

УДК 004.822

ИНКРЕМЕНТАЛЬНАЯ АКТУАЛИЗАЦИЯ СОДЕРЖИМОГО СЕМАНТИЧЕСКОЙ СЕТИ

Ильинская О.В. (Университет ИТМО)

Научный руководитель – старший преподаватель факультета программной инженерии и компьютерной техники, Цопа Е.А.
(Университет ИТМО)

Введение. Актуализация семантической сети нужна для того, чтобы обновлять ее знания и улучшать ее способность к анализу и пониманию текстов и других данных. Семантическая сеть представляет собой структуру, которая связывает понятия и их отношения в форме графа. Эта структура может использоваться для анализа текста, поиска информации и других задач, которые требуют понимания смысла и связей между различными элементами данных. Однако знания, которыми располагает семантическая сеть, могут устаревать или стать неполными в результате изменений в языке, культуре, технологиях и других областях знаний. Поэтому актуализация семантической сети является важным процессом, который позволяет ей адаптироваться к новым условиям и оставаться актуальной и полезной. В рамках процесса актуализации семантической сети могут использоваться различные подходы, включая анализ новых текстов и данных, обучение на новых примерах, ручное обновление знаний и другие методы. Эти подходы могут помочь расширить семантическую сеть и обеспечить ее способность к анализу новых данных и ситуаций.

Основная часть. На первом этапе работы необходимо было выбрать метод для актуализации графово-реляционной структуры. В ходе этого этапа были рассмотрены следующие подходы: дифференциальная актуализация, инкрементальная актуализация, транзакционная актуализация и алгоритм актуализации, основанный на машинном обучении. В результате был выбран инкрементальный подход, так как он позволяет обновлять структуру постепенно, только в тех местах, где были внесены изменения, минимизируя при этом количество необходимых вычислений и времени на обработку данных. Этот подход позволяет быстро и эффективно обрабатывать изменения в структуре и избегать пересчета всей структуры заново при каждом изменении, что является значительным преимуществом по сравнению с другими подходами. Кроме того, инкрементальный подход позволяет более гибко управлять актуализацией структуры и учитывать особенности конкретной задачи или контекста.

Далее были рассмотрены принципы инкрементальной актуализации на примере двух известных технологий — системы контроля версий Git и реляционной базы данных PostgreSQL. Git использует инкрементальный подход для хранения и обновления данных, где каждый коммит содержит только те изменения, которые были внесены в файлы после предыдущего коммита. При этом, при выполнении команды `git add`, файлы добавляются в индекс, и коммит содержит только те изменения, которые были внесены в эти файлы. Такой подход позволяет Git быстро обновлять файлы и хранить только те данные, которые действительно изменились. В свою очередь, реляционная база данных PostgreSQL использует инкрементальную актуализацию для обновления данных. Данные в PostgreSQL хранятся в виде таблиц, где при обновлении данных только измененные строки обновляются, а не все строки таблицы. Кроме того, PostgreSQL использует индексы для ускорения доступа к данным, и обновляет только те строки, которые были изменены. Этот подход позволяет PostgreSQL быстро и эффективно обновлять данные в таблицах и ускорять работу с базой данных.

В результате анализа существующих решений была предложена собственная реализация инкрементального подхода к актуализации данных, основанная на концепциях, используемых в обоих рассмотренных программных продуктах. Эта реализация была адаптирована к использованию в системе хранения на базе семантической сети, характерными особенностями которой является использование при хранении данных элементов как графовых, так и реляционных моделей.

Предложенная концепция была реализована в виде программного модуля на языке программирования C++, который был интегрирован в состав существующей системы хранения данных на основе семантической сети и позволил осуществлять инкрементальную синхронизацию содержимого этой сети в внешними источниками данных. Проведённое функциональное и нагрузочное тестирование доказало пригодность разработанного модуля к использованию в продукционных системах обработки текста на естественном языке.

Вывод. В результате исследования был предложен алгоритм инкрементальной актуализации графово-реляционной структуры, представляющей собой семантическую сеть. Предложенный алгоритм был реализован в виде программного модуля, интегрированного в состав системы хранения на базе семантической сети. Проведённое тестирование доказало эффективность предложенного подхода и возможность его использования в реальных системах.

Список используемых источников:

1. John F. Sowa. Semantic Networks. [Электронный ресурс]. Режим доступа: <https://www.jfsowa.com/pubs/semnet.htm>
2. John F. Sowa, John A. Goguen. Conceptual Structures: Information Processing in Mind and Machine
3. O'Reilly. Graph databases, 2nd edition

Ильинская О.В. (автор)

Подпись



Цопа Е.А. (научный руководитель)

Подпись