходит вручную и занимает значительное время за счёт этого.

Таким образом, передо мной стала задача развития текущего тезауруса корпоративного поиска. Мною было принято решение выбрать одно из направлений деятельности компании «Газпром нефть» для которого я составлю отдельный тезаурус предметной области, который потом станет частью большого общего тезауруса в корпоративном поиске. Для написания дипломной работы я проходила практику в компании, выполняла различные задачи, которые напрямую не были связаны с разработкой тезауруса в корпоративном поиске. Так одной из задач было проведение интервью для представителей разных отделов компании для получения общих сведений об использовании онтологий в компании и степени их развития. На основе этого интервью была составлена таблица со списком заинтересованных сторон в развитии онтологий.

В этом интервью приняла участие Идрисова Анастасия Геннадьевна, являющаяся на момент проведения интервью руководителем по разработке продуктов. Анастасия Геннадьевна рассказала о работе над проектом по геологоразведке и о дальнейших планах по возможном использовании онтологии в нём. У представителей проекта есть большая заинтересованность в использовании онтологии, и они в дальнейшем хотят использовать её, но пока не понимают, как устроен этот инструмент и каким образом он может быть полезен по причине начальной стадии развития проекта. В целом, руководитель проекта выразил большую заинтересованность в развитии онтологии и планы на будущее сотрудничество с представителями корпоративного поиска в этом вопросе.

Таким образом, мною была предметная область геологоразведки для создания тезауруса. Это случилось по нескольким причинам:

* Тезаурус содержит маленькое количество связанных концепций по направлению геологоразведки
* Сами сотрудники компании считают это направление работы компании одним из основных
* Сотрудники отдела являются активными пользователями корпоративного поиска
* Представители отдела выразили заинтересованность в сотрудничестве
* Мои разработки могут быть полезны не только для развития системы корпоративного поиска, но и могут быть полезны для проекта по геологоразведки.

### 3.3.1. Методика построения

Прежде чем строить онтология мне было необходимо выбрать методику её построения. Как уже было упомянуто выше, нет единого универсального стандарта, по которому можно строить онтологии, но тем не менее, есть общие шаги, которые должны быть выполнены. «Методология и технология проектирования, разработки и использования онтологий для структурирования и тиражирования знаний в различных предметных областях и приложениях» называются онтологическим инжинирингом.

Так, Т.А. Гаврилова предлагает следующий алгоритм создания онтологии[[35]](#footnote-35):

1. Формирование глоссария предметной области.
2. Установление связей между понятиями глоссария и их визуализация.
3. Категоризация понятий и формирование мета-понятий (снизувверх).
4. Детализация (сверху-вниз).
5. Ре-инжиниринг (уточнение, разрешение противоречий, синонимии, избыточности, перестройка, дополнение).

Д.В. Кудрявцев рекомендует начинать онтологический инжиниринг с определения области и масштаба онтологии, предлагая следующие шаги:

1. Определение области и масштаба онтологии
2. Рассмотрение вариантов повторного использования существующих онтологий
3. Перечисление важных терминов в онтологии
4. Определение классов и иерархии классов
5. Определение свойств классов (слотов)
6. Определение ограничений на значение свойств
7. Создание экземпляров.

Также аспирант Высшей Школы Менеджмента Алёна Беглер разработала рекомендации по построению онтологий в компании «Газпром нефть» специально для сотрудников отдела системы распространения знаний.

Рекомендованные шаги по разработки включают в себя следующие шаги[[36]](#footnote-36):

1. Определить цель разработки онтологии (должна определяться от цели проекта)
2. Определить, как онтология будет использоваться в проекте
   * В форме сценариев использования
   * В форме вопросов проверки компетентности (вопросов, на которые должна быть способна ответить онтология)
3. Определить связь с другими онтологиями в компании
   * Того же уровня из смежных областей
   * Более абстрактных и конкретных из той же области
4. Определить принципы внедрения будущей онтологии   
   (какие задачи и кем будут решаться с помощью неё на местах
5. Оценить количество времени и ресурсов
6. Оценить необходимость вовлечения других заинтересованных лиц
   * Конечные пользователи
   * Менеджеры и аналитики
   * Архитекторы (от баз данных до системных)
7. Определить используемые ресурсы
8. Выбрать стиль моделирования
9. Подготовка полу-формальной модели
10. Реализовать в формальном языке

Все описанные методы выше носят общий характер и не полностью соответствуют условиям создания тезауруса для корпоративного поиска. Однако, именно на основании этих рекомендаций я определила для себя список вопросов необходимых мне для понимания этапов построения тезауруса и выделила основные этапы.

Таким образом, я выделила основные принципы, на которые полагалась при построении:

* Определить цели разработки онтологии
* Определить, как онтология будет использоваться в проекте
* Определить связь с другими онтологиями в компании
* Определить внутренние и внешние ресурсы
* Выбор стиля моделирования словаря и само моделирование
* Разработка неформальной модели онтологии
* Корректировка онтологии и последующее её утверждение.

Мною была составлена таблица с общей информацией об этом.

1. Общая информация касательно разработки тезауруса

|  |  |
| --- | --- |
| Цель разработки тезауруса | Развитие корпоративного поиска за счёт добавление в тезаурус новых концепций, которые помогут пользователю быстрее находить нужный документ. |
| Область | Предметная область «Геологоразведка» |
| Роль онтологии: | Надстройка над глоссарием |
| Масштаб | Тезаурус не должен быть слишком большим и перегруженным, поэтому я ставила перед собой цель в 100-150 концептов. Не был передан файл со списком из 300 терминов этой предметной области, разработанный специалистами корпоративного поиска ещё на начальном этапе его создания. |
| Возможность использования повторной онтологии | В компании нет онтологии подходящей для повторного использования |
| Связь с другими онтологиями или проектами | Корпоративный поиск связан с проектом системы распространения связей, но на момент прохождения практики онтология для системы распространения связей находилась на этапе разработки. |
| Вовлечение других лиц в разработке | Специалисты отдела корпоративного поиска для получения доступа к администрированию корпоративного поиска, эксперты отдела геологоразведки для верификации тезауруса. |
| Структура тезауруса | Тезаурус должен включать в себя:   * Концепт * Перевод концепта на английском языке * Определение * Синоним * Отношение «шире-уже». При этом гипероним (более широкое понятие) может быть только одно в этой связи * Отношение «связан с». Связанных терминов должно быть не более 15. |
| Cтилm моделирования | Из середины или от критических терминов –позволяет создавать сбалансированные по уровню абстракции онтологии. Так как был обозначены критические термины для тезауруса, был выбран именно этот способ. |
| Определение используемых ресурсов | Внутренние: бизнес глоссарий, запросы пользователей, внутренняя документация компании, эксперты геологоразведки  Внешние: большая энциклопедия нефти и газа, учебники по геологоразведки, Библиотека нефтегазового инжиниринга, cловари Шлюмберже, cтандарты ISO по нефтедобыче, разработки POSC Caesar, открытые источники, относящиеся к теме |
| Процесс формализации тезауруса | Тезаурус будет вручную добавлен в систему корпоративного поиска благодаря механизму для ручного ввода новых слоев онтологии и редактирования связей и API (Американский институт нефти – одна из авторитетных коммерческих организаций, которым принадлежит ведущая роль в классификации моторных масел и установлении требований к ним.) , позволяющий получать информацию из онтологии и добавлять информацию в неё. |

[Составлено автором по проведенному анализу с использованием корпоративных данных компании «Газпром нефть»]

Таким образом, в своей работе я выделила 3 главных этапа, которые стояли передо мной:

1. Подготовка неформальной модели, которую затем будут верифицировать эксперты из геологоразведки
2. Создание онтологии в Protégé для визуализации онтологии и предтавленияя её комиссии
3. Добавление онтологии в корпоративный поиск.

