

# Экскурс в графы знаний

И.А. Волкова, Е.Д. Шамаева

**Аннотация** — В данном обзоре рассматриваются различные аспекты графов знаний, ставшей популярной в последнее время технологии хранения информации. Концепция графа используется для хранения информации о сущностях и отношениях между ними. Графы знаний могут быть полезны как для обогащения программных систем структурированной информацией о мире, так и в качестве справочной информации для человека.

Раскрыто понятие графа знаний, отмечены их разновидности.

Рассмотрены некоторые особенности внутреннего устройства графов знаний, в том числе аспекты, связанные с хранением данных и их получением, а также вывод новых знаний на основе уже имеющихся.

Затронуты вопросы, связанные с построением графов знаний, добавлением в них новой информации (в том числе автоматически) и предоставлением пользователям модифицированной версии графа знаний.

Указаны также сферы применения графов знаний. Особое внимание уделено их применению в различных системах автоматической обработки текстов, рекомендательных системах, в машинном обучении. Отдельно рассматривается процесс создания векторного представления графа знаний. Кроме того, перечислены не решенные на данный момент задачи, связанные с графами знаний.

Данный обзор будет наиболее интересен специалистам, начинающим знакомство с графами знаний.

**Ключевые слова** — граф знаний, понятие, сущность, отношение.

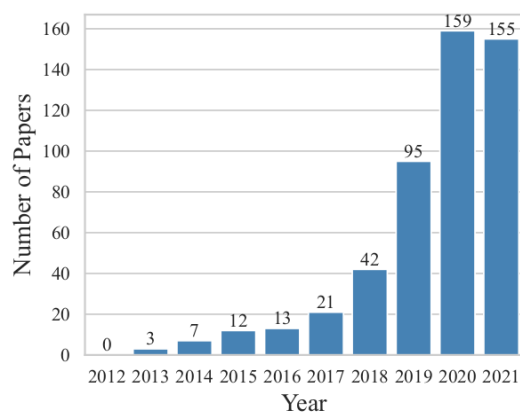
Статья получена 27 декабря 2022

Волкова И.А., Московский государственный университет имени М.В. Ломоносова (e-mail: irina.a.volkova@gmail.com). Шамаева Е.Д., аспирантура, Московский государственный университет имени М.В. Ломоносова (e-mail: derinhelm@yandex.ru).

## I. Введение

На текущий момент достаточно актуальным является создание приложений, управляемых знаниями (*knowledge-driven application*) [1]. Ядром таких приложений являются *графы знаний*, которые используются для хранения фактографических знаний о мире. Согласно [2], термин граф знаний привлек внимание научного сообщества после того, как в 2012 году в блоге Google появилась публикация, в которой было описано использование графа знаний для дополнения результатов поискового запроса. Графы знаний активно используются во многих приложениях искусственного интеллекта, в частности, при создании систем автоматической обработки текстов на естественном языке.

На Рис. 1 представлена диаграмма из [2], отражающая увеличение количества статей, связанных одновременно и с графами знаний, и с обработкой естественного языка.



**Рисунок 1.** Динамика количества статей, связанных с графами знаний и обработкой естественного языка<sup>1</sup>

<sup>1</sup> На Рис.1 не учитываются многие проведенные в 2021, 2022 годах исследования, которые были внесены в базы данных после сбора статистики для этого графика.

В графах знаний часто содержится достаточно большое количество объектов. Например, в соответствии с [3] в Wikidata [4] содержится более 65 миллионов объектов<sup>2</sup>. Информация может храниться и на разных языках. В частности, согласно [5], DBpedia [6] содержит более 6 млрд связанных фактов на всех используемых в Википедии языках.

#### А. Определение графа знаний

В настоящее время существуют разные определения термина граф знаний. Но, как отмечено в [2], на данный момент многие исследователи пользуются следующим определением, введенным в [7], где граф знаний определен как «*граф данных, предназначенный для накопления и передачи знаний о реальном мире, узлы которого представляют интересующие объекты, а ребра представляют отношения между этими объектами*». А, например, в [8] приводится другое определение: «*Граф знаний — это мультиреляционный граф, состоящий из сущностей и отношений, которые представлены узлами и ребрами разных типов соответственно*».

Вслед за [9] в настоящем обзоре система будет позиционироваться как граф знаний, если она рассматривалась в качестве графа знаний в признанной научным сообществом публикации.

#### В. Виды графов знаний

Графы знаний подразделяются на:

- **общие**, которые хранят информацию, общую для разных предметных областей;
- **специализированные**, содержащие информацию из одной предметной области.

Примером общего графа знаний является Wikidata [4]. Пример специализированного графа знаний — BioPortal [10], граф знаний, в котором содержится более 140 миллиардов фактов о биотехнологиях и медицине.

