

Секция «Системная биология, логика и искусственный интеллект»

Подсекция «Логика и искусственный интеллект»

DOI 10.18699/sblai2023-01

О построении графа знаний научной предметной области в семантической библиотеке

Атаева О.М.*, Серебряков В.А.** , Тучкова Н.П.***

ФИЦ «Информатика и управление» РАН, Москва, Россия

* oli.ataeva@gmail.com, ** serebrvas@gmail.com, *** natalia_tuchkova@mail.ru

В работе рассматривается подход к построению графа знаний на основе онтологического представления научных предметных областей средствами семантической библиотеки LibMeta [1]. Изложение опирается на понятия, связанные с информацией и интеллектуальным анализом данных, такие как знание, извлечение знаний, онтология предметной области, научная предметная область, тезаурус, семантическая цифровая библиотека, информационная потребность пользователя, метод онтологического проектирования и собственно граф знаний.

Граф знаний представляет собой базу знаний, которая использует модель данных или топологию, структурированную графом, для интеграции данных. Назначение графа знаний состоит в извлечении знаний из источников данных [2, 3] и представлении этих знаний в виде модели (графа), которая определяет сущности и отношения между ними. К числу актуальных задач относятся построение графа знаний научной предметной области и выбор технологии его построения. Новая вершина тезауруса предметной области появляется, если при анализе (предобработке) первичных данных (статей) выявлен новый термин, обозначающий понятие этой предметной области, но ранее он не встречался (или отсутствовал) в тезаурусе. Задача пополнения может трансформироваться в задачу встраивания в граф [4], когда один из узлов заменяется подграфом.

На основе LibMeta есть решение проблем построения графа знаний в онтологическом проектировании [5] с достраиванием на основании информационного запроса от пользователя. Для математических знаний аналогичная идея была заложена еще задолго до появления технологии построения графа знаний как способ обеспечения доступа исследователей к научным результатам во Всемирной математической библиотеке (WDMML [6]). Онтология цифровой библиотеки LibMeta определяет структуру данных контента библиотеки. Каждому элементу данных, загруженных в библиотеку, можно сопоставить вершину онтологии, определяющую положение элемента данных в онтологии. На основании связей онтологии и связей, определенных на этапе проектирования, можно построить граф данных. Данные предметной области можно представить как граф, структура которого определяется онтологией, вершины (статьи) – экземпляры элементов онтологии, связи – связи тезауруса. Дальнейшие исследования посвящены применению алгоритмов искусственного интеллекта для извлечения знаний на основе энциклопедии и наполнения библиотеки.

Работа частично поддержана Российским научным фондом, проект 22-21-00449.

Литература

1. Ataeva O., Serebryakov V., Tuchkova N. Creating the Applied Subject Area Ontology by Means of the Content of the Digital Semantic Library, Lobachevskii J. Mathematics, 2022, vol. 43, no. 7. P. 1795-1804. Doi 10.1134/S1995080222100043.

2. Wang P., Jiang H., Xu J., Zhang Q., Knowledge graph construction and applications for web search and beyond. *Data Intelligence*. 2019;1(4):333-349.
3. Wang J. Math-KG: Construction and Applications of Mathematical Knowledge Graph, arXiv:2205.03772, 2022. Doi 10.48550/arXiv.2205.03772.
4. Yang H., Zhang L., Wang B., Yao T., Liu J. Cycle or Minkowski: Which is More Appropriate for Knowledge Graph Embedding? *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2021;2301-2310.
5. Lange C. Ontologies and languages for representing mathematical knowledge on the Semantic Web. *Semantic Web*. 2013;4(2):119-158.
6. Ion P.D.F., Watt S.M. The Global Digital Mathematics Library and the International Mathematical Knowledge. In: Geuvers H., England M., Hasan O., Rabe F., Teschke O. (Eds.). *Intelligent Computer Mathematics. CICM 2017. Lecture Notes in Computer Science*, Springer, Cham, 2017, vol. 10383. P. 56-69. Doi 10.1007/978-3-319-62075-6_5.

Building a knowledge graph of a scientific subject area in a semantic library

Ataeva O.M.*, Serebryakov V.A.***, Tuchkova N.P.***

FRC "Computer Science and Control" RAS, Moscow, Russia

* oli.ataeva@gmail.com, ** serebrvas@gmail.com, *** natalia_tuchkova@mail.ru

The paper considers an approach to building a knowledge graph based on the ontological representation of scientific subject areas using the semantic library LibMeta [1]. This work is based on concepts related to information and data mining, such as knowledge, knowledge extraction, domain ontology, scientific domain, thesaurus, semantic digital library, user information need, ontological design method and, in fact, the knowledge graph.