### 3.3.3. Подготовка неформальной онтологии

После того, как мною были определены цели, этапы, заинтересованные стороны и условия построения онтологии, я могла начинать разработку неформальной онтологии, то есть онтологии на естественном языке. Для её построения мне необходимо было заполнить следующую таблицу в Excel:

1. Форма для неформального тезауруса

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Определение | Концепт | Гипероним | Синонимы | Связаные  термины |
| Эквивалент Доказанных запасов; наименьшая оценка Запасов | 1Р |  |  |  |
| proved + probable - эквивалент суммы Доказанных и Вероятных запасов; оптимальная оценка Запасов | 2Р |  |  |  |
| эквивалент суммы Доказанных, Вероятных и Возможных запасов; наибольшая оценка Запасов | 3Р |  |  |  |
| Field Size Distribution\r\nфункция распределения месторождений по величине запасов | FSD |  |  |  |

[Составлено автором с использованием корпоративных данных компании

«Газпром нефть»]

Как мы видим, таблица состоит из 5 столбцов:

* Определение
* Концепт
* Гипероним
* Синонимы
* Связаные термины

Определения

Разработка онтологии началась с такого простого способа по нескольким причинам. Во-первых, это наиболее простой способ визуализировать все элементы тезауруса. Во-вторых, экспертам предстояло верифицировать, то есть проверить на правильность, как понятия, так и связи, синонимичность и определения, им некогда разбираться в сложных программах, а Excel им понятен и доступен.

В эту таблицу входило 336 концептов, каждому из которых был дано определения. Эти концепты представляли из себя как названия внутренних документов, так и общие понятия, характерные для предметной области геологоразведки. В системе Газпром нефти было несколько тезаурусов, которые объединили. Один - в системе распространения знаний, другой - в системе закупок, и ещё в системе операционной деятельности. Это были глоссарий, по сути, без каких-либо связей, просто термины и определения. Которые были объединены, затем были найдены похожие термины, найдена связь между ними и началась разработка иерархической структуры.

Как я уже упоминала выше, мною был выбран метод моделирования из середины, суть которого состоит в том, что процесс построения онтологии начинается с выделения ключевых или критических терминов. Список этих концептов и представлял из себя списком с критическими терминами, на которые я ориентировалась и которые являются по своей сути являются каркасом будущей онтологии.

Важно упомянуть, что в этой таблицы нет такого столбца как перевод на английский язык, который, тем не менее есть в корпоративном поиске, поскольку были использованы не все термины и для экономии времени перевод был произведен только для тех концептов, которые были добавлены в тезаурус.

Таким образом, началась самая трудоёмкая и затратная по времени часть моей работы, поиск связей и отношений между выделенными концептами. Для того, чтобы корректно определить связи необходимо было понимать значение терминов, в этом мне помогли представленные определения концептов, а также внутренняя документация компании вместо с другими внешними источниками: большая энциклопедия нефти и газа, учебники по геологоразведки, статьи и многие другие открытые источники. Несмотря на то, что этот файл был отправлен экспертам геологоразведки для верификации и утверждения, важно было также оценивать достоверность источников при их выборе.

Мой алгоритм работы по нахождения связей между концептами состоял из следующего: я читала определение термина, основываясь на него находила гипероним, то есть более широкий термин среди представленных, либо добавляла новый концепт, который стал новым концептом. Затем через поиск я находила определение, содержащее в своём определение концепт. Например, если я знала, что карта концепт «карта нефтегазоносности» связан с концептом «оперативный подсчет запасов», то я понимала, что нужно теперь смотреть информацию по концепту «оперативный подсчет запасов» и он точно будет уже, чем концепт «подсчёт запасов», который также был в списке критических терминов. Таким образом, я несколько раз перечитывала все концепты, многие из них были расположены в списке рядом не просто так, а потому что тоже были связаны, что в целом помогало поиску связей.

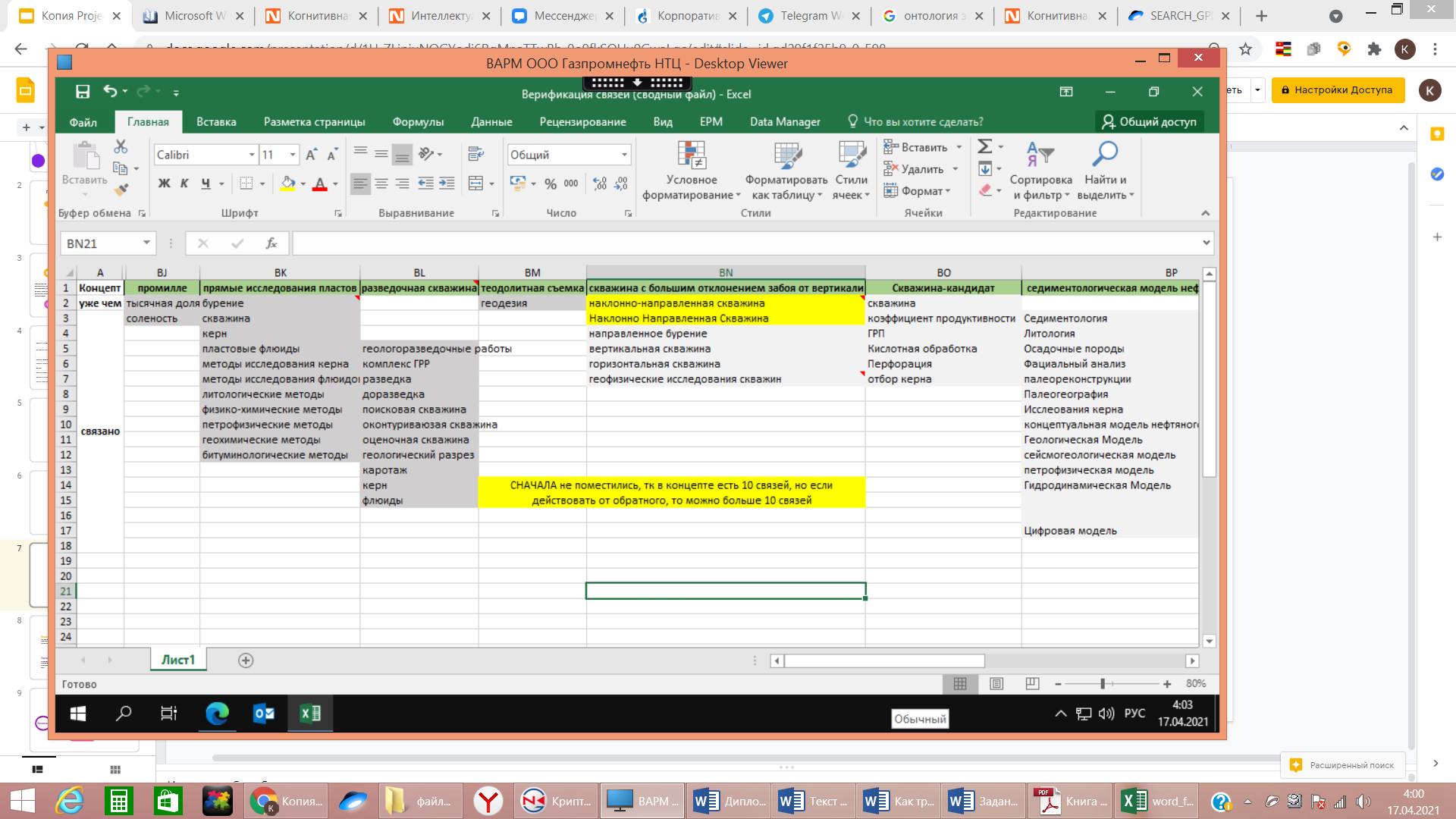
1. Отрывок заполненной таблицы для формального тезауруса

