

Онтологический Саммит 2020 Коммюнике: Графы знаний Ontology Summit 2020: Knowledge Graphs¹

Декабрь 2020

Кен Баклавски, Северо-Восточный университет, Бостон, Массачусетс, США

Майкл Беннет, Hupercube Limited, Лондон, Великобритания

Гэри Берг-Кросс, руководитель ESIP по семантической гармонизации

Тодд Шнайдер, Engineering Semantics, Fairfax, Вирджиния, США

Рави Шарма, старший архитектор предприятия, Elk Grove, Калифорния, США

Джанет Сингер, INCOSE, Скоттс-Вэлли, Калифорния, США

Аннотация:

В настоящее время все больше данных доступно из государственных и частных источников. Кроме того, типы, форматы и количество источников данных также увеличиваются. В последние несколько лет были разработаны методы извлечения, хранения, обработки и анализа таких данных для управления этим удивительным разнообразием на основе структуры, называемой графом знаний. Компании приложили много усилий для разработки графов знаний, и теперь они имеют решающее значение для функций интеллектуальных виртуальных помощников, таких как *Siri*, *Alexa* и *Google Assistant*. Целью Онтологического саммита 2020 было понять не только, что такое графы знаний, но и то, откуда они возникли, почему они так популярны, текущие проблемы и их перспективы на будущее. На сессиях саммита было рассмотрено множество примеров графов знаний, а также соответствующие существующие и разрабатываемые стандарты для них. Цель этого Коммюнике - обобщить наше понимание, чтобы способствовать исследованиям и развитию графов знаний.

Ключевые слова: граф знаний, архитектура графа знаний, онтология, семантика, семантические сети.

Цитирование: Баклавски, К. Онтологический Саммит 2020. Коммюнике: Графы знаний / К. Баклавски, М. Беннет, Г. Берг-Кросс, Т. Шнайдер, Р. Шарма, Д. Сингер. Перевод с англ. Д. Боргест // Онтология проектирования. – 2020. – Т.10, №4(38). – С.540-555. – DOI: 10.18287/2223-9537-2020-10-4-540-555.

1 Введение

Существует долгая история использования графов знаний (ГЗ) в различных областях, однако за последние несколько лет они оказались особенно важным инструментом для семантических технологий и областей исследований. Как структурированные представления семантических знаний, которые хранятся в графе, ГЗ представляют собой облегченные версии семантических сетей, которые потенциально могут масштабироваться до массивных репозиториях данных, таких как вся Всемирная паутина ["Semantic Network", 2020]. Промышленность приложила много усилий для разработки ГЗ, и теперь они имеют решающее значение для функций интеллектуальных виртуальных помощников, таких как *Siri*, *Alexa* и *Google Assistant*. Исследовательские направления, в которых актуальны ГЗ, включают прикладную онтологию, большие данные, связанные данные, открытую сеть знаний, искусственный интеллект (ИИ) и глубокое обучение.

¹ Перевод финальной версии Коммюнике Онтологического Саммита за 2020 год, опубликованного в декабре 2020 года. - <https://ontologforum.s3.amazonaws.com/OntologySummit2020/Communique/OntologySummit2020Communique.pdf>.

Предварительная краткая информация о Саммите представлена в 2-ом номере 10-го тома журнала. - [https://www.ontology-of-designing.ru/article/2020_2\(36\)/9_Ontology_Summit_2020_Knowledge_Graphs.pdf](https://www.ontology-of-designing.ru/article/2020_2(36)/9_Ontology_Summit_2020_Knowledge_Graphs.pdf).

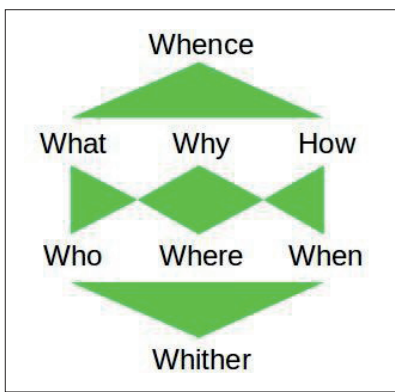


Рисунок 1 - Контекстные вопросы

В ходе виртуальных сессий, проведённых с сентября 2019 по июнь 2020 года ГЗ были рассмотрены с нескольких точек зрения. В коммюнике обобщаются и резюмируются результаты этих сессий. Онтологический саммит 2020 базировался на вопросах, ответы на которые считаются основными для сбора информации, решения проблем или установления контекста. Эти вопросы включают традиционные пять W, к которым добавили «Как», «Откуда» и «Куда», как показано на рисунке 1. Соответственно, это коммюнике построено на основе этих вопросительных слов.

Чтобы придерживаться постоянства в терминологии понятия ГЗ, раздел 2 начинается с предложения практического ответа на вопрос «Что?» для определения ГЗ, основанного на определениях, опубликованных в литературе, а также на выступлениях докладчиков