Согласно [9], существуют открытые (общедоступные) и закрытые (коммерческие) графы знаний. Например,

- общедоступными графами знаний являются DBpedia [6], Wikidata [4], YAGO [11], NELL [12],
- коммерческими и закрытыми — Cys [13], Google Knowledge Graph [14], Google Knowledge Vault [15].

## II. Внутреннее устройство графов знаний

В узлах графов знаний находятся понятия и именованные сущности, ребра соответствуют отношениям между этими сущностями. У сущностей могут быть введены некоторые атрибуты. Например, если в виде графа знаний представлена некоторая социальная сеть, то пользователи этой сети являются узлами графа, местоположение пользователя — атрибутом узла, а пользователи связываются некоторыми отношениями «дружбы» [16].

Набор отношений, используемый в графе знаний, зависит от проблемной области. Однако во многих графах знаний используются схожие системы отношений.

Для хранения графов знаний используются различные форматы. Согласно [9], большинство общедоступных графов знаний для хранения данных использует RDF-триплеты<sup>3</sup>, предложенные организацией World Wide Web Consortium (W3C). RDF-триплеты представляют собой тройки (*субъект, отношение, сущность*). Сущности являются узлами графа знаний, а отношения — ребрами.

На Рис. 2 и Рис. 3 показан фрагмент графа знаний<sup>4</sup>. На Рис. 2 фрагмент графа знаний представлен в графическом виде, на Рис. 3 — в виде триплетов (*сущность, отношение, сущность*).

<sup>2</sup> Состояние на август 2019 года.

<sup>3</sup> RDF — Resource Description Framework.

<sup>4</sup> Рис. 2 и Рис. 3 взяты из [7].

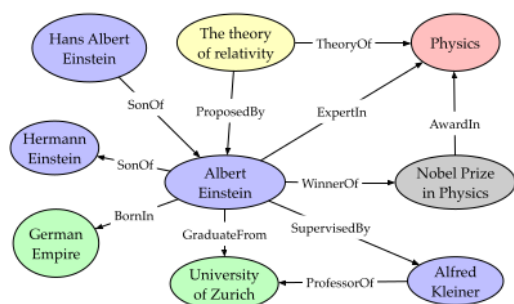


Рисунок 2. Фрагмент графа знаний

- (Albert Einstein, **BornIn**, German Empire)
- (Albert Einstein, **SonOf**, Hermann Einstein)
- (Albert Einstein, **GraduateFrom**, University of Zurich)
- (Albert Einstein, **WinnerOf**, Nobel Prize in Physics)
- (Albert Einstein, **ExpertIn**, Physics)
- (Nobel Prize in Physics, **AwardIn**, Physics)
- (The theory of relativity, **TheoryOf**, Physics)
- (Albert Einstein, **SupervisedBy**, Alfred Kleiner)
- (Alfred Kleiner, **ProfessorOf**, University of Zurich)
- (The theory of relativity, **ProposedBy**, Albert Einstein)
- (Hans Albert Einstein, **SonOf**, Albert Einstein)

Рисунок 3. Граф знаний в виде триплетов

С помощью декларативного языка запросов SPARQL [17] можно получать данные, хранящихся в RDF-формате. Так, согласно [3], DBpedia и Wikidata для доступа к своим графам знаний предоставляют данные в формате RDF и конечные точки (*endpoint*) SPARQL<sup>5</sup>.

Согласно [18], графы знаний (например, DBpedia, Wikidata) содержат таксономическую основу, которая позволяет упорядочивать и структурировать концепции в соответствии с гипо-гиперонимными отношениями<sup>6</sup>.

Многие графы знаний содержат ссылки на сущности из внешних открытых источников данных и других графов знаний. Например, согласно [19], в графе знаний DBpedia содержатся ссылки на сущности из 179 внешних источников, в том числе из таких графов знаний, как Wikidata, YAGO, W3 и других. А, согласно [3], Google's Knowledge Graph интегрирует в свой граф знаний

данные из таких источников, как Wikipedia, World Bank, Eurostat.

### А. Векторное представление графов знаний

В задачах машинного обучения используются векторные представления графов знаний (*embedding*).

Согласно [8], создание векторного представления графов знаний, при котором используются только содержащиеся в графах знаний факты, часто происходит в три этапа:

1. Сущности и отношения конвертируются в элементы непрерывного векторного пространства. Сущности обычно представляются в качестве векторов, а отношения — в виде операций в векторном пространстве.
2. Для каждого факта  $(h, r, t)$ <sup>7</sup> определяется оценочная функция (*scoring function*)  $f_r(h, t)$ , показывающая правдоподобие (*plausibility*) этого факта. Факты, которые содержатся в графе знаний, имеют более высокую оценку, чем факты, которые отсутствуют в графе знаний.
3. Для получения векторных представлений этих сущностей и отношений решается оптимизационная задача максимизации общего правдоподобия фактов, наблюдаемых в графе знаний.