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Определение | Концепт | Гипероним | Синонимы | Связаные термины |
| Эквивалент Доказанных запасов; наименьшая оценка Запасов | 1Р | Классифи-  кация PRMS | Доказанные запасы, Наименьшая оценка запасов | 2P, 3P, Индекс Точности Определения Запасов |
| proved + probable - эквивалент суммы Доказанных и Вероятных запасов; оптимальная оценка Запасов | 2Р | Классифи-  кация PRMS | Оптимальная оценка запасов | вероятные запасы |
| эквивалент суммы Доказанных, Вероятных и Возможных запасов; наибольшая оценка Запасов | 3Р | Классифи-  кация PRMS | Наибольшая оценка запасов | вероятные запасы |
| NetPresentValue; перев. с англ. чистая приведенная стоимость свободного денежного потока'; 'NetPresentValue | NPV | Геолого Экономическая Оценка | NetPresentValue | Геолого Экономическая Оценка, AC (Actual cost), Технико Экономические Показатели |
| уточнение фактических объемов количества и качества ПНГ на месторождениях | аудит  добычи  ПНГ | Попутный нефтяной газ |  | нефтеотдача пластов |
| то природный углеводородный газ, растворенный в нефти или находящийся в «шапках» нефтяных и газоконденсатных месторождений. | Попутный нефтяной  газ | природный углеводород-ный газ | ПНГ | нефтяные месторождения, газоконденсатные местоположения, конденсат |
| форма отчетности «Сведения о состоянии и изменении запасов нефти газа конденсата этана пропана бутанов серы гелия азота углекислого газа за \_\_\_ год» утвержденная постановлением Госкомстата России от 18 июня 1999 № 44 (ред. Постановлений Госкомст | баланс  запасов  УВС и сопутст-вующих компонентов | подсчет  запасов |  | объект оперативного изменения запасов УВС, плей, конденсат |
| вероятность того, что залежь УВ будет обнаружена, запасов достаточно для получения притока УВ при испытании скважины. Оценивается для пласта, структуры, участка, кластера, плея. Для структуры, участка, кластера оценивается как «1-шанс общего неуспеха»; для плея gCoS оценивается как среднее значение | вероятность геологического  успеха |  | Pg, gCos | плей,, объект оперативного изменения запасов УВС, скважина, риск геологический, месторождение, подготовленное к разработке |
| вероятность того, что скопление УВ будет обнаружено, качество и объем запасов будут достаточными для приобщения к общему профилю добычи. Оценивается для структуры, лицензионного участка | Вероятность экономического  успеха локальная | вероятность геологического успеха |  | Геолого Экономическая Оценка, AC (Actual cost), |
| недоказанные запасы, которые, исходя из анализа инженерно- геологических данных, более вероятно извлечь. В этом смысле при использовании вероятностных методов должна существовать вероятность не менее 50 % того, что фактически извлеченные объемы будут равны или больше суммы оценок доказанных и вероятных запасов | вероятные запасы | вероятность геологического успеха |  | 1Р, 2Р, вероятность экономического успеха локальная, категория запасов, коэффициент нефте/газоотдачи |

[Составлено автором самостоятельно на основе проведенного анализа]

Важно сказать, что в целом создание тезауруса и поиск связи в нём достаточно трудоёмкий процесс и тот формат, в котором это сейчас реализовано в компании делает его ещё более трудозатратным. Возможно, в этом и есть одна из основных причин почем развитие тезауруса занимает в компании не один месяц и эксперты тратят несколько месяцев на поиск связей среди сотни концептов. Именно нехватка времени экспертов повлияло на то, что разработкой тезауруса этой предметной области занялась я.

После того, как такая неформальная форма тезауруса была создана, была необходима верификация со стороны экспертов. Несмотря на то, что экспертом было необходимо только проверить правильность связей этот процесс занял не одну неделю в виду их большой занятости. Эксперты внесли ряд корректировок, после чего неформальная модель была готова.



1. Скриншот файла, отправленного экспертам на экспертизу

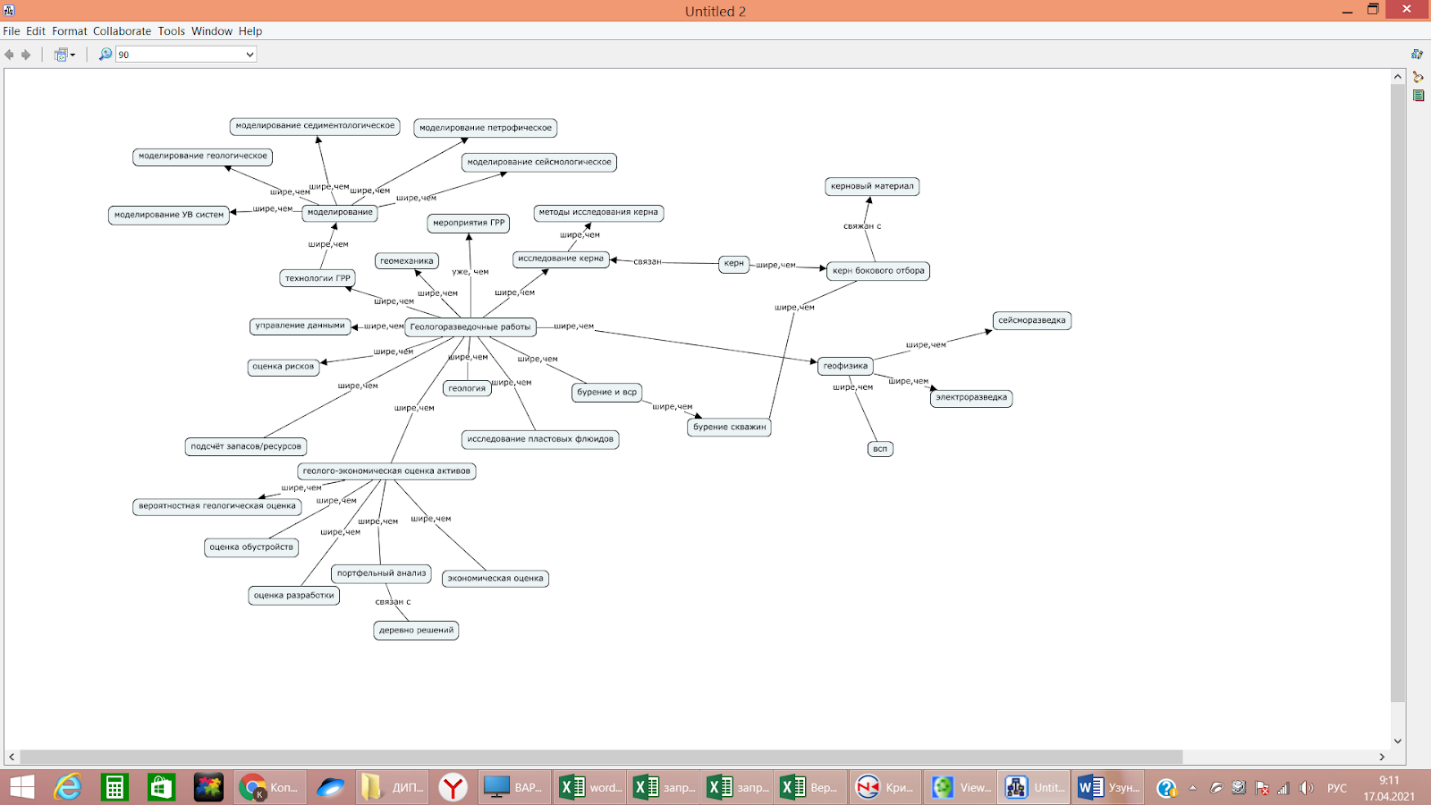
[Источник: внутренняя документация компании «Газпром нефть»]

Важно отметить, что созданный мною тезауруса в cотню концептов не является конечным и, безусловно, может быть и должен быть в дальнейшем доработан. Во-первых, я была ограничена во времени, во-вторых, необходимо учитывать уровень моей экспертизы в предметной области геологоразведки, в-третьих, важно понимать, что несмотря на корректировки, произведенные экспертами, разработка тезауруса, а именно определение связей, особенно типа «шире-уже» субъективен. На момент начала разработки мною тезауруса он состоял только из 500 связанных концептов, которые разрабатывались экспертами предметных областей несколько месяцев.

