

ПЕРЕХОД ОТ ОБРАБОТКИ ДАННЫХ К ОБРАБОТКЕ ЗНАНИЙ КАК КЛЮЧЕВОЙ ЭТАП В РАЗВИТИИ ТЕХНОЛОГИЙ ДЛЯ РАЗРАБОТКИ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ НОВОГО ПОКОЛЕНИЯ

Зотов Н.В.

*Белорусский государственный университет информатики и радиоэлектроники,
Минск, Республика Беларусь*

Голенков В.В. – д-р техн.наук, профессор,
профессор кафедры ИИТ

Аннотация. В статье обосновывается необходимость перехода от современных баз данных к базам знаний и моделям представления и обработки знаний для разработки интеллектуальных систем нового поколения. Обосновывается преимущество графовых моделей данных в плане выразительной силы, гибкости, производительности и семантического поиска над реляционными моделями. В работе сформулированы требования к технологиям для создания баз знаний и интеллектуальных систем на их основе.

Ключевые слова: представление данных и знаний, модель данных, база данных, графовая модель данных, графовая база данных, база знаний.

Введение

В настоящее время информационные системы подвергаются интенсивной интеллектуализации. В первую очередь, это вызвано повышением уровня сложности решаемых задач. Интеллектуализация информационных систем требует от технологий разработки программных систем учета слабоформализуемой, возможно не полностью определенной, нечеткой, темпоральной, пространственно-распределенной информации и, как следствие, получения структурированных, слабоструктурированных и неструктурированных данных [1, 2]. Увеличение числа интеллектуальных задач обработки больших объемов данных во всех сферах деятельности человека приводит к потребности создания универсальных средств хранения, представления и обработки сложноструктурированной информации.

Цель данной работы заключается в обосновании необходимости перехода от современных баз данных и принципов организации и обработки данных в них к базам знаний и моделям представления и обработки знаний, а также определению требований к технологиям для разработки интеллектуальных систем.

Основная часть

Хранилища данных, используемые в большинстве программных компьютерных систем, в настоящее время основаны на реляционной модели данных. Примерами таких систем являются: Платформа SMILA (SeMantic Information Logistic Architecture), Платформа Teradata (Teradata Aster Discovery Platform), реализующие реляционную, поколоночную и гибридную модели хранения записей в базе данных с массово-параллельной архитектурой и другие.

Необходимость решения высокоинтеллектуальных задач стимулирует переход от традиционных реляционных баз данных к их графовым аналогам. Это объясняется не столько эффективностью организации памяти и обработки данных в графовых базах данных, сколько важностью представления конфигураций связей (то есть смысла) между ними. Подробное изъяснение принципов организации графовых данных в базах данных можно найти в работе авторов популярной графовой СУБД Neo4j [3], Базы знаний DBpedia [4] и других.

Мотивация перехода от реляционных баз данных к графовым базам данных также объясняется преимуществами организации модели памяти и обработки данных в них [1-3, 5-8]:

Графовые модели данных имеют огромную выразительную силу. Графовые базы данных предлагают гибкую модель данных и способ их представления. Графы аддитивны, это обеспечивает гибкость добавления новых связей между данными, новые узлы и новые подграфы к уже существующей структуре графа, не нарушая ее целостности и связности.

Многообразие форм представления данных минимизируется за счет уменьшения количества синтаксических аспектов, учитываемых при хранении данных и использовании их

в базах данных, поскольку графовые модели данных позволяют записывать различные виды знаний одним и тем же способом.

Чтобы понимать смысл знаний, необходимо представлять эти знания в понятной форме для каждого: и для человека, и для системы. Говоря об унификации представления всех видов знаний, важным считается использование графовых баз данных не просто как средств для хранения структурированных данных, а для хранения семантически целостных и связанных между собой знаний.

Производительность обработки данных улучшается на один или более порядков при представлении данных в виде графов, что объясняется свойствами самого графа. В отличие от реляционных баз данных, где производительность запросов с увеличением интенсивности запросов ухудшается по мере увеличения набора данных, производительность графовой модели данных остается постоянной, даже когда набор данных растет. Это связано с тем, что обработка данных локализуется в некоторой части графа. В результате время выполнения каждого запроса равно пропорционально только размеру части графа, пройденной для удовлетворения этого запроса, а не размеру всего графа.

Графовые базы данных позволяют эффективно выполнять семантический поиск, то есть находить данные и информацию на основе связей между ними, что способствует улучшению качества и точности поисковых запросов, а также обеспечивает более глубокое понимание этих связей и зависимостей между данными.