Для создания векторного представления могут также использоваться типы сущностей, пути отношений (*relation path*), текстовые описания и логические правила, содержащиеся в графе знаний.

### В. Механизм вывода

Помимо информации граф знаний содержит вспомогательные инструменты. Как отмечено в [20], граф знаний, в частности, должен сопровождаться механизмом вывода (*reasoning engine*). Согласно [21], в графах знаний *вывод знаний* — это получение на основе графа знаний триплета, ранее не принадлежащего графу

<sup>5</sup> Это соответствует стандартам Semantic Web.

<sup>6</sup> Отношениями «класс-подкласс».

<sup>7</sup>  $h$  и  $t$  — сущности графа знаний, а  $r$  — отношение между ними.

знаний. Например, если в графе знаний есть факты (`Microsoft`, `IsBasedIn`, `Seattle`), (`Seattle`, `StateLocatedIn`, `Washington`) и (`Washington`, `CountryLocatedIn`, `USA`), можно получить пропущенную связь (`Microsoft`, `HeadquarterLocatedIn`, `USA`).

Согласно [21], механизм вывода в графах знаний используется для выявления ошибок и получения новых знаний из существующих в графе знаний данных.

Там же выделяются три типа механизма вывода, использующегося в графах знаний:

1. на основе логических правил,
2. на основе распределенных представлений (*distributed representation*),
3. на основе нейронных сетей.

### III. Построение и развитие графов знаний

Согласно [9], графы знаний могут быть построены либо нисходящим способом (*top-down*), либо восходящим (*bottom-up*). В первом случае сначала моделируется структура графа знаний, а затем граф знаний заполняется данными. Во втором случае структура графа знаний выводится на основе данных.

Для построения графа знаний может использоваться:

- информация, собранная вручную, с привлечением экспертов;
- информация, автоматически выделенная из текста;
- информация, полученная из внешних источников (например, из других графов знаний).

Может быть применена и комбинация этих подходов.

Например, согласно [3], для построения графов знаний DBpedia и YAGO извлекалась информация из Википедии, а для графа знаний NELL с помощью методов машинного обучения были получены данные из 500 миллионов веб-страниц.

После создания необходимо поддерживать актуальность графа знаний,

пополняя и модифицируя содержащиеся в нем данные.

В [2] отмечено, что для автоматического построения и развития графов знаний часто используются методы, разработанные для таких задач, как извлечение сущностей (*entity extraction*), извлечение связей (*relation extraction*), извлечение атрибутов (*attribute extraction*). Важное значение имеет также задача связывания извлеченных сущностей с сущностями базы знаний (*entity linking*) и извлеченных отношений с отношениями базы знаний (*relation linking*).

Новые данные могут быть также внесены вручную.

Помимо этого, для повышения качества графа знаний необходимо производить объединение повторяющихся сущностей, которые могли появиться, например, при автоматическом добавлении информации из различных текстовых источников или других графов знаний.

После обновления графа знаний необходимо предоставить пользователям возможность получить доступ к новой версии. Например, согласно [3], DBpedia периодически осуществляет публикацию новой версии, синхронизированной с Википедией. Кроме того, последнюю версию DBpedia можно получить также с помощью отправки запросов к конечной точке (*endpoint*) с помощью языка SPARQL.

### IV. Сферы применения

#### A. Системы автоматической обработки текстов

Одно из основных направлений в области искусственного интеллекта — создание систем автоматической обработки текстов. Графы знаний в первую очередь полезны в тех системах обработки естественного языка, где требуется выход на семантический уровень его обработки, требуется то или иное «понимание» текста.

На Рис. 4 показано<sup>8</sup>, в каких предметных областях, согласно [2]<sup>9</sup>, для решения тех или

<sup>8</sup> Рис. 4 взят из [1].

иных задач использовались и графы знаний, и системы автоматической обработки текстов. Наибольшее количество публикаций было отмечено в здравоохранении.

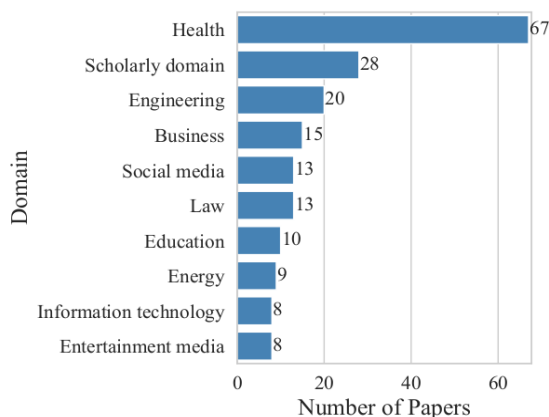


Рисунок 4. Графы знаний и системы автоматической обработки естественного языка в различных предметных областях

естественном языке. Согласно [25], графы знаний можно использовать и для получения (вывода) ответа из фактов, хранящихся в графе знаний, и в качестве источника знаний о мире для генерации ответов на основе текстов (*textual QA*)<sup>10</sup>. Например, в [3] описано применение диалоговых систем, использующих графы знаний, в сфере туризма.