### 3.3.4. Создание онтологии в Protégé

После того, как неформальный тезаурус был готов мне было необходимо визуализировать свою работу, а именно оформить граф с изображением концептов и двух видов отношений между ними: «шире чем – уже чем» и «связан с». Прежде всего такое визуальное представление тезауруса мне необходимо для представления результатов работы на защите диплома.

Изначально мною рассматривался вариант построения тезауруса, а именно концептов и отношений между ними в стандартной программе CmapTools[[37]](#footnote-37).



1. Концептуальная карта из программы CmapTools

[Составлено автором самостоятельно на основе проведенного анализа]

Но, сделав в ней часть концептов и связей в ней, я столкнулась с тем, что это:

* Долго из-за ручного ввода связей
* Программа нестабильно работает
* Концептуальная карта выходит очень нагруженной даже тогда, когда я добавила только один вид отношений «шире чем»

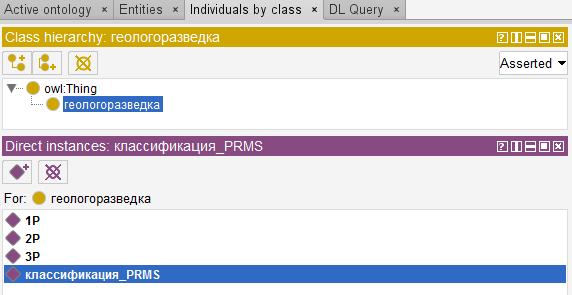
Поэтому мною было принято решений выбрать программу, онтосяющуюся напрямую к онтологическому инжинирингу.

Мне кажется излишним перегружать изображение тезауруса синонимами, определениями и переводом. Мне необходимо было выбрать программу, которая позволит мне изобразить концепты и отношения между нами, в то же время мне было необходимо формализовать тезаурус для самой системы корпоративного поиска. Загрузка в него прошлого тезауруса происходила вручную, что, опять же, крайне затратно по времени.

Мною была выбрана программа Protégé[[38]](#footnote-38). Protégé – это бесплатная, находащаяся в открытом доступе платформа, которая предоставляет ряд инструментов для построения моделей предметной области, основанных на онтологиях. Protégé является проектом «Semantic Web» Университета Стэнфорд Protégé является одной из самых известных программ для создания и редактирований онтологий различного вида. Программа имеет два формата: в виде web-приложения, позволяющем строить онтологии в формате совместного использования и в виде стандартного приложения на компьютер для различных операционных систем. Также данная программа позволяет пользователям импортировать и экспортировать онтологии в форматах RDF и OWL. Главным для себя достоинством Protégé я выделила удобный для использования графический интерфейс, содержащий справки и примеры. Мне, как начинающему пользователю в области построения онтологий было важно легко разобраться с такой программой. Поимого прочего, мне показалось интересным изучить именно программу за счёт её распространённости и широких возможностей, хотя я понимала, что они не пригодятся в случае загрузки в Protégé тезауруса.

Передо мной стояла задача изучить интерфейс и правила работы в программе Protégé, для этого мною были просмотрены обучающие видео на Youtube и различные статьи.

Первым делом мне нужно было загрузить концепты. Для этого я создала класс «Геологоразведка». Но, к сожалению, я не могу использовать классы для отображения концептов. Поэтому мне нужно было создать экземпляры, то есть «individuals», которые и будут отображать концепты и связи между ними. Для этого я зашла во вкладку «Individuals by classes», в окне иерархией классов выбрала класс «геологоразведка». Внутри этого класса и будут храниться все концепты. Затем внизу, в окне «direct instances» я добавила концепт. Для начала я решила добавить 4 концепта: 1Р,2Р,3Р и классификация PRMS, по причине того, что между собой они имеют оба вида отношений «шире чем – уже чем» и «связан с».



1. Загрузка концептов в Protégé

[Составлено автором по проведенному анализу с использованием корпоративных данных компании «Газпром нефть»]

После того, как концепты были загружены, следующем шагом стало добавление отношений вида «шире чем – уже чем» и «связан с». Для этого мне необходимо было открыть вкладку «object properties». Эта вкладка служит для добавления предикатов или свойств, то есть те свойства, у которых объектом выступает класс или экземпляр. Через эти свойства и были добавлены отношения «связан\_с» и «уже\_чем», которые нужны для отображения отношений между концептами. Для отображения связи «шире чем – уже чем» мною было создано именно свойство «уже\_чем» для того, чтобы обозначить его функциональным. Например, концепт «1P» уже, чем концепт «классификация PRMS». Отображая свойство «уже\_чем» функциональным, мы говорим о том, что концепт «1P» не может быть уже какого-либо ещё концепта. При этом концепт «классификация PRMS» может быть шире нескольких концептов. Помимо этого, свойство «уже\_чем» ассиметрично. А вот связь «связан\_с», наоборот, является симметричной, потому что если концепт «1P» связан с концептом «2P», то и концепт «2P» связан с концептом «1P». Важно также упомянуть и про добавление свойства «шире\_чем», которое обратно «уже\_чем». Это необходимо для того, чтобы при запуске «reasoner» мы могли видеть двухстороннюю связь концепций с отношением «уже чем, шире чем».

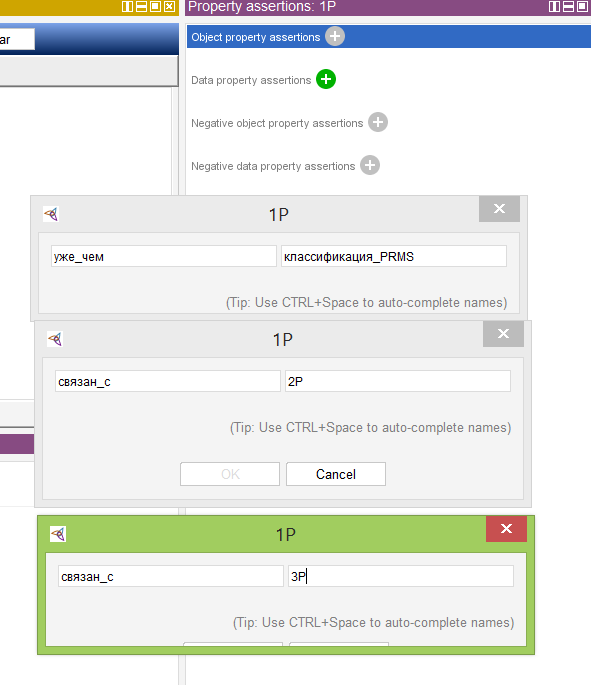


1. Добавленные отношения в Protégé

[Составлено автором по проведенному анализу с использованием корпоративных данных компании «Газпром нефть»]

Помимо этого, при добавлении свойства необходимо указывать «Domains» и «Ranges». Domains показывает из какого класса берутся субъекты для этого свойства. «Ranges» же является противоположностью «Domains» и определяет область допустимых значений, другими словами, откуда эти значения могут браться. Для обоих свойств что в «Domains», что в «Ranges» был выбран самый верхний класс «Things», хотя в этом случае не имеет значение какой из двух классов был бы выбран.

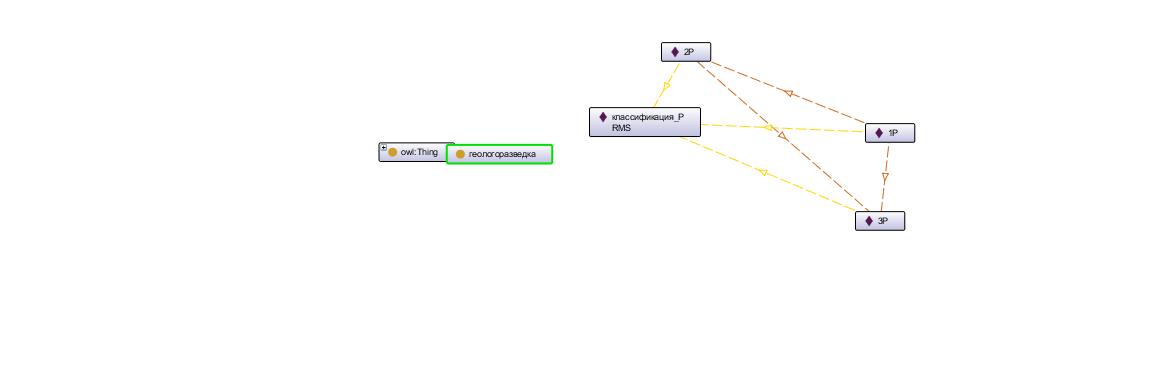
После того, как были определены свойства, их можно добавлять к концептам. Для этого нужно выбрать нужный концепт, зайти в окно «Property assertions» и нажать на «Object property assertions». Откроется диалоговое окно, где сначала нужно выбрать свойство, а затем экземпляр, который нужно связать этим свойством.