A knowledge graph is a knowledge base that uses a graph-structured data model or topology to integrate data. The purpose of a knowledge graph is to extract knowledge from data sources [2, 3] and represent this knowledge in the form of a model (graph) that defines entities and relationships between them. Among the urgent tasks are the construction of a knowledge graph of a scientific subject area and the choice of technology for its construction. A new top of the subject area thesaurus appears if, during the analysis (preprocessing) of primary data (articles), a new term denoting the concept of this subject area is identified, but it has not been previously encountered (or was absent) in the thesaurus. The replenishment problem can be transformed into a graph embedding problem [4] when one of the nodes is replaced by a subgraph.

For a knowledge graph based on LibMeta, there is a solution to the construction problems in ontological design [5] with completion based on an information request from the user. For mathematical knowledge, a similar idea was laid down long before the advent of knowledge graph technology as a way to provide researchers with access to scientific results in the World Mathematical Library WDMML [6]. The LibMeta digital library ontology defines the data structure of the library content. Each data element loaded into the library can be associated with an ontology vertex that determines the position of the data element in the ontology. Based on the ontology links and the links defined at the design stage, you can build a data graph. Domain data can be represented as a graph whose structure is determined by the ontology, vertices (articles) are instances of ontology elements, links are thesaurus links. Further research is devoted to the use of artificial intelligence algorithms for extracting knowledge from the subject area based on the encyclopedia and filling the library.

This work was supported by budget topics of the Ministry of Science and Higher Education of the Russian Federation 'Mathematical methods for data analysis and forecasting' and particular the Russian Science Foundation, grant No. 22-21-00449.

References

1. Ataeva O., Serebryakov V., Tuchkova N. Creating the Applied Subject Area Ontology by Means of the Content of the Digital Semantic Library, Lobachevskii J. of Mathematics, 2022, vol. 43, no. 7. P. 1795-1804. Doi 10.1134/S1995080222100043.
2. Wang P., Jiang H., Xu J., Zhang Q., Knowledge graph construction and applications for web search and beyond. *Data Intelligence*. 2019;1(4):333-349.

3. Wang J. Math-KG: Construction and Applications of Mathematical Knowledge Graph, arXiv:2205.03772, 2022. Doi 10.48550/arXiv.2205.03772.
4. Yang H., Zhang L., Wang B., Yao T., Liu J. Cycle or Minkowski: Which is More Appropriate for Knowledge Graph Embedding? Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021;2301-2310.
5. Lange C. Ontologies and languages for representing mathematical knowledge on the Semantic Web. Semantic Web. 2013;4(2):119-158.
6. Ion P.D.F., Watt S.M. The Global Digital Mathematics Library and the International Mathematical Knowledge. In: Geuvers H., England M., Hasan O., Rabe F., Teschke O. (Eds.). Intelligent Computer Mathematics. CICM 2017. Lecture Notes in Computer Science, Springer, Cham, 2017, vol. 10383. P. 56-69. Doi 10.1007/978-3-319-62075-6_5.

在語義庫中構建科學主題領域的知識圖譜

阿塔耶娃

Ataeva O.M.*, Serebryakov V.A.** , Tuchkova N.P.***

FRC CSC RAS, 莫斯科

* oli.ataeva@gmail.com, ** serebrvas@gmail.com, *** natalia_tuchkova@mail.ru

本文考慮了一種使用語義庫 LibMeta [1]

基於科學主題領域的本體表示來構建知識圖譜的方法。

這項工作基於與信息和數據挖掘相關的概念，例如知識、知識抽取、領域本體、科學領域、詞庫、語義數字圖書館、用戶信息需求、本體設計方法以及知識圖譜。

知識圖譜是使用圖結構的數據模型或拓撲來集成數據的知識庫。

知識圖譜的目的是從數據源 [2, 3]

中提取知識，並以定義實體及其之間關係的模型（圖）的形式表示這些知識。

緊迫的任務包括構建科學學科領域的知識圖譜及其構建技術的選擇。

如果在分析（預處理）原始數據（文章）期間，識別出表示該主題領域概念的新詞，但該主題詞庫之前未曾遇到（或不存在）詞庫。

當其中一個節點被子圖替換時，補貨問題可以轉化為圖嵌入問題[4]。

對於基於 LibMeta 的知識圖譜，本體設計 [5]

