

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

УДК 004.896

Борщ
Егор Владимирович

Средства семантического анализа вводимых фрагментов проектируемой
базы знаний

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

Научный руководитель
Шункевич Даниил Вячеславович
канд. технических наук, доцент

Минск 2021

ВВЕДЕНИЕ

Сокращение трудоемкости и продолжительность разработки баз знаний является и будет являться одной из основных задач поставленных перед разработчиками баз знаний. Также повышение оценки качества баз знаний является одной из приоритетных задач разработчиков.

Проблема трудоемкости и продолжительности разработки интеллектуальных систем состоит в том, что, когда разработчик вводит новый фрагмент знаний, эксперт, перед тем как утвердить фрагмент на интеграцию в базу знаний, должен проверить этот фрагмент на наличие ошибок, правильности знаний и др.. Эксперт по какой-либо причине может пропустить какую-либо ошибку и утвердить фрагмент на интеграцию в базу знаний, что в свою очередь приведет к уменьшению оценки качества базы знаний.

Разработка средства семантического анализа вводимых фрагментов знаний позволит не искать эксперту ошибки в вводимом фрагменте знаний. Разработчик, перед тем как отправить фрагмент знаний эксперту на утверждение, просит систему провести семантический анализ разработанными средствами. Система ищет ошибки в вводимом фрагменте знаний и формирует конструкции с указаниями найденных ошибок. Таким образом средства семантического анализа вводимых фрагментов знаний находят все ошибки в вводимом фрагменте знаний, разработчик видит какие ошибки допустил и исправляет их, после чего направляет эксперту на утверждение. Такой алгоритм работы с средствами семантического анализа вводимых фрагментов знаний позволяет сократить время разработки базы знаний за счет уменьшения времени на поиск ошибок экспертом, а также увеличить оценку качества базы знаний за счет того, что фрагменты знаний будут интегрироваться без ошибок.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Целью данной диссертационной работы является разработка средств семантического анализа вводимого фрагмента знаний.

Для достижения указанной цели необходимо выполнить следующие задачи:

- Проанализировать существующие принципы коллективной разработки баз знаний;
- Спроектировать онтологии семантических ошибок
- Спроектировать алгоритмы работы средств поиска семантических ошибок;
- Реализовать спроектированные средства поиска семантических ошибок.

Научная новизна обусловлена отсутствием средств семантического анализа вводимых фрагментов знаний в современных моделях интеллектуальных систем.

Объектом исследования являются вводимые фрагменты баз знаний.
Предметом исследования являются модели, методы, алгоритмы и программные средства анализа вводимых фрагментов знаний.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационной работы, показана необходимость проведения исследования по данной теме.

Первая глава содержит аналитический обзор существующих средств разработки баз знаний и принципов коллективной разработки баз знаний.

В первом разделе рассматривается понятие интеллектуальной системы. Рассматривается структура интеллектуальной системы и откуда интеллектуальная система может получать знания. Также рассматривается понятия базы знаний, а также интеграция новых знаний в существующую базу знаний и какие возможные проблемы при этом могут возникнуть.

Во втором разделе осуществляется анализ моделей представления знаний. Приводятся языки представления знаний. Приводятся сильные и слабые стороны языков представления знаний.

В третьем разделе проводится анализ средств коллективной разработки. Приводятся существующие подходы и инструментальные средства. Приводятся пользователи каждого из существующих подходов и инструментальных средств, а также роли этих пользователей. Также приводятся достоинства и недостатки каждого из существующих подходов и инструментальных средств к коллективной разработке баз знаний.

В четвертом разделе приводится структура баз знаний Технологии OSTIS (Open Semantic Technology for Intelligent Systems). Рассказывается декомпозиция унифицированной семантической модели базы знаний. Приводится структуризация, с точки зрения разработки, на разделы и подробно рассматривается каждый из этих разделов. Также рассматривается механизм согласования фрагментов знаний.

В пятом разделе уточняется постановка задачи на конкретные задачи разработки средств.

Во второй главе рассматривается из чего состоит подсистема анализа вводимых фрагментов. Производится разбиение семантических ошибок на более подробные. Проектируются онтологии семантических ошибок.

В первом разделе приводится структура предметных областей базы знаний подсистемы анализа вводимых фрагментов. Показывается

разбиение семантических ошибок на более подробные: онтологии неполноты, онтологии противоречий.

Во втором разделе производится детальная классификация ошибок неполноты. Проектируются онтологии ошибок неполноты. Рассматриваются примеры каждой из ошибок неполноты в вводимых фрагментах знаний. Рассматривается ответ системы разработчику о наличии ошибки неполноты в вводимом фрагменте знаний.

В третьем разделе производится детальная классификация ошибок противоречий. Проектируются онтологии ошибок противоречий. Рассматриваются примеры каждой из ошибок противоречий в вводимых фрагментах знаний. Рассматривается ответ системы разработчику о наличии ошибки противоречия в вводимом фрагменте знаний.

В третьей главе рассматривается решатель задач подсистемы анализа вводимых фрагментов знаний. Производится декомпозиция решателя задач на абстрактные агенты. Проектируется пошаговый алгоритм работы для каждого абстрактного агента.

В первом разделе приводится пошаговый алгоритм работы абстрактного sc-агента семантического анализа фрагмента знаний.

Во втором разделе приводится декомпозиция абстрактного агента поиска ошибок в предложении. Также приводится пошаговый алгоритм работы каждого из агентов декомпозиции абстрактного агента поиска ошибок в предложении.

В третьем разделе приводится пример сценария добавления разработчиком нового фрагмента знаний. Показывается как система должна реагировать на добавление агента, как происходит поиск ошибок, а также приводится ответ системы для разработчика о выполнении семантического анализа.

В заключению подведены итоги к каждой главе. Приведены выводы чего позволят достичь разработанные средства семантического анализа вводимых фрагментов знаний. Поставлены дальнейшие планы развития средств семантического анализа вводимых фрагментов знаний.

ЗАКЛЮЧЕНИЕ

В рамках данной диссертационной работы были проанализированы существующие модели разработки баз знаний и принципы коллективной разработки баз знаний.

Был проведен анализ семантических ошибок и были выделены общие семантические ошибки в вводимых фрагментах знаний. Разработаны онтологии этих ошибок в вводимых фрагментах знаний, а также разработаны пошаговые алгоритмы работы абстрактных агентов поиска семантических ошибок в вводимых фрагментах знаний.

Разработанные средства позволяют в разы сократить время разработки баз знаний и уменьшить количество ошибок в базе знаний, что приводит к повышению качества баз знаний, а значит и самих интеллектуальных систем.

При дальнейшем развитии будут добавлены новые специализированные онтологии ошибок. Также будут добавлены алгоритмы работы агентов поиска этих ошибок. Также значительное внимание планируется уделить общему процессу автоматизации семантического анализа вводимого фрагмента знаний.