1. Присвоение свойств экземплярам в Protégé

[Составлено автором по проведенному анализу с использованием корпоративных данных компании «Газпром нефть»]

Я проделала эти же действия с оставшимися 3 концептами. В итоге 2 строчка Таблицы 4 может быть представлена следующим образом (взято только 4 концепта):



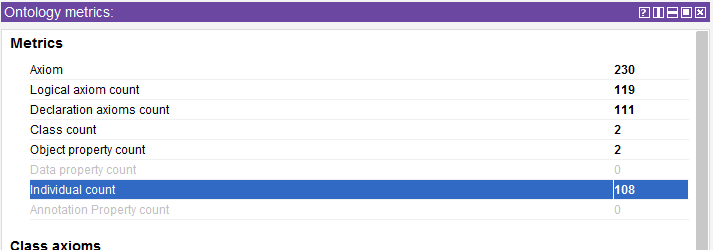
1. Пример графа концептов в Protégé

[Составлено автором по проведенному анализу с использованием корпоративных данных компании «Газпром нефть»]

Мы видим граф, на котором пунктирной линией желтого цвета изображены отношения «уже\_чем», а оранжевой пунктирной линией – «связан\_с».

Тебе мне было необходимо повторить вышеописанные действия с оставшимися экземплярами. Только на этот раз алгоритм был такой:

1. Я добавила все концепты как экземпляры (individuals) в Protégé

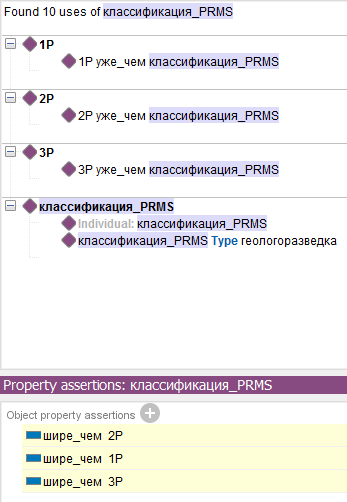


1. Онтологические метрики до добавления связей

[Составлено автором по проведенному анализу с использованием корпоративных данных компании «Газпром нефть»]

1. Затем заходила в каждый концепт и добавляла ему необходимые свойства: «уже\_чем», показывающее отношение «шире чем – уже чем» и свойство «связан с».

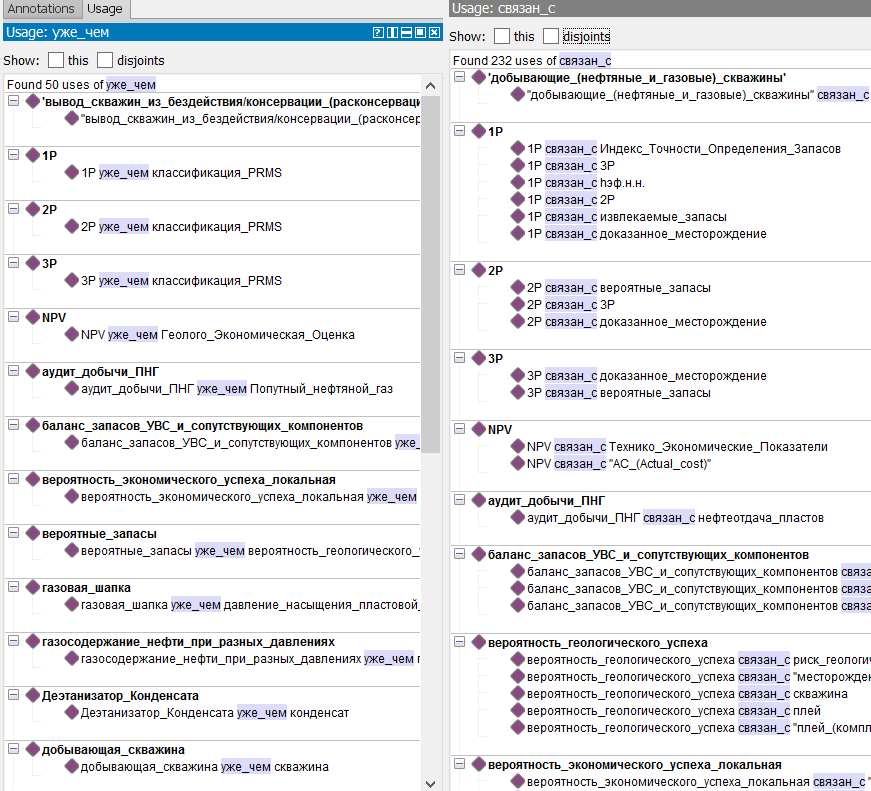
Благодаря тому, что в свойстве «связан\_с» я выбрала характеристику симметричности, мне не пришлось делать двойную работу, дублируя связи ли тратя время на проверку. Аналогично, за счёт того, что свойства «уже\_чем» и «шире\_чем» были обозначены как обратные, мне было достаточно выбирать одно из них.



1. Связь свойств «уже\_чем» и «шире\_чем» в Protégé

[Составлено автором по проведенному анализу с использованием корпоративных данных компании «Газпром нефть»]

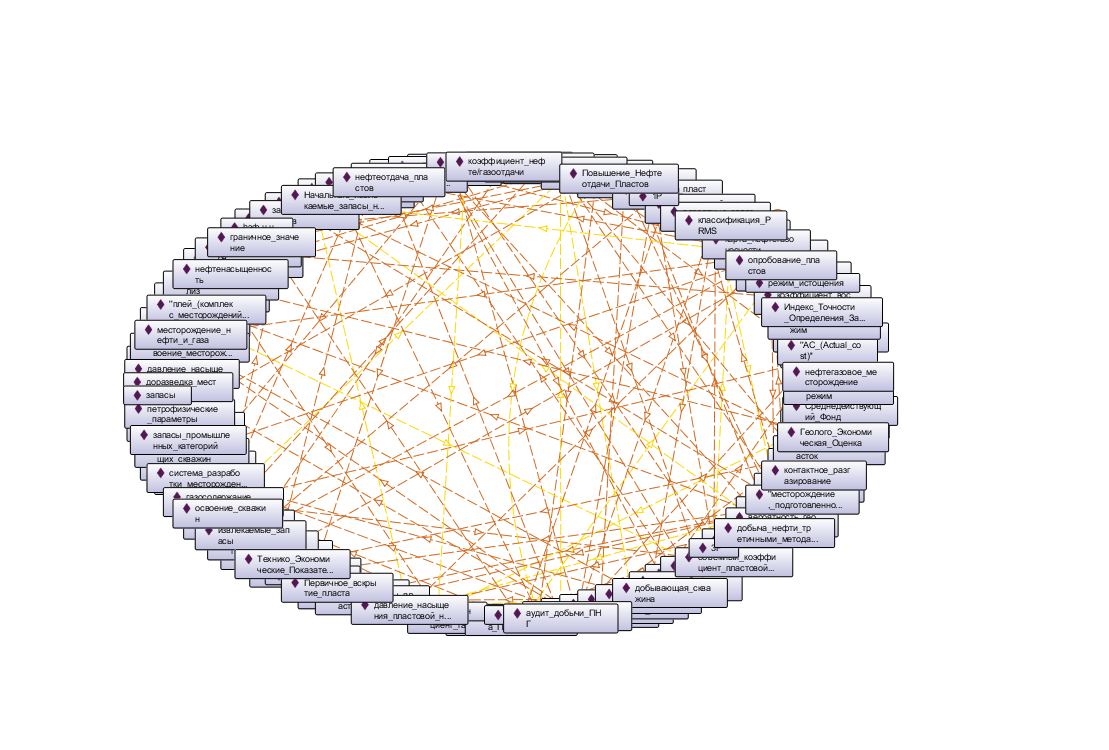
Таким образом, мною был составлен тезаурус предметной области «Геологоразведка». Тезаурус включает в себя 108 концептов, между которыми 232 раза было использовано свойство «связан\_с», как можно заметить на Рисунке ниже, эти связи уникальны и не повторяются между собой за счёт симметричности, выбранной в характеристике свойства. Помимо этого, была также использовано свойство «уже\_чем», которое было использовано мною при установлении отношения «шире чем – уже чем».



