

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

УДК 004.82:004.89

Джум
Владислав Егорович

Модели анализа баз знаний

АВТОРЕФЕРАТ

на соискание степени магистра технических наук

по специальности 1-31 80 10 – Теоретические основы информатики

Научный руководитель
Гулякина Наталья Анатольевна
кандидат физико-математических наук,
доцент

Минск 2019

КРАТКОЕ ВВЕДЕНИЕ

База знаний – ядро любой интеллектуальной системы, совокупность знаний предметной области. Как правило, базы знаний имеют большой объём, хранят знания различных предметных областей. При накоплении больших объёмов знаний появляется огромное количество самых различных понятий и возникает проблема структуризации и систематизации этого многообразия понятий.

Современные базы знаний — один из тех объектов в сфере информатизации, от которых требуется особенно высокое качество и наличие возможности его оценки. Для понимания, значения качества баз знаний, требований предъявляемых к их качеству, какими характеристиками можно описывать качество, могут быть полезны методы и стандарты, разработанные для анализа сложных программных средств.

В силу того, что разработка интеллектуальных систем является распределенным и параллельным процессом, человеку трудно оценить корректность функционирования каждой из взаимодействующих компонент системы.

Как и любой программный продукт, базу знаний необходимо проверять на предмет качества. Анализ качества базы знаний имеет огромное значение в процессе разработки и эксплуатации интеллектуальных систем, так как ошибки и неточности в базе знаний могут привести к непредсказуемому и некорректному поведению системы, а также нарушить целостность базы знаний.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является разработка методов анализа баз знаний интеллектуальных систем, основанных на семантических сетях.

Для достижения поставленной цели необходимо решить следующие задачи:

- определить характеристики качества баз знаний интеллектуальных систем;
- проанализировать основные существующие методы анализа программного обеспечения;
- выделить основные типы проблемных фрагментов базы знаний;
- описать методы анализа баз знаний;
- разработать методы и алгоритмы анализа базы знаний.

Объектом исследования являются базы знаний интеллектуальных систем.

Предметом исследования являются процессы проектирования и пополнения баз знаний, а принципы разработки программных средств для решения задач анализа баз знаний, методы и алгоритмы анализа онтологий, целостности, непротиворечивости, полноты баз знаний.

Связь работы с приоритетными направлениями научных исследований

Работа выполнялась в соответствии научно-техническими заданиями и планами работ кафедры «Интеллектуальных информационных технологий».

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя Н. А. Гулякиной, заключается в формулировке целей и задач исследования, определении возможных путей решения, оценке результатов.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников. Общий объем магистерской диссертации составляет 63 страницы, включая 30 иллюстраций, библиографический список из 21 наименований.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В **первой главе** описан процесс проектирования баз знаний, определены характеристики качества баз знаний, выведены основные характеристики знаний для формирования критериев, по которым требуется проводить анализ баз знаний.

Основные характеристики знаний:

1 Внутренняя интерпретируемость. Все информационные объекты должны иметь уникальный идентификатор. В случае отсутствия имен, данные невозможно идентифицировать системой. В такой ситуации данные можно извлечь только при помощи разработчика.

2 Структурированность. Структура информационных объектов должна быть гибкой. Каждый объект может находиться в составе любой другой структуры и из каждого объекта можно выделить некоторые его составляющие. То есть информационный объект должен обладать способностью как и самостоятельного существования, так и в составе других структур (часть - целое).

3 Связность. Между информационными объектами должна быть возможность устанавливать различные связи, характеризующих отношения между ними. Такими отношениями могут быть: отношения структуризации (иерархия), функциональные отношения (вычисление одних информационных объектов через другие), казуальные отношения (причинно-следственные связи), семантические отношения.

4 Семантическая метрика. Установление отношения релевантности между информационными объектами. Отношение релевантности предоставляют возможность нахождения знаний, близких к уже найденным.

В рамках **второй главы** произведён анализ структуры баз знаний, построенных на основе технологии OSTIS, определена предметная область проблемных фрагментов базы знаний, выделены наиболее часто встречающиеся ошибки в процессе разработки базы знаний, выявлены причины, по которым в базе знаний появляется информационный мусор, определены методы оценки онтологий.

Наиболее часто встречающиеся ошибки в процессе разработки базы знаний:

– неполнота - недостаточный объём информации, входящей в какую-либо предметную область, означает неудовлетворение запросов пользователей в пределах области использования базы знаний;

– противоречивость - возможность вывода в одной системе двух взаимоисключающих утверждений;

- нецелостность - несогласованность по данным либо правилам вывода;
- зацикленность - ситуация, при которой действия над знаниями могут каким-то образом привести к утверждению тех же самых знаний;
- ошибки вывода - совпадение результатов вывода из существующих знаний, в случае, когда эти знания имеют различные несхожие состояния;
- избыточность - наличие нескольких цепочек рассуждений, позволяющих сделать один и тот же вывод.

В третьей главе описан принцип работы sc-агентов, произведена классификация абстрактных sc-агентов анализа базы знаний, а также произведена их декомпозиция.

Агенты анализа базы знаний, можно классифицировать по следующим типам:

- проверка на наличие циклов;
- проверка на непротиворечивость;
- проверка на целостность;
- проверка на наличие информационного мусора;
- проверка на полноту;
- оценка качества онтологий.

ЗАКЛЮЧЕНИЕ

По итогам работы решены следующие задачи:

1 Изучены этапы создания базы знаний, определены характеристики качества баз знаний.

2 На основе этого выделены типы ошибочных ситуаций, которые допускаются при проектировании и создании баз знаний. Основными из них являются противоречивость, неполнота, нецелостность, зацикленность, избыточность. Описана предметная область проблемных фрагментов базы знаний.

3 Изучены основные методы обнаружения ошибок.

4 Разработана модель анализа базы знаний интеллектуальных систем, в основе которой лежит анализ структуры базы знаний с помощью методов оценки онтологий и верификации. Суть методов состоит в том, что анализ проводится машиной обработки знаний проверяемой системы, путём запуска проверочных алгоритмов. Результат работы программы может быть передан экспертам для дальнейшего анализа или оценки. Таким образом, отсутствует проблема интеграции разработанных программных средств с базой знаний, так как они являются её частью.

5 Определены правила, которым должна соответствовать структура база знаний. Основываясь на эти правила, разработаны алгоритмы работы агентов.

6 Разработаны агенты анализа базы знаний. Составлена спецификация агентов с описанием условий инициирования и результата работы. Описаны алгоритмы работы агентов, с пошаговым описанием хода выполнения.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

[1] Джум, В. Е. Коллектив агентов оценки полноты баз знаний / В. Е. Джум, П. Н. Лось, А. Г. Шалёв // Информационные технологии и управление: 53-я научная конференция аспирантов, магистрантов и студентов учреждения образования «Белорусский государственный университет информатики и радиоэлектроники». – Минск: БГУИР, 2017. – С. 20.

[2] Джум, В. Е. Агенты верификации баз знаний / В. Е. Джум, В. В. Трунц, А. Г. Шалёв // Информационные технологии и управление: материалы 52-й научной конференции аспирантов, магистрантов и студентов.. – Минск: БГУИР, 2016. – С. 28 – 29.