Кроме того, в соответствии с [2], графы знаний активно используются для информационного поиска, учитывающего цель поиска (*search intent*) или контекст запроса<sup>11</sup>. Они также успешно применяются для дополнения поисковой выдачи. На Рис. 5 показан<sup>12</sup> пример поисковой выдачи, дополненной информацией, которая была получена с помощью графа знаний<sup>13</sup> (подробнее см. в [14]).

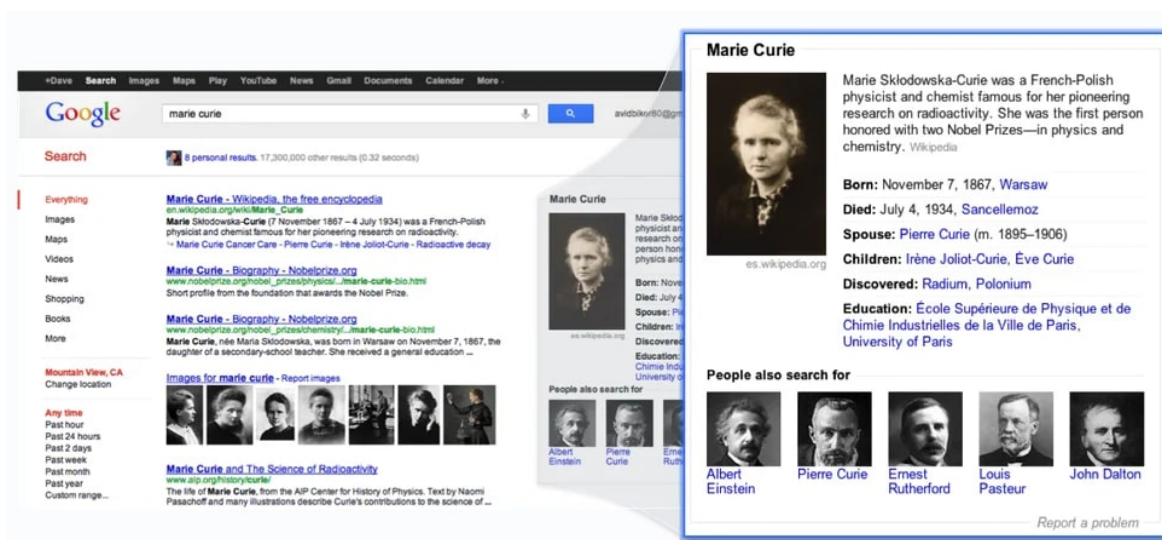


Рисунок 5. Применение графа знаний для расширения поисковой выдачи

Согласно [2], графы знаний активно используются для определения тональности (*sentiment detection*) [22], моделирования темы (*topic modeling*) [23], снятия омонимии слов [24] и во многих других задачах анализа текстов.

Графы знаний также используются в диалоговых системах общения на

## В. Рекомендательные системы

Графы знаний успешно используются и в рекомендательных системах. Например,

- в фармакологической рекомендательной системе лекарственных средств [26] —

<sup>10</sup> В таком случае, согласно [36], способ получения ответов становится более понятным для пользователя.

<sup>11</sup> Применяются для поиска, подбора рекомендаций и аналитики.

<sup>12</sup> Рис. 5 взят из [14].

<sup>13</sup> Эта информация сгруппирована в карточке, находящейся в правой части страницы.

<sup>9</sup> В [1] были специальным образом отобраны статьи, опубликованные за последние 10 лет.

для хранения в графе знаний GNBR<sup>14</sup> объектов, связанных с лекарствами, болезнями и генами;

- в образовательной рекомендательной системе онлайн-курсов [27] — для хранения информации о курсах и выявления взаимосвязи между знаниями разных курсов.
- в рекомендательной системе [28] — для агрегирования междисциплинарных знаний относительно улучшения психического здоровья в графе эмоциональных знаний (EmoKG).

Использование графов знаний в рекомендательных системах повышает интерпретируемость результата.

### C. Графы знаний и машинное обучение

Согласно [29], графы знаний могут быть использованы для увеличения объема обучающих выборок, состоящих из текста.

В [30] отмечается, что графы знаний могут быть также использованы для интерпретации результатов, полученных с помощью машинного обучения.