1. Отчёт по использованию «связан\_с» и «уже\_чем» в Protégé

[Составлено автором по проведенному анализу с использованием корпоративных данных компании «Газпром нефть»]

Несмотря на то, что изначально программа Protégé была выбрана мною по причине визуализации тезауруса в виде графа, нельзя не упомнить множеством плюсов работы в этой программе для создания тезауруса и удобство при его редактировании, хотя очевидно, что сама система Protégé ориентирована, скорее, на таксономии. Тем не менее, интерфейс системы показался мне очень удобным и нативным, а большое число открытых источников.



1. Граф построенный на основе тезауруса » в Protégé

[Составлено автором по проведенному анализу с использованием корпоративных данных компании «Газпром нефть»]

В данный момент редактирование и развитие тезауруса происходит через административную модель в самом корпоративном поиске, что крайне неудобно. Необходимо вручную добавлять каждый концепт, нет возможности перехода между страницами концепта, что крайне затрудняет общее понимание иерархии и отношений между концептами. В данный момент это значительно затрудняет процесс построение общего единого тезауруса для всей системы, возможно. Именно поэтому развитие проекта происходит так медленно. Внутренняя система редактирования хоть и понятно, но не удобна и не позволяет свободно перемещаться между концептами. Мне известно о том, что в системе предусмотрена загрузка триплетов, которые, однако, не удовлетворяют параметрам текущей онтологии с синонимичными связями и альтернативными назвния на английском языке.

Также стоит отметить непродуманность процесса разработки неформальной онтологической модели, построение которой происходит в таблицах Excel, что, безусловно, является доступным вариантом для экспертов, но совершенна не продумана в технологическом плане. В такой таблице сложно ориентироваться между понятиями, для того, чтобы понять, была ли уже составлена связь между концептами или нет приходится постоянно возвращаться к ним. Возможно именно причина неудобства и непродуманности составления неформальной модели играет ключевую роль в пассивном участие представителей предметной области в развитии тезауруса. Так, получив для обработки более 700 концептов из предметной области геологоразведки были верифицированы только 100 из них.

# Заключение

В первой главе данной выпускной квалификационной работы мною были рассмотрены торические материалы, статьи, учебники, различная литература как на тему базовых знаний в предметной области онтологии, так и более глубокие и специфические моменты, которые были необходимы для прохождения практики в компании «Газпром нефт» для написания данной дипломной работы, поскольку мне приходилось общаться с представителями компании с достаточным глубоким уровнем экспертизы в данной области. Важно отметить, что мою были просмотрены записи лекций преподавателей Высшей Школы Менеджмента специально для сотрудников компании «Газпром Нефть», материалы которых не были представлены в рамках ни одного из курсов. Таким образом, мною были описаны как общие понятия, так и различные подходы к классификации онтологий и методов онтологического инжиниринга. В целом считаю анализ литературы и изучение источников, поиск информации на эту тему самым важным этапом этой работы, поскольку от глубины уровня понимания темы зависла дальнейшая работа. Я приобрела много новых знаний в онтологическом инжиниринге, а качественное изучение литературы помогло сформировать комплексное понимание изученной тематике.

Во второй главе мною была рассмотрена компания в рамках деятельности которой и строилась моя выпускная квалификационная работа. Мною были описаны как общая информация о компании «Газпром Нефть», так и тенденция на цифровизацию бизнеса, которая, безусловна, находит отражение в деятельности всей компании, что позволяет лучше понять стратегию компанию. Помимо этого, мною была описана и проанализирована и описана деятельность двух дочерних организаций компании «Газпром Нефть»: «Газпромнефть-Цифровые решения» и «Газпром нефть НТЦ», что также позволило сформировать общее представление о всей корпорации.

Следующая, третья глава является основной и заключительной. Сначала мною был изучен корпоративный поиск, который является выдающимся проектом компании «Газпром нефть». На основании собственного опыта работы в нём и изучении источников компании, я коротко, но емко писала основные моменты устройства и принципов работы когнитивного поиска в компании. Затем мне было необходимо понять роль и значение тезауруса в корпоративном поиске, для этого мною было проведено интервью с представителями отдела корпоративного поиска, а также я непосредственно работала в панели администрирования тезауруса корпоративного поиска.

Таким образом, я описала принципы устройства тезауруса, роль концепта в системе корпоративного поиска, я описала процесс создания, который непосредственно повлиял на текущий его вид. Онтология корпоративного поиска была мною проанализирована и классифицирована.

Во второй половине 3 главы я приступила к непосредственной работе над созданием тезауруса. Мною была выбрана предметная область «геологоразведки», нуждающаяся в доработке. На основании нескольких методик построения онтологии я составила подходящий мне способ онтологического инжиниринга, который должен удовлетворять два критериям:

* Применимость результатов на практике
* «Защищабельность» результатов перед комиссией

Важным шагом была разработка неформально предметной области, мне потребовалось много времени на внимательное изучение и выбор ресурсов После того как неформальная онтология была сформулирована, необходимо было добавит её в программу Protégé. Для этого мною были изучение различные видео-материалы, статьи. Особое внимание я уделила характеристикам свойств, которые представляли собой связи. Я выбрала их формальность, симметричности или асимметричностью, в зависимости от типа связи, а также настроила обратную зависимость для более полной картины тезауруса. В результате, тезаурус принял полуформальный вид, который можно как презентовать на защите, так и внедрять в корпоративную систему поиска. Помимого этого мною были формулированы рекомендации по доработки процесса создания неформальной модели тезауруса.

Таким образом, результатом моей выпускной квалификационной работы является тезаурус предметной области «Геологоразведка», который включает в себя 108 концептов, между которыми 232 раза было использовано свойство «связан\_с», и 50 раз связь «уже\_чем». Этот тезаурус представлен в неформальном виде для компании «Газпром нефть» и добавлен в программу Protégé для формализации модели.

