

Министерство образования Республики Беларусь  
Учреждение образования  
«Белорусский государственный университет  
информатики и радиоэлектроники»

УДК 004.896

Грузинова  
Любовь Игоревна

Средства анализа качества и структуризации баз знаний

**АВТОРЕФЕРАТ**

на соискание степени магистра технических наук  
по специальности 1-40 80 06 «Искусственный интеллект»

---

Научный руководитель

Гулякина Наталья Анатольевна

Кандидат физико-математических  
наук, доцент, зам. зав. каф. по  
научной работе

---

Минск 2021

## ВВЕДЕНИЕ

Экспоненциальный рост объемов информации во многих областях человеческой деятельности делает задачу эффективного информационного обеспечения все более актуальной. Проблема такого характера зачастую рассматривается в контексте создания некоего единого источника информации, в частности, баз знаний, их модификации и последующего расширения, а также систематизации с целью облегчения их использования и обработки. Разработка систем, основанных на знаниях, является одним из наиболее перспективных направлений в данной области. Однако это является довольно сложным и трудоемким процессом.

При разработке базы знаний важно не только обеспечить возможность хранения знаний, реализовать возможность поиска и навигации по ним, но и предоставить возможность работать над созданием и модификацией базы знаний целой группе разработчиков. Коллективная работа позволяет многократно ускорить разработку базы знаний, однако приводит к новым проблемам. Каждый разработчик имеет субъективное представление и понимание разрабатываемой в базе знаний предметной области, свою схему представления необходимых к формализации знаний, что может привести к понижению связанности, усложнению последующей поддержки, усложнению чтения и навигации как для разработчиков, так и для пользователей системы.

Для проектирования и последующей разработки базы знаний не существует единой унифицированной системы ее структурирования и наполнения. Основной проблемой является выделение основных характеристик качества базы знаний и последующий их анализ. Поэтому необходима некоторая система качеств и структуризации, общих для всех существующих и разрабатываемых баз знаний список характеристик, по которым можно будет оценить уровень качества разрабатываемой системы, а также средства для их анализа.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Целью диссертационной работы является исследование существующих и разработка новых средств, которые позволят анализировать качество и структуризацию баз знаний.

Объект исследования: критерии качества и структуризации баз знаний.

Предмет исследования: методы и средства анализа качества и структуризации баз знаний.

Задачи исследования включают:

1. Обзор существующих средств анализа качества и структуризации баз знаний;
2. Анализ существующих подходов к анализу качества и структуризации баз знаний, а также выявление основных характеристик качества базы знаний, по которым может быть произведён анализ;
3. Исследование выбранных характеристик качества и разработка алгоритмов, на их основе.

Научная новизна обусловлена отсутствием при проектировании и последующей разработке баз знаний единой унифицированной системы ее структурирования и наполнения

## СОДЕРЖАНИЕ РАБОТЫ

В первой главе магистерской диссертации был произведен анализ существующих подходов к структуризации и анализу качества баз знаний были рассмотрены следующие направления:

Рассмотрена технология OSTIS (Open Semantic Technology for Intelligent Systems), а также основанные на этой технологии ostis-системы.

Была рассмотрена общая структура интеллектуальной системы и структура знаний в базах знаний. Рассмотрены существующие подходы к структурированию знаний. На стадии структурирования знаний предметной области необходимо решить следующие задачи: составление словаря используемых терминов; выявление понятий и их атрибутов; выявление связей и определение отношений между понятиями; детализация и обобщение понятий; построение обобщенной структуры знаний предметной области. Рассмотрено понятие качества знаний и определены критерии качества знаний системы, а также основные показатели качества, присущие системе. Из вышесказанного вытекает необходимость постоянного контроля качества данных как со стороны получателя, так и со стороны поставщика. Что в свою очередь вынуждает разрабатывать и применять специальные контрольно-измерительные инструменты.

Во второй главе были рассмотрены существующие критерии оценки качества базы знаний. Выбраны основные критерии, по которым в дальнейшем будут разработаны средства анализа качества базы знаний системы. Было представлено содержимое базы знаний проектируемой системы в виде онтологии критериев качества реализованное на языке SCn. С помощью проектируемой системы на примере фрагмента ИИС по химии был проведён качественный анализ фрагмента базы знаний, были показаны примеры сильного и слабого показателя каждого из набора критериев (полнота, связность и информативность). Также было выявлено, что уровень структурированности рассматриваемой системы напрямую зависит от показателя отдельно взятых критериев качества, например, связности, что позволяет производить анализ качества и структурированности параллельно и совместно друг с другом. Был спроектирован общий алгоритм действий при структуризации знаний со стороны разработчика баз знаний, который позволяет выстроить иерархию знаний передаваемой в качестве входного параметра области знаний.

В третьей главе была спроектирована система SCP-агентов для оценки качества базы знаний по каждому из набора характеристик. Были представлены условия инициирования и алгоритм работы каждого агента. Агенты позволяют отобрать в выбранной системе узлы, которые не соответствуют таким критериям как полнота, связность и информативность для последующей модификации системы.

Также был реализован SCP-агент поиска всех узлов без идентификатора узла пояснения. Этот агент позволяет найти в базе знаний понятия, у которых отсутствует идентификатор узла определения/пояснения. Наличие идентификаторов у данных узлов улучшают семантическую связанность между узлами базы знаний, повышают понятность представленной информации для конечного пользователя, а также позволяют повысить частотность выдачи понятия в поиске.

## ЗАКЛЮЧЕНИЕ

Задача эффективного информационного обеспечения с каждым днем становится все более актуальной. Поэтому для решения этой проблемы создаются некие единые источники информации, в частности базы знаний. Разработка систем, основанных на знаниях, является одним из наиболее перспективных направлений в данной области. Однако это является довольно сложным и трудоемким процессом.

Так как для проектирования и последующей разработки базы знаний не существует единой унифицированной системы ее структурирования и наполнения, необходима некоторая система качеств и структуризации, общих для всех существующих и разрабатываемых баз знаний, по которым можно будет оценить уровень качества разрабатываемой системы, а также средства для их анализа.

Главной задачей магистерской диссертации является разбор и обработка сведений о существующих средствах анализа качества и структуризации баз знаний, а также проектирование алгоритмов и последующая реализация средств анализа качеств баз знаний на базе *ostis*-системы, позволяющих оценить представленную для примера и тестирования область базы знаний по полученному набору характеристик качества и структуризации, таких как полнота, связность и информативность.

В процессе работы был произведен разбор существующих подходов к анализу качества и структуризации баз знаний. Определена базовые стратегии структурирования базы знаний и инструменты ее структуризации. Так же были выделены основные характеристики качества базы знаний системы.

На каждый из критериев (полнота, связность, информативность) были сформулированы логика и условия инициализации, а также спроектирован алгоритм работы SCP-агента для оценки и анализа, а также реализован один из SCP-агентов.

Полученный в результате набор критериев не является конечным и подразумевает дальнейшее расширение (выделение и оценка дополнительных критериев), как и возможность оценки каждого из уже выделенных критериев не одним, а комплексной системой агентов (оценка критерия с разных сторон), что говорит о возможности дальнейшего расширения и модификации системы в плане ее анализа.