

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

УДК 004.424.42+004.89

Бруцкий
Дмитрий Сергеевич

Алгоритмы поиска полностью представленных фрагментов знаний для
модели унифицированного семантического представления знаний

Автореферат
на соискание степени магистра технических наук
по специальности 1-31 80 10 «Теоретические основы информатики»

Научный руководитель
Ивашенко Валерьян Петрович
кандидат технических наук

Минск 2020

ВВЕДЕНИЕ

В эпоху бурного развития сферы интеллектуального труда и увеличения объемов информационных потоков автоматизация интеллектуального труда имеет большое экономическое и промышленное значение. Традиционные подходы к автоматизации становятся менее эффективными из-за роста сложности и разнообразия решаемых задач. Применение методов, средств искусственного интеллекта призвано помочь в решении таких сложностей за счет повышения гибкости систем автоматизации и применения алгоритмов самообучения.

Важной задачей при внедрении эффективных программных систем обработки знаний является поиск в базах знаний. Кроме эффективности поиска не менее важной задачей является выбор наглядного и выразительного представления знаний. В качестве такого представления в статье используется SC-код предложенный в рамках технологии OSTIS.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Цель работы заключается во внедрении программных средств для поддержки поиска полностью представленных фрагментов знаний.

Ниже перечислены **задачи**, поставленные в рамках работы.

1. Произвести аналитический обзор существующих подходов к поиску знаний и программных систем поиска знаний.
2. Спроектировать модель программных средств и алгоритмов поиска полностью представленных фрагментов знаний.
3. Реализовать компоненты программных средств поддержки поиска полностью представленных фрагментов баз знаний.

Объект исследования - программные системы обработки знаний.

Предмет исследования - методы и алгоритмы поиска полностью представленных фрагментов знаний.

Предложенные в работе модели и алгоритмы интеграции знаний разработаны на базе графодинамической модели обработки знаний, в основе которой лежат однородные семантические сети с теоретико-множественной интерпретацией. Для решения поставленных задач использовались также методы теории графов и теории групп.

Новизна полученных результатов:

1. Двухфазный алгоритм вычисления канонической формы sc-фрагментов. Данный алгоритм основан на алгоритме предложенным Ивашенко В.П. Отличиями от базового алгоритма являются исключение фазы выделения вершин с петлями и использование алгоритма Nauty для вычисления разметки произвольного графа. Данный алгоритм применен к поиску в БЗ OSTIS систем.
2. Коллектив агентов, реализующих поиск в БЗ OSTIS систем. Особенностью агентов является использование алгоритма вычисления канонической формы sc-фрагментов для ускорения поиска в БЗ.

Положения, выносимые на защиту

1. Разработан двухфазный алгоритм вычисления канонической формы sc-фрагментов. Данный алгоритм основан на алгоритме предложенным Ивашенко В.П. Отличиями от базового алгоритма являются исключение фазы выделения вершин с петлями и использование алгоритма Nauty для вычисления разметки произвольного графа. Данный алгоритм применен к поиску в БЗ OSTIS систем.
2. Разработан коллектив агентов, реализующих поиск в БЗ OSTIS систем. Особенностью агентов является использование алгоритма вычисления канонической формы sc-фрагментов для ускорения поиска в БЗ.

Апробация результатов диссертации

Материалы работы докладывались и обсуждались на 56-й Научной конференции аспирантов, магистрантов и студентов учреждения образования «Белорусский Государственный Университет Информатики и Радиоэлектроники» (Минск, 2020).

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

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

Первая глава включает 2 раздела.

В **первом разделе** осуществляется анализ современных систем поиска знаний и анализ подходов применяемых при их реализации. Показано, что задача нахождения изоморфного вложения графа и определения изоморфизма графов является наиболее общей формой поиска знаний.

Во **втором разделе** осуществляется анализ существующих алгоритмов решающих задачи нахождения изоморфного вложения и определения изоморфизма графом.

Во **второй главе** производится преобразование программных компонентов поиска полностью представленных фрагментов знаний.

В **первом разделе** осуществляется проектирование компонентов, которые позволяют вычислять каноническую форму заданного sc-фрагмента.

Во **втором разделе** осуществляется проектирование программных компонентов использующих каноническую форму sc-фрагментов для поиска в БЗ.

Целью **третьей главы** является реализация компонентов спроектированных в предыдущей главе.

В **первом разделе** приводится перечень технологий используемых при разработке.

Во **втором разделе** рассмотрены детали реализации алгоритма вычисления канонической формы sc-фрагментов. Показана структура данных оптимизированная для работы с ними.

В **третьем разделе** показан сценарий работы системы поиска полностью представленных sc-фрагментов, интегрирующийся с OSTIS системой.

В **четвертой главе** представлены результаты вычислительных экспериментов показывающие временные характеристики работы разработанных компонентов поиска полностью представленных фрагментов знаний.

ЗАКЛЮЧЕНИЕ

В рамках работы произведен анализ систем поиск знаний. Показано, что задача нахождения изоморфного вложения графа и определения изоморфизма графов являются ключевыми. Проведен анализ известных

алгоритмов обнаружения изоморфного вложения графов и сравнение графов (обнаружение изоморфизма).

Обнаружено, что вычислительные сложности лучших алгоритмов сравнения графов и вычисления канонической формы графа одинаковы.

Учитывая это, показано, что поиск в библиотеке размеченных графов работает быстрее, чем поиск основанный на прямом сравнении графов ($O(f(n)+n^2\log(m))$ против $O(mf(n))$, где $f(n)$ – обозначение вычислительной сложности сравнения двух графов и сложности вычисления канонической разметки графа, n -количество вершин в графе шаблоне, m - количество графов в библиотеке).

Спроектированы и реализованы программные средства вычисления канонических форм sc-фрагментов. Разработан гибридный алгоритм, первая ступень которого имеет полиномиальную вычислительную сложность и может запускаться клиентской стороне серверного приложения. Данный алгоритм способен вычислить каноническую форму для большого, но не полного множества графов, поэтому второй ступенью выступает экспоненциальный алгоритм, решающий задачу для любого входного графа.

Спроектированы и реализованы компоненты программных средств поддержки поиска полностью представленных фрагментов баз знаний. Данные компоненты реализуются в виде sc-агентов, что позволяет интегрировать их в коллектив агентов OSTIS системы. Каноническая форма размеченных sc-фрагментов сохраняется в виде строкового содержимого sc-ссылки, что позволяет находить изоморфные фрагменты за минимальное количество элементарных операций ассоциативного поиска sc-памяти.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Бруцкий, Д.С. Алгоритмы поиска полностью представленных фрагментов знаний для модели унифицированного семантического представления знаний / Д.С. Бруцкий // 56-я научная конференция аспирантов, магистрантов, студентов учреждения образования "БГУИР" : материалы конференции по направлению 2: Информационные технологии и управления - г.Минск, 2020 – С. 7