# Список использованной литературы

1. Гаврилова Т. А. Онтологический подход к управлению знаниями при разработке корпоративных информационных систем //Новости искусственного интеллекта. – 2003. – №. 2. – С. 24-30. (<http://bigc.ru/publications/bigspb/km/ontol_podhod_to_uz.php#1> )
2. Гаврилова Т.А. Модели и методы структурирования знаний: визуальный подход [Электронный ресурс]/ Сайт spbu.ru, URL: [https://spbu.ru/sites/default/files/20160328\_protocol\_us\_sp](https://spbu.ru/sites/default/files/20160328_protocol_us_spbu_material_presentation_gavrilova.pd), свободный. – Загл. с экрана (дата обращения - 13.03.2021)
3. Газпром нефть создает собственный ИТ-кластер для ускорения цифровой трансформации [Электронный ресурс]/ Сайт Neftegaz.RU, <URL:https://neftegaz.ru/news/tsifrovizatsiya/549554-gazprom-neft-sozdaet-sobstvennyy-it-klaster-dlya-uskoreniya-tsifrovoy-transformatsii/> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
4. Газпром нефть» создает собственный ИТ-кластер для ускорения цифровой трансформации [Электронный ресурс]/ Сайт компании «Газпром нефть», <URL:https://www.gazprom-neft.ru/press-center/news/gazprom_neft_sozdaet_sobstvennyy_it_klaster_dlya_uskoreniya_tsifrovoy_transformatsii/> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
5. Главный НПЗ России [Электронный ресурс]/ Сайт Омского нефтеперерабатывающего завода «Газпром нефть», <URL:https://onpz.gazprom-neft.ru/> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
6. Годовой отчет 2018 год [Электронный ресурс]//Газпром нефть. Режим доступа <https://ir.gazprom-neft.ru/fileadmin/user_upload/documents/annual_reports/gpn_ar18_rus_1909.pdf> (дата обращения 12.02.2020)
7. Горелова А. А., Попова О. А., Филиппова З. М. Понятие системы онтологий и ее применение для характеристики условий терригенного осадконакопления //Территория Нефтегаз. – 2019. – №. 12.
8. Добров Б. В. и др. Онтологии и тезаурусы: учебно-методическое пособие //Казань: Изд-во Казанского ГУ. – 2006.
9. Завод для города [Электронный ресурс]/ Сайт Сайт Московского нефтеперерабатывающего завода «Газпром нефть», <URL:https://mnpz.gazprom-neft.ru/> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
10. Корпоративный поиск – единая точка доступа к знаниям компании [Электронный ресурс]/ Сайт GlobalCIO, <URL:https://globalcio.ru/live/projects/10524/> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
11. Кудрявцев Д. В. Системы управления знаниями и применение онтологий: учебное пособие. – 2010.
12. Лаврищева Е. М., Карпов Л. Е., Томилин А. Н. Семантические ресурсы для разработки онтологии научной и инженерной предметных областей //Труды. XVIII Всероссийский научной конференции «Научный сервис в сети Интернет. – 2016. – С. 223-239.
13. Митрофанова О. А., Константинова Н. С. Онтологии как системы хранения знаний. – 2008.
14. Награждение «Проект Года 2018» [Электронный ресурс]/ Сайт GlobalCIO, <URL:https://globalcio.ru/discussion/1863/> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
15. Нариньяни А.С. ТЕОН-2: ОТ ТЕЗАУРУСА К ОНТОЛОГИИ И ОБРАТНО [Электронный ресурс]/ Сайт, HYPERLINK URL:http://www.dialog-21.ru/digest/2002/articles/narinyani/ , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
16. НАУЧНАЯ ДЕЯТЕЛЬНОСТЬ [Электронный ресурс]/ Сайт научно-технического центра компании «Газпром нефть», <URL:https://ntc.gazprom-neft.ru/research-and-development/science/> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
17. НАУЧНО-ТЕХНИЧЕСКИЙ ЦЕНТР «ГАЗПРОМ НЕФТИ» [Электронный ресурс]/ Сайт научно-технического центра компании «Газпром нефть», [URL:https://ntc.gazprom-neft.ru/about/hello-ntc /](URL:https://ntc.gazprom-neft.ru/about/hello-ntc%20/) , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
18. НАУЧНО-ТЕХНИЧЕСКИЙ ЦЕНТР «ГАЗПРОМ НЕФТИ» [Электронный ресурс]/Сайт, URL: https://www.naumen.ru/digital-transformation/gazpromneft.php , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
19. НПЗ Панчево (NIS, Сербия) [Электронный ресурс]/ Сайт компании «Газпром нефть», <URL:https://www.gazprom-neft.ru/company/oil-refining/nis-refinery/> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
20. Объявлены победители конкурса "Проект Года 2018" [Электронный ресурс]/ Сайт COMNEWS, <URL:https://www.comnews.ru/content/117426/2019-01-31/obyavleny-pobediteli-konkursa-proekt-goda-2018> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
21. Сладкова Е., Зорина С. Умный поиск [Электронный ресурс]/ Сайт компании «Газпром нефть», URL: <https://www.gazprom-neft.ru/press-center/sibneft-online/archive/2020-november/5117433/> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
22. Филиппович А. Тезаурус vs. Онтология [Электронный ресурс]/ Персональный блог Филипповича Андрея,URL: [http://blogs.it-claim.ru/andrey/2011/01/18/](http://blogs.it-claim.ru/andrey/2011/01/18/%D1%82%D0%B5%D0%B7%D0%B0%D1%83%D1%80%D1%83%D1%81-vs-%D0%BE%D0%BD%D1%82%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%8F/), свободный. – Загл. с экрана (дата обращения - 10.02.2021)
23. Электронный ресурс] / Сайт posp.raai.org, URL: <http://posp.raai.org/data/posp2005/gavrilova/gavrilova.html> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)]
24. Cmap software is a result of research conducted at the Florida Institute for Human & Machine Cognition (IHMC). It empowers users to construct, navigate, share and criticize knowledge models represented as concept maps [Электронный ресурс]/ Сайт Cmap, URL:https://cmap.ihmc.us/, свободный. – Загл. с экрана (дата обращения - 13.03.2021)
25. Controlled vocabulary thesaurus onthology [Электронный ресурс]/ Сайт, URL: <http://www.dianebowen.net/controlled-vocabulary-thesaurus-ontology.html> , свободный. – Загл. с экрана (дата обращения - 10.02.2021)
26. Gómez-Pérez A., Fernández-López M., Corcho O. Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. – Springer Science & Business Media, 2006.
27. Gruber T. R. A translation approach to portable ontology specifications //Knowledge acquisition. – 1993. –
28. Gruber T. R. A translation approach to portable ontology specifications //Knowledge acquisition. – 1993. – Т. 5. – №. 2. – С. 199-220.
29. Gruber T. R. A translation approach to portable ontology specifications //Knowledge acquisition. – 1993. – Т. 5. – №. 2. – С. 199-220.
30. Lassila O, McGuiness D, The role of frame-based representation on the semantic web //Linköping Electronic Articles in Computer and Information Science. – 2001
31. Noy N. F. et al. Ontology development 101: A guide to creating your first ontology. – 2001.
32. Noy N. F. et al. Ontology development 101: A guide to creating your first ontology. – 2001.
33. Protégé is a free, open-source platform that provides a growing user community with a suite of tools to construct domain models and knowledge-based applications with ontologies [Электронный ресурс]/ Сайт Protégé, <URL:https://protege.stanford.edu/products.php> , свободный. – Загл. с экрана (дата обращения - 13.03.2021)
34. Unified Medical Language System (UMLS) [Электронный ресурс]/ Сайт, URL:https://www.nlm.nih.gov/research/umls/index.html, свободный. – Загл. с экрана (дата обращения - 10.02.2021)
35. Uschold M. et al. Ontologies: Principles, methods and applications //TECHNICAL REPORT-UNIVERSITY OF EDINBURGH ARTIFICIAL INTELLIGENCE APPLICATIONS INSTITUTE AIAI TR. – 1996.

