

УДК 004.65

ПОБУДОВА ПРЕДСТАВЛЕННЯ ЗНАНЬ НА ОСНОВІ ФАКТОР-ГРАФУ З УРАХУВАННЯМ ЧАСТОТИ ДОСТУПУ ДО ЙОГО ВЕРШИН

Дородних Д. О.

Науковий керівник – проф. Чалий С. Ф.

Харківський національний університет радіоелектроніки, каф. ШІ
м. Харків, Україна

тел.: +38(096) 537-09-02, e-mail: dimadorodnykh@gmail.com

In today's fast-paced world, everyone wants easy access to accurate information. Users would rather use knowledge base and find an immediate solution to their problems than contact the help desk. Formal representation of knowledge bases is usually based on the use of graphs. One of the important areas of development of knowledge is the construction of factor graphs. The factor graph concept is used, in particular, in the DeepDive knowledge base. However, the existing architecture does not consider the frequency of use of the vertices of the graph. The proposed combined method increases the efficiency of the graph knowledge base built on factor graphs.

У сучасному онлайн-світі користувачі широко використовують бази знань для вирішення своїх задач, вважаючи такий спосіб більш ефективним, ніж взаємодія із службами підтримки споживачів [1]. Тому, на сьогодні, значна увага приділяється розробці нових автоматизованих методів побудови і використання бази знань (БЗ). Відмінність сучасних автоматизованих баз знань від традиційних полягає в можливостях створення та використання знань програмними засобами, з мінімальною участю людини. Участь експертів-людей зазвичай передбачається лише для уточнення уже сформованих знань. Автоматизована база знань є основою для побудови інтелектуальних інформаційних систем різних типів: пошукових, систем електронної комерції, інформаційних систем управління підприємством, тощо. Можливості автоматизації роботи такої БЗ значною мірою залежать від вибраного представлення знань [2]. Сучасні автоматизовані БЗ використовують переважно графове представлення знань. Вперше підхід до представлення знань у формі графу (knowledge graph) був представлений компанією Google у 2012 році [3].

Одним із важливих напрямків розробки графового представлення знань є побудова й використання фактор-графів. Концепція фактор-графу застосована, зокрема, в базі знань DeepDive [4]. Фактор-граф є різновидом імовірнісної графічної моделі [5]. Він має два типи вузлів: змінні та фактори. Змінні відображають вузли які можуть бути або змінними- "доказами", коли їхнє значення відоме, або змінними запиту, якщо їхнє значення потрібно передбачити. Фактори – це вузли, що визначають зв'язки між змінними графу. Кожен фактор може бути пов'язаний з

багатьма змінними. Важливість зв'язку між змінними задається через фактор-функцію. Кожна фактор-функція має вагу, пов'язану з нею. Вага описує ступінь впливу фактору на свої змінні у відносному виразі. Позитивна велика вага свідчить про впевненість у залежності, яку кодує фактор. Якщо вага велика і негативна, то знання, представлені відповідним фрагментом графу, швидше за все є неправильними. Вага може бути отримана з тренувальних даних або призначена вручну, на основі досвіду експерту. Такий підхід забезпечує пошук знань на графі і оцінку отриманих знань з урахуванням ваг факторів.

Однак існуюча схема представлення знань не враховує частоту використання вершин графу. При врахуванні частоти вершин виникає можливість підвищити ефективність пошуку знань у БЗ. В даному випадку використовується припущення, що дані та знання, які більш часто використовуються користувачами, є більш релевантними стосовно запиту. Такі дані повинні отримати пріоритет при формуванні результатів у БЗ.

Сучасні підходи до опису фактор-графів не дозволяють включити частотну інформацію у модель графу. Тому збереження інформації про частоту використання тих чи інших вершин потребує удосконалення моделі та реалізується через створення додаткових структур даних. Показником релевантності результатів для кінцевого користувача буде відносна вага результату, яка зберігається у дузі фактор-графу. У даній роботі запропоновано змінити розрахунок ваги системи DeepDive, розраховуючи її не лише із ваги фактор-графу, але також врахувавши фактор частоти використання дуг. Обом вагам призначається коефіцієнт, який можна вирахувати вручну чи змінювати під час роботи програми.

Запропонований комбінований метод підвищує ефективність використання графової бази знань, що побудована на фактор-графах.

Список використаних джерел:

1. Tian, L., Zhou, X., Wu, Y. P., Zhou, W. T., Zhang, J. H., & Zhang, T. S. (2022). Knowledge graph and knowledge reasoning: A systematic review. *Journal of Electronic Science and Technology*, 100159.
2. Kejriwal, M. (2019). *Domain-specific knowledge graph construction*. New York: Springer International Publishing.
3. Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., & Taylor, J. (2019). Industry-scale Knowledge Graphs: Lessons and Challenges: Five diverse technology companies show how it's done. *Queue*, 17(2), 48-75.
4. De Sa, C., Ratner, A., Ré, C., Shin, J., Wang, F., & Wu, S. (2016). DeepDive: A data management system for automatic knowledge base construction. *ACM Transactions on Database Systems (TODS)*.
5. Ratner, A., De Sa, C., Wu, S., Selsam, D., Ré, C., & Duchi, J. (2015). *Incremental Knowledge Base Construction Using DeepDive*. VLDB Endowment.