Вообще говоря, современные базы данных сильно ограничивают развитие существующих систем и технологий их разработки. При помощи обычных баз данных невозможно интегрировать существующие системы и их компоненты, поскольку это порождает не только дублирование информации на стыках между системами, но и влечёт за собой высокую сложность сопровождения этих систем, что требует как высоких требований к их разработчикам, так и серьёзных ресурсов для решения таких задач [2, 6, 7, 8].

Для преодоления проблем, связанных с использованием баз данных, необходима разработка и внедрение:

- моделей и средств, ориентированных на унифицированное семантическое представление и обработку знаний;
- стандартов представления и согласования знаний в базах знаний;
- стандартов обработки и совместного использования знаний в базах знаний;
- баз знаний со стандартизированными принципами и правилами идентификации, спецификации и размещения знаний в них;
- баз знаний с единой унифицированной семантической основой, с помощью которой интеграция знаний сводится к "склейке" знаний независимо от их вида и сложности;
- баз знаний, позволяющих описывать разные виды знаний независимо от программных и аппаратных средств, на которых реализованы данные базы знаний;
- баз знаний, позволяющих представлять и обрабатывать знания процедурного вида;
- моделей и средств для обработки знаний, позволяющих быстро и гибко внедрять новые модели решения задач;
- моделей и средств, позволяющих гибко создавать и расширять существующие компоненты интеллектуальных систем;
- моделей и средств для реализации согласованного и целенаправленного взаимодействия коллективов интеллектуальных систем;
- человеко-ориентированных моделей и средств визуализации знаний в базах знаний в понятной и доступной форме для любого человека;
- методов, методик и стандартов поддержки разработки баз знаний и моделей обработки знаний в них.

Другими словами, необходимо создание технологий, объединяющих все необходимые методы и средства для разработки семантически совместимых баз знаний и интеллектуальных систем, использующих их, а также методик и средств, позволяющих создавать с помощью этих технологий такие базы знаний.

Заключение

Переход от обработки данных к обработке информации является закономерным этапом развития информационных технологий, а также является ключевым этапом в развитии

технологий для разработки интеллектуальных систем нового поколения. Одной из технологий, позволяющих разрабатывать такие системы, является Технология OSTIS [9]. На базе Технологии OSTIS уже реализован прототип Платформы OSTIS [10], позволяющей создавать базы знаний со всеми выше перечисленными свойствами.

Список использованных источников:

1. Ryen, V. Building semantic knowledge graphs from (semi-) structured data: a review / Ryen V., Soylyu A., Roman D. // *Future Internet*. – 2022. – Т. 14. – №. 5. – С. 129.
2. Barnaghi, P. From data to actionable knowledge: Big data challenges in the web of things [Guest Editors' Introduction] / Barnaghi P., Sheth A., Henson C. // *IEEE Intelligent Systems*. – 2013. – Т. 28. – №. 6. – С. 6-11.
3. Robinson, I. Graph databases / Robinson, I. – O'Reilly Media, Inc., 2015.
4. Lehmann, J. Dbpedia—a large-scale, multilingual knowledge base extracted from wikipedia / Lehmann J. et al. // *Semantic web*. – 2015. – Т. 6. – №. 2. – С. 167-195.
5. Han, J. Semantic networks for engineering design: state of the art and future directions / Han J. et al. // *Journal of Mechanical Design*. – 2022. – Т. 144. – №. 2. – С. 020802.
6. Kellogg, C. From data management to knowledge management / Kellogg C. // *Computer*. – 1986. – Т. 19. – №. 01. – С. 75-84.
7. Alavi, M. Knowledge management systems: issues, challenges, and benefits / Alavi M., Leidner D. // *Communications of the Association for Information systems*. – 1999. – Т. 1. – №. 1. – С. 7.
8. Yahya, M. I. Semantic web and knowledge graphs for industry 4.0 / Yahya M., Breslin J. G., Ali M. // *Applied Sciences*. – 2021. – Т. 11. – №. 11. – С. 5110.
9. Golenkov, V. Artificial intelligence standardization is a key challenge for the technologies of the future / Golenkov, V. et al. // *International Conference on Open Semantic Technologies for Intelligent Systems*. – Cham : Springer International Publishing, 2020. – С. 1-21.
10. Zotov, N. Software platform for next-generation intelligent computer systems / N. Zotov // *Open Semantic Technologies for Intelligent Systems (OSTIS-2022)* / БГУИР ; редкол.: В. В. Голенков [и др.]. – Минск, 2022. – Вып. 6. – С. 297–326.