Информация из графов знаний может быть также внедрена в языковые модели (например, BERT, GPT). Таким образом, структурированные данные из графов знаний объединяются с неструктурированными данными, на основе которых происходило обучение языковых моделей. Полученные модели называются дополненными языковыми моделями (*augmented language model*). Примерами таких моделей являются ERNIE [31], COMET [32], K-BERT [33] и KEPLER [34].

### D. Другие применения графов знаний

В [9] и [35] отмечено использование графов знаний для анализа данных с производственных линий, генерации кода, генерации тестовых примеров на основе спецификаций, создания цифровых двойников, автоматизированного моделирования (документирования) киберфизических систем и имитации

процессов, а также для аналитики и визуализации данных<sup>15</sup> и других задач.

## V. Заключение

Несмотря на относительную новизну, графы знаний заняли значимое место не только в научных исследованиях, но и в реальных промышленных проектах. Однако, согласно [9], многие первичные исследования (*primary study*) еще находятся на ранней стадии разработки и не были оценены в промышленных условиях, поэтому в этой сфере остается еще множество нерешенных задач. Например,

- автоматическое пополнение имеющихся графов знаний новыми понятиями и именованными сущностями с контролем непротиворечивости и логической совместимости новых знаний с уже имеющимися;
- адаптация знаний, хранящихся в графах знаний, к изменяющимся внешним реалиям;
- автоматическое выделение и включение в графы знаний новых отношений, как общих, так и для конкретных предметных областей.

На сегодняшний день графы знаний — развивающийся способ представления и хранения фактографической информации о предметных областях, а также способ построения семантических моделей той или иной сферы деятельности человека.

## VI. Библиография

- [1] T. Hubauer, S. Lamparter, P. Haase, and D. M. Herzig, “Use Cases of the Industrial Knowledge Graph at Siemens,” 2018.
- [2] P. Schneider, T. Schopf, J. Vladika, M. Galkin, E. Simperl, and F. Matthes, “A Decade of Knowledge Graphs in Natural Language Processing: A Survey,” *arXiv preprint arXiv:2210.00105*, 2022.
- [3] D. Fensel *et al.*, *Knowledge graphs*. Springer, 2020.
- [4] D. Vrandečić and M. Krötzsch, “Wikidata: A Free Collaborative Knowledgebase,” *Commun. ACM*, vol. 57, no. 10, pp. 78–85, Sep. 2014, doi: 10.1145/2629489.
- [5] Д. Муромцев, Д. Волчек, and А. Романов, “Индустриальные

<sup>14</sup> Global Network of Biomedical Relationships

<sup>15</sup> Например, в области целевой рекламы, бизнес-аналитики, оценки рисков, контроля за предприятием, управления логистикой.

