

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

УДК 004.896

Коршунов
Руслан Андреевич

Модели и средства автоматического наполнения баз знаний из открытых
структурированных источников

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

Научный руководитель
Шункевич Даниил Вячеславович
кандидат технических наук, доцент

Минск 2022

Сегодня базы знаний играют важную роль и используются в различных сферах жизнедеятельности общества. Однако успешность использования базы знаний во многом зависит от ее наполненности и актуальности. Наполнение базы знаний — это трудоемкий процесс, требующий человеческих и временных затрат. Источником знаний могут выступать неструктурированные и структурированные источники. И если при работе с неструктурированными источниками необходимо наличие эксперта для их успешного внедрения в базу знаний, то процесс обработки уже структурированных источников может быть сокращен и автоматизирован.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель работы: снизить трудоемкость наполнения БЗ из открытых структурированных источников за счет автоматизации указанного процесса.

Задачи исследования:

1. анализ существующих и наиболее эффективных моделей и средств наполнения БЗ;
2. разработка подсистемы автоматизации процесса наполнения БЗ из структурированных источников;
3. реализация подсистемы автоматизации процесса наполнения БЗ из структурированных источников;
4. анализ полученных результатов.

Объект исследования: модели и средства автоматического наполнения БЗ из открытых структурированных источников.

Предмет исследования: существующие подходы к автоматизации наполнения БЗ.

Апробация результатов диссертации

Результаты исследований были продемонстрированы на:

- 57-ой научной конференции аспирантов, магистрантов и студентов БГУИР;
- международной научно-технической конференции OSTIS-2021.

Опубликованность результатов исследования

Результаты исследований публиковались в:

- информационные технологии и управление: материалы 57-й научной конференции аспирантов, магистрантов и студентов (2021);
- сборник статей “OSTIS-2021”.

Структура и объем диссертации

Диссертация имеет следующий состав:

- введение;
 - анализ существующих моделей и средств наполнения баз знаний;
 - реализация подсистемы автоматизации процесса наполнения базы знаний знаниями из заданного структурированного источника;
 - оценка эффективности и необходимости разработанной системы.
- Полный объем диссертации составляет 52 страницы, 44 иллюстрации, 26 библиографических источников (включая собственные публикации).

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первая глава рассматривает существующие подходы к наполнению баз знаний. В рамках главы было сделано следующее:

- дано определение базы знаний и сформулированы основные проблемы ее наполнения;
- приведена классификация основных типов источников данных, сделан вывод о простоте использования структурированных форматов представления знаний в качестве источника нового для базы знаний;
- рассмотрены несколько подходов к наполнению базы знаний;
- проведен сравнительный анализ подходов и на его основе сформулирован вывод.

Вторая глава описывает разработку подсистемы автоматизации процесса наполнения БЗ знаниями из заданного структурированного источника. В рамках главы было выполнено следующее:

- сформулирован предлагаемый подход;
- описана предметная область понятий и отношений, которые будут использоваться системой для представления полученных знаний;
- описана архитектура системы, ее основные элементы.

Третья глава содержит результаты реализации подсистемы. Глава содержит следующее:

- перечень используемых технологии с подробным описанием каждой из них;
- описание процесса работы модуля загрузки знаний из заданного структурированного источника;
- описание процесса интеграции знаний из внешнего источника.

Четвертая глава содержит оценку эффективности и необходимости разработанной системы.

ЗАКЛЮЧЕНИЕ

Целью работы являлось снижение трудоемкости наполнения БЗ из открытых структурированных источников за счет автоматизации указанного процесса.

Для ее достижения были решены задачи, по результатам которых можно сделать следующие выводы.

Рассмотренные подходы к наполнению базы знаний связаны с понятием RDF, т.к. либо он является форматом представления полученных знаний, либо лежит в основе других форматов. RDF, при всех своих преимуществах имеет ряд недостатков, которые впоследствии могут стать помехой для наполнения базы знаний.

Был сформулирован собственный подход и на его основе разработан прототип системы для автоматизации процесса наполнения базы знаний. В качестве примера источника был выбран ресурс WikiData как наиболее простой и широкодоступный, имеющий формат представления знаний, основанный на тройке “субъект-предикат-объект”. В качестве базы знаний была выбрана база знаний ostis-системы.

Предложенный подход был реализован. Его можно разбить на два основных этапа. На первом этапе специальный модуль “подтягивает” полученные из WikiData знания и трансформирует их во внутренние конструкции ostis-системы. Затем агент интеграции знаний на основе имеющихся правил встроил полученные знания в уже существующую иерархию.

Библиотека

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Коршунов, Р. А. Средство автоматизированного пополнения базы знаний OSTIS-системы на основе знаний из внешних источников / Р. А. Коршунов, А. Г. Загорский // Информационные технологии и управление : материалы 57-ой научной конференции аспирантов, магистрантов и студентов по направлению 2, Минск, 19-23 апреля 2021 года / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск, 2021. – С. 14.

2. Korshunov, R. A. Ontological approach to the integration of knowledge from external sources / R. A. Korshunov, M. E. Sadouski, A. G. Zagorskij // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2021) : сборник научных трудов / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : В. В. Голенков [и др.]. – Минск, 2021. – Вып. 5. – С.117–122.

Библиотека БГУИР