中的構建問題有一個解決方案，它基於用戶的信息請求完成。

對於數學知識，早在知識圖譜技術出現之前就提出了類似的想法，作為一種方式為研究人員提供訪問世界數學圖書館 WDM [6] 中的科學成果的途徑。LibMeta

數字圖書館本體定義了數據結構 庫內容。

加載到庫中的每個數據元素可以與確定數據元素在本體中的位置的本體頂點相關聯。

基於本體鏈接和設計階段定義的鏈接，可以構建數據圖。

領域數據可以表示為一個圖，其結構由本體決定，頂點（文章）是本體元素的實例，鏈接是詞庫鏈接。

進一步的研究致力於使用人工智能算法從基於百科全書的學科領域中提取知識並填充圖書館。

文學

1. Ataeva O., Serebryakov V., Tuchkova N. Creating the Applied Subject Area Ontology by Means of the Content of the Digital Semantic Library, Lobachevskii J. of Mathematics, 2022, vol. 43, no. 7. P. 1795-1804. Doi 10.1134/S1995080222100043.
2. Wang P., Jiang H., Xu J., Zhang Q., Knowledge graph construction and applications for web search and beyond. Data Intelligence. 2019;1(4):333-349.
3. Wang J. Math-KG: Construction and Applications of Mathematical Knowledge Graph, arXiv:2205.03772, 2022. Doi 10.48550/arXiv.2205.03772.
4. Yang H., Zhang L., Wang B., Yao T., Liu J. Cycle or Minkowski: Which is More Appropriate for Knowledge Graph Embedding? Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021;2301-2310.
5. Lange C. Ontologies and languages for representing mathematical knowledge on the Semantic Web. Semantic Web. 2013;4(2):119-158.

DOI 10.18699/sblai2023-02

***Логические и вычислительные аспекты искусственного интеллекта
в семантическом моделировании***

Гончаров С.С.

Институт математики им. С.Л. Соболева СО РАН, Новосибирск, Россия

s.s.goncharov@math.nsc.ru

Вопросы развития исследований в области искусственного интеллекта (ИИ) с момента создания вычислительных машин всегда занимали внимание ученых и разработчиков вычислительной техники с целью ее применения в различных сферах человеческой деятельности и развитии современного общества. Начало исследованиям было положено в работах Я.М. Барздиня, Э. Блюм по извлечению предельных функций по последовательностям начальных данных, ими были построены первые контрпримеры, а также были развиты другие ослабленные требования поиска, которые вылились в машинное обучение. В последнее время на пути успехов в задачах распознавания образов и построения систем обработки естественного языка, систем перевода, систем обработки символьной информации вновь возникли идеи расширения возможностей применения машин в проблемах науки, инженерии и других ранее присущих лишь человеку видах деятельности.

Логические исследования лежат в основе современных ИТ-технологий и развития методов доверительного ИИ. Высокая оценка и признание выдающейся роли логики в развитии научного подхода к исследованию окружающего нас мира и в развитии новых технологий послужили основанием объявления ЮНЕСКО 14 января Днем логики. Генеральным директором ЮНЕСКО Одрэ Азуле было в 2018 году отмечено, что «В XXI веке логика как дисциплина оказалась, как никогда прежде, современной и актуальной для наших обществ и наших экономик. Так, вычислительная техника и цифровая среда представляют собой технологии, изначально основанные на принципах логического и алгоритмического мышления». На заседании Президиума РАН по вопросу ИИ была поставлена проблема доверительного ИИ в разработке научных основ и технологий ИИ, которая требует комплексных исследований, начиная с методологии и заканчивая проблемами безопасности использования сложных программных и машинных комплексов.

В ИМ СО РАН под руководством академиков Ю.Л. Ершова и С.С. Гончарова были инициированы исследования [1] по построению не императивного, а дескриптивного языка программирования на основе логически обоснованного задачного подхода и построения вычислимости на абстрактных моделях. Построен логический язык программирования и его полиномиальная версия [2] на основе развития теории неподвижных точек монотонных логических операторов Барвайса [3], что позволило успешно реализовать программные системы, получить успешно реализованные программы для современных задач в телекоммуникациях, управлении системами платежей, анализе генетической информации и других, которые имеют ясную логическую основу и позволяют делать корректные их модификаций для новых условий и задач.

Сама постановка задачи начинается с построения онтологии предметной области и ее трех этапов формализации в рамках и средствами логического языка, предложенных алгоритмов и их программных реализаций, а также анализа взаимоотношений этих трех уровней формализаций. В основу построения математической теории семантического подхода были положены идеи языка программирования Пролог, теория допустимых множеств, классическая теория вычислимости, включая относительную вычислимость с