- графы знаний-интеллектуальное ядро цифровой экономики,” *Control Engineering Россия*, no. 5, pp. 32–39, 2019.
- [6] J. Lehmann *et al.*, “DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia,” *Semant Web*, vol. 6, no. 2, pp. 167–195, 2015, doi: 10.3233/SW-140134.
- [7] A. Hogan *et al.*, “Knowledge graphs,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–37, 2021.
- [8] Q. Wang, Z. Mao, B. Wang, and L. Guo, “Knowledge graph embedding: A survey of approaches and applications,” *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 12, pp. 2724–2743, Dec. 2017, doi: 10.1109/TKDE.2017.2754499.
- [9] G. Buchgeher, D. Gabauer, J. Martinez-Gil, and L. Ehrlinger, “Knowledge graphs in manufacturing and production: A systematic literature review,” *IEEE Access*, vol. 9, pp. 55537–55554, 2021.
- [10] “BioPortal,” *BioPortal*. <http://bioportal.bioontology.org> (Дата доступа: 20 декабря, 2022).
- [11] F. M. Suchanek, G. Kasneci, and G. Weikum, “Yago: A core of semantic knowledge,” in *Proceedings of the 16th international conference on World Wide Web - WWW '07*, New York, New York, USA, May 2007, p. 697, doi: 10.1145/1242572.1242667.
- [12] A. Carlson, J. Betteridge, R. C. Wang, E. R. Hruschka, and T. M. Mitchell, “Coupled semi-supervised learning for information extraction,” in *Proceedings of the third ACM international conference on Web search and data mining - WSDM '10*, New York, New York, USA, Feb. 2010, p. 101, doi: 10.1145/1718487.1718501.
- [13] D. B. Lenat, “CYC: a large-scale investment in knowledge infrastructure,” *Commun. ACM*, vol. 38, no. 11, pp. 33–38, Nov. 1995, doi: 10.1145/219717.219745.
- [14] G. O. Blog, “Introducing the knowledge graph: thing, not strings,” *Introducing the Knowledge Graph: things, not strings*, 2012.
- [15] X. Dong *et al.*, “Knowledge vault: A web-scale approach to probabilistic knowledge fusion,” in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*, New York, New York, USA, Aug. 2014, pp. 601–610, doi: 10.1145/2623330.2623623.
- [16] Ш. Бхатт, Ц. Чжао, А. Сетх, and В. Шалин, “Графы знаний как средство улучшения искусственного интеллекта,” *Открытые системы. СУБД*, no. 3, pp. 24–26, 2020.
- [17] S. Harris, A. Seaborne, and E. Prud’hommeaux, “SPARQL 1.1 query language,” *W3C recommendation*, vol. 21, no. 10, p. 778, 2013.
- [18] I. Nikishina, M. Tikhomirov, V. Logacheva, Y. Nazarov, A. Panchenko, and N. Loukachevitch, “Taxonomy enrichment with text and graph vector representations,” *Semantic Web*, no. Preprint, pp. 1–35, 2022.
- [19] “DBpedia Snapshot 2022-09 Release,” *DBpedia*. <https://www.dbpedia.org/blog/dbpedia-snapshot-2022-09-release/> (Дата доступа: 20 декабря, 2022).
- [20] L. Ehrlinger and W. Wöb, “Towards a definition of knowledge graphs,” *SEMANTiCS (Posters, Demos, SuCCESS)*, vol. 48, no. 1–4, p. 2, 2016.
- [21] X. Chen, S. Jia, and Y. Xiang, “A review: Knowledge reasoning over knowledge graph,” *Expert Syst. Appl.*, vol. 141, p. 112948, Mar. 2020, doi: 10.1016/j.eswa.2019.112948.
- [22] A. Kumar, D. Kawahara, and S. Kurohashi, “Knowledge-enriched two-layered attention network for sentiment analysis,” *arXiv preprint arXiv:1805.07819*, 2018.
- [23] D. Li, S. Zamani, J. Zhang, and P. Li, “Integration of knowledge graph embedding into topic modeling with hierarchical dirichlet process,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 940–950.
- [24] S. Kumar, S. Jat, K. Saxena, and P. Talukdar, “Zero-shot Word Sense Disambiguation using Sense Definition Embeddings,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA, 2019, pp. 5670–5681, doi: 10.18653/v1/P19-1568.
- [25] B. Fu, Y. Qiu, C. Tang, Y. Li, H. Yu, and J. Sun, “A survey on complex question answering over knowledge base: Recent advances and challenges,” *arXiv preprint arXiv:2007.13069*, 2020.
- [26] D. N. Sosa, A. Derry, M. Guo, E. Wei, C. Brinton, and R. B. Altman, “A Literature-Based Knowledge Graph Embedding Method for Identifying Drug Repurposing Opportunities in Rare Diseases,” *BioRxiv*, Aug. 2019, doi: 10.1101/727925.
- [27] J. Zhou, G. Jiang, W. Du, and C. Han, “Profiling temporal learning interests with time-aware transformers and knowledge graph for online course recommendation,” *Electronic Commerce Research*, pp. 1–21, 2022.
- [28] A. Gyraud and K. Boudaoud, “Interdisciplinary IoT and Emotion Knowledge Graph-Based Recommendation System to Boost Mental Health,” *Applied Sciences*, vol. 12, no. 19, p. 9712, 2022.
- [29] S. Sharifirad, B. Jafarpour, and S. Matwin, “Boosting text classification performance on sexist tweets by text augmentation and text generation using a combination of knowledge graphs,” in *Proceedings of the 2nd workshop on abusive language online (ALW2)*, 2018, pp. 107–114.
- [30] I. Tiddi and S. Schlobach, “Knowledge graphs as tools for explainable machine learning: A survey,” *Artif. Intell.*, vol. 302, p. 103627, Jan. 2022, doi: 10.1016/j.artint.2021.103627.
- [31] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu, “ERNIE: Enhanced language representation with informative entities,” *arXiv preprint arXiv:1905.07129*, 2019.
- [32] A. Bosselut, H. Rashkin, M. Sap, C. Malaviya, A. Celikyilmaz, and Y. Choi, “COMET: Commonsense transformers for automatic knowledge graph construction,” *arXiv preprint arXiv:1906.05317*, 2019.
- [33] W. Liu *et al.*, “K-BERT: Enabling Language Representation with Knowledge Graph,” *AAAI*, vol. 34, no. 03, pp. 2901–2908, Apr. 2020, doi: 10.1609/aaai.v34i03.5681.
- [34] X. Wang *et al.*, “KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation,” *Trans. Assoc. Comput. Linguist.*, vol. 9, pp. 176–194, Mar. 2021, doi: 10.1162/tacl\_a\_00360.
- [35] C. Billy, W. K. Eric, and W. T. Eric, “Knowledge Graph Construction and Applications in E-Retailing: A Review of Literature,” 2021.
- [36] S. Zhu, X. Cheng, and S. Su, “Knowledge-based question answering by tree-to-sequence learning,” *Neurocomputing*, vol. 372, pp. 64–72, 2020.