1. Сладкова Е., Зорина С. Умный поиск [Электронный ресурс]/ Сайт компании «Газпром нефть», URL: <https://www.gazprom-neft.ru/press-center/sibneft-online/archive/2020-november/5117433/> , свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-1)
2. Gruber T. R. A translation approach to portable ontology specifications //Knowledge acquisition. – 1993. – Т. 5. – №. 2. – С. 199-220. [↑](#footnote-ref-2)
3. Гаврилова Т. А. Онтологический подход к управлению знаниями при разработке корпоративных информационных систем //Новости искусственного интеллекта. – 2003. – №. 2. – С. 24-30. (<http://bigc.ru/publications/bigspb/km/ontol_podhod_to_uz.php#1> ) [↑](#footnote-ref-3)
4. Gruber T. R. A translation approach to portable ontology specifications //Knowledge acquisition. – 1993. – [↑](#footnote-ref-4)
5. Горелова А. А., Попова О. А., Филиппова З. М. Понятие системы онтологий и ее применение для характеристики условий терригенного осадконакопления //Территория Нефтегаз. – 2019. – №. 12. [↑](#footnote-ref-5)
6. Митрофанова О. А., Константинова Н. С. Онтологии как системы хранения знаний. – 2008. [↑](#footnote-ref-6)
7. Gómez-Pérez A., Fernández-López M., Corcho O. Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. – Springer Science & Business Media, 2006. [↑](#footnote-ref-7)
8. Noy N. F. et al. Ontology development 101: A guide to creating your first ontology. – 2001. [↑](#footnote-ref-8)
9. Gruber T. R. A translation approach to portable ontology specifications //Knowledge acquisition. – 1993. – Т. 5. – №. 2. – С. 199-220. [↑](#footnote-ref-9)
10. Гаврилова Т. А. и др. Инженерия знаний. Модели и методы. – 2016. [↑](#footnote-ref-10)
11. Лаврищева Е. М., Карпов Л. Е., Томилин А. Н. Семантические ресурсы для разработки онтологии научной и инженерной предметных областей //Труды. XVIII Всероссийский научной конференции «Научный сервис в сети Интернет. – 2016. – С. 223-239. [↑](#footnote-ref-11)
12. Unified Medical Language System (UMLS) [Электронный ресурс]/ Сайт, URL:https://www.nlm.nih.gov/research/umls/index.html, свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-12)
13. Кудрявцев Д. В. Системы управления знаниями и применение онтологий: учебное пособие. – 2010. [↑](#footnote-ref-13)
14. Lassila O, McGuiness D, The role of frame-based representation on the semantic web //Linköping Electronic Articles in Computer and Information Science. – 2001 [↑](#footnote-ref-14)
15. Филиппович А. Тезаурус vs. Онтология [Электронный ресурс]/ Персональный блог Филипповича Андрея,URL: [http://blogs.it-claim.ru/andrey/2011/01/18/](http://blogs.it-claim.ru/andrey/2011/01/18/%D1%82%D0%B5%D0%B7%D0%B0%D1%83%D1%80%D1%83%D1%81-vs-%D0%BE%D0%BD%D1%82%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%8F/), свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-15)
16. 6. Нариньяни А.С. ТЕОН-2: ОТ ТЕЗАУРУСА К ОНТОЛОГИИ И ОБРАТНО [Электронный ресурс]/ Сайт, HYPERLINK URL:http://www.dialog-21.ru/digest/2002/articles/narinyani/ , свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-16)
17. Добров Б. В. и др. Онтологии и тезаурусы: учебно-методическое пособие //Казань: Изд-во Казанского ГУ. – 2006. [↑](#footnote-ref-17)
18. 7. Controlled vocabulary thesaurus onthology [Электронный ресурс]/ Сайт, URL: <http://www.dianebowen.net/controlled-vocabulary-thesaurus-ontology.html> , свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-18)
19. Noy N. F. et al. Ontology development 101: A guide to creating your first ontology. – 2001. [↑](#footnote-ref-19)
20. Uschold M. et al. Ontologies: Principles, methods and applications //TECHNICAL REPORT-UNIVERSITY OF EDINBURGH ARTIFICIAL INTELLIGENCE APPLICATIONS INSTITUTE AIAI TR. – 1996. [↑](#footnote-ref-20)
21. Годовой отчет 2018 год [Электронный ресурс]//Газпром нефть. Режим доступа <https://ir.gazprom-neft.ru/fileadmin/user_upload/documents/annual_reports/gpn_ar18_rus_1909.pdf> (дата обращения 12.02.2020) [↑](#footnote-ref-21)
22. Главный НПЗ России [Электронный ресурс]/ Сайт Омского нефтеперерабатывающего завода «Газпром нефть», URL:https://onpz.gazprom-neft.ru/ , свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-22)
23. Завод для города [Электронный ресурс]/ Сайт Сайт Московского нефтеперерабатывающего завода «Газпром нефть», URL:https://mnpz.gazprom-neft.ru/ , свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-23)
24. НПЗ Панчево (NIS, Сербия) [Электронный ресурс]/ Сайт компании «Газпром нефть», URL:https://www.gazprom-neft.ru/company/oil-refining/nis-refinery/, свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-24)
25. «Газпром нефть» создает собственный ИТ-кластер для ускорения цифровой трансформации [Электронный ресурс]/ Сайт компании «Газпром нефть», URL:https://www.gazprom-neft.ru/press-center/news/gazprom\_neft\_sozdaet\_sobstvennyy\_it\_klaster\_dlya\_uskoreniya\_tsifrovoy\_transformatsii/, свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-25)
26. Газпром нефть создает собственный ИТ-кластер для ускорения цифровой трансформации [Электронный ресурс]/ Сайт Neftegaz.RU, URL:https://neftegaz.ru/news/tsifrovizatsiya/549554-gazprom-neft-sozdaet-sobstvennyy-it-klaster-dlya-uskoreniya-tsifrovoy-transformatsii/ , свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-26)
27. НАУЧНАЯ ДЕЯТЕЛЬНОСТЬ [Электронный ресурс]/ Сайт научно-технического центра компании «Газпром нефть», URL:https://ntc.gazprom-neft.ru/research-and-development/science/, свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-27)
28. НАУЧНО-ТЕХНИЧЕСКИЙ ЦЕНТР «ГАЗПРОМ НЕФТИ» [Электронный ресурс]/ Сайт научно-технического центра компании «Газпром нефть», URL:https://ntc.gazprom-neft.ru/about/hello-ntc/, свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-28)
29. Корпоративный поиск – единая точка доступа к знаниям компании [Электронный ресурс]/ Сайт GlobalCIO, URL:https://globalcio.ru/live/projects/10524/, свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-29)
30. Внутренняя информация компании «Наконец-то новости Корпоративного поиска!» от 6 ноября 2020 [↑](#footnote-ref-30)
31. Объявлены победители конкурса "Проект Года 2018" [Электронный ресурс]/ Сайт COMNEWS, URL:https://www.comnews.ru/content/117426/2019-01-31/obyavleny-pobediteli-konkursa-proekt-goda-2018, свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-31)
32. Награждение «Проект Года 2018» [Электронный ресурс]/ Сайт GlobalCIO, URL:https://globalcio.ru/discussion/1863/, свободный. – Загл. с экрана (дата обращения - 10.02.2021) [↑](#footnote-ref-32)
33. Внутренняя информация компании «Новые возможности Корпоративного поиска» [↑](#footnote-ref-33)
34. Американский институт нефти – одна из авторитетных коммерческих организаций, которым принадлежит ведущая роль в классификации моторных масел и установлении требований к ним. [↑](#footnote-ref-34)
35. Гаврилова Т.А. Модели и методы структурирования знаний: визуальный подход [Электронный ресурс]/ Сайт spbu.ru, URL: [https://spbu.ru/sites/default/files/20160328\_protocol\_us\_sp](https://spbu.ru/sites/default/files/20160328_protocol_us_spbu_material_presentation_gavrilova.pd), свободный. – Загл. с экрана (дата обращения - 13.03.2021) [↑](#footnote-ref-35)
36. Внутренняя документация компании «Газпром нефть» из системы распространения знаний в рабочей зоне «Онтологии» [↑](#footnote-ref-36)
37. Cmap software is a result of research conducted at the Florida Institute for Human & Machine Cognition (IHMC). It empowers users to construct, navigate, share and criticize knowledge models represented as concept maps [Электронный ресурс]/ Сайт Cmap, URL:https://cmap.ihmc.us/, свободный. – Загл. с экрана (дата обращения - 13.03.2021) [↑](#footnote-ref-37)
38. Protégé is a free, open-source platform that provides a growing user community with a suite of tools to construct domain models and knowledge-based applications with ontologies [Электронный ресурс]/ Сайт Protégé, URL:https://protege.stanford.edu/products.php, свободный. – Загл. с экрана (дата обращения - 13.03.2021) [↑](#footnote-ref-38)