# Excursus in knowledge graphs

I.A. Volkova, E.D. Shamaeva

**Abstract** — This review examines various aspects of knowledge graphs, which have recently become a popular information storage technology. The graph concept is used to store information about entities and the relationships between them. Knowledge graphs can be useful both for enriching software systems with structured information about the world, and as reference information for people.

The knowledge graph concept is revealed, their varieties are noted.

Some features of the knowledge graph internal structure are considered, including aspects related to data storage and acquisition, as well as the new knowledge reasoning based on existing one.

The issues related to the knowledge graph construction, adding new information to them (including automatic one) and providing users with a modified knowledge graph version are touched upon.

The knowledge graph application fields are also indicated. Special attention is paid to their application in various automatic text processing systems, recommendation systems, and machine learning. The process of creating a knowledge graph vector representation is considered separately. In addition, the tasks related to knowledge graphs that have not been solved at the moment are listed.

This review will be most interesting for specialists, beginning to get acquainted with knowledge graphs.

**Keywords** — knowledge graph, concept, entity, relation.

## REFERENCES

- [1] T. Hubauer, S. Lamparter, P. Haase, and D. M. Herzig, “Use Cases of the Industrial Knowledge Graph at Siemens.,” 2018.
- [2] P. Schneider, T. Schopf, J. Vladika, M. Galkin, E. Simperl, and F. Matthes, “A Decade of Knowledge Graphs in Natural Language Processing: A Survey,” *arXiv preprint arXiv:2210.00105*, 2022.
- [3] D. Fensel *et al.*, *Knowledge graphs*. Springer, 2020.
- [4] D. Vrandečić and M. Krötzsch, “Wikidata: A Free Collaborative Knowledgebase,” *Commun. ACM*, vol. 57, no. 10, pp. 78–85, Sep. 2014, doi: 10.1145/2629489.
- [5] Д. Муромцев, Д. Волчек, and А. Романов, “Индустриальные графы знаний-интеллектуальное ядро цифровой экономики,” *Control Engineering Россия*, no. 5, pp. 32–39, 2019.
- [6] J. Lehmann *et al.*, “DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia,” *Semant Web*, vol. 6, no. 2, pp. 167–195, 2015, doi: 10.3233/SW-140134.
- [7] A. Hogan *et al.*, “Knowledge graphs,” *ACM Computing Surveys* (*CSUR*), vol. 54, no. 4, pp. 1–37, 2021.
- [8] Q. Wang, Z. Mao, B. Wang, and L. Guo, “Knowledge graph embedding: A survey of approaches and applications,” *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 12, pp. 2724–2743, Dec. 2017, doi: 10.1109/TKDE.2017.2754499.
- [9] G. Buchgeher, D. Gabauer, J. Martinez-Gil, and L. Ehrlinger, “Knowledge graphs in manufacturing and production: A systematic literature review,” *IEEE Access*, vol. 9, pp. 55537–55554, 2021.
- [10] “BioPortal,” *BioPortal*. <http://biportal.bioontology.org> (accessed Dec. 20, 2022).
- [11] F. M. Suchanek, G. Kasneci, and G. Weikum, “Yago: A core of semantic knowledge,” in *Proceedings of the 16th international conference on World Wide Web - WWW '07*, New York, New York, USA, May 2007, p. 697, doi: 10.1145/1242572.1242667.
- [12] A. Carlson, J. Betteridge, R. C. Wang, E. R. Hruschka, and T. M. Mitchell, “Coupled semi-supervised learning for information extraction,” in *Proceedings of the third ACM international conference on Web search and data mining - WSDM '10*, New York, New York, USA, Feb. 2010, p. 101, doi: 10.1145/1718487.1718501.
- [13] D. B. Lenat, “CYC: a large-scale investment in knowledge infrastructure,” *Commun. ACM*, vol. 38, no. 11, pp. 33–38, Nov. 1995, doi: 10.1145/219717.219745.
- [14] G. O. Blog, “Introducing the knowledge graph: thing, not strings,” *Introducing the Knowledge Graph: things, not strings*, 2012.
- [15] X. Dong *et al.*, “Knowledge vault: A web-scale approach to probabilistic knowledge fusion,” in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*, New York, New York, USA, Aug. 2014, pp. 601–610, doi: 10.1145/2623330.2623623.
- [16] SH. Bhatt, C. CHzhao, A. Seth, and V. SHalin, “Графы знаний как средство улучшения искусственного интеллекта,” *Otkrytye sistemy. SUBD*, no. 3, pp. 24–26, 2020.
- [17] S. Harris, A. Seaborne, and E. Prud'hommeaux, “SPARQL 1.1 query language,” *W3C recommendation*, vol. 21, no. 10, p. 778, 2013.
- [18] I. Nikishina, M. Tikhomirov, V. Logacheva, Y. Nazarov, A. Panchenko, and N. Loukachevitch, “Taxonomy enrichment with text and graph vector representations,” *Semantic Web*, no. Preprint, pp. 1–35, 2022.
- [19] “DBpedia Snapshot 2022-09 Release,” *DBpedia*. <https://www.dbpedia.org/blog/dbpedia-snapshot-2022-09-release/> (accessed Dec. 19, 2022).
- [20] L. Ehrlinger and W. Wöß, “Towards a definition of knowledge graphs.,” *SEMANTiCS (Posters, Demos, SuCESS)*, vol. 48, no. 1–4, p. 2, 2016.
- [21] X. Chen, S. Jia, and Y. Xiang, “A review: Knowledge reasoning over knowledge graph,” *Expert Syst. Appl.*, vol. 141, p. 112948, Mar. 2020, doi: 10.1016/j.eswa.2019.112948.
- [22] A. Kumar, D. Kawahara, and S. Kurohashi, “Knowledge-enriched two-layered attention network for sentiment analysis,” *arXiv preprint arXiv:1805.07819*, 2018.
- [23] D. Li, S. Zamani, J. Zhang, and P. Li, “Integration of knowledge graph embedding into topic modeling with hierarchical dirichlet



- process,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 940–950.
- [24] S. Kumar, S. Jat, K. Saxena, and P. Talukdar, “Zero-shot Word Sense Disambiguation using Sense Definition Embeddings,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA, 2019, pp. 5670–5681, doi: 10.18653/v1/P19-1568.
- [25] B. Fu, Y. Qiu, C. Tang, Y. Li, H. Yu, and J. Sun, “A survey on complex question answering over knowledge base: Recent advances and challenges,” *arXiv preprint arXiv:2007.13069*, 2020.
- [26] D. N. Sosa, A. Derry, M. Guo, E. Wei, C. Brinton, and R. B. Altman, “A Literature-Based Knowledge Graph Embedding Method for Identifying Drug Repurposing Opportunities in Rare Diseases,” *BioRxiv*, Aug. 2019, doi: 10.1101/727925.
- [27] J. Zhou, G. Jiang, W. Du, and C. Han, “Profiling temporal learning interests with time-aware transformers and knowledge graph for online course recommendation,” *Electronic Commerce Research*, pp. 1–21, 2022.
- [28] A. Gyrard and K. Boudaoud, “Interdisciplinary IoT and Emotion Knowledge Graph-Based Recommendation System to Boost Mental Health,” *Applied Sciences*, vol. 12, no. 19, p. 9712, 2022.
- [29] S. Sharifirad, B. Jafarpour, and S. Matwin, “Boosting text classification performance on sexist tweets by text augmentation and text generation using a combination of knowledge graphs,” in *Proceedings of the 2nd workshop on abusive language online (ALW2)*, 2018, pp. 107–114.
- [30] I. Tiddi and S. Schlobach, “Knowledge graphs as tools for explainable machine learning: A survey,” *Artif. Intell.*, vol. 302, p. 103627, Jan. 2022, doi: 10.1016/j.artint.2021.103627.
- [31] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu, “ERNIE: Enhanced language representation with informative entities,” *arXiv preprint arXiv:1905.07129*, 2019.
- [32] A. Bosselut, H. Rashkin, M. Sap, C. Malaviya, A. Celikyilmaz, and Y. Choi, “COMET: Commonsense transformers for automatic knowledge graph construction,” *arXiv preprint arXiv:1906.05317*, 2019.
- [33] W. Liu *et al.*, “K-BERT: Enabling Language Representation with Knowledge Graph,” *AAAI*, vol. 34, no. 03, pp. 2901–2908, Apr. 2020, doi: 10.1609/aaai.v34i03.5681.
- [34] X. Wang *et al.*, “KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation,” *Trans. Assoc. Comput. Linguist.*, vol. 9, pp. 176–194, Mar. 2021, doi: 10.1162/tacl\_a\_00360.
- [35] C. Billy, W. K. Eric, and W. T. Eric, “Knowledge Graph Construction and Applications in E-Retailing: A Review of Literature,” 2021.
- [36] S. Zhu, X. Cheng, and S. Su, “Knowledge-based question answering by tree-to-sequence learning,” *Neurocomputing*, vol. 372, pp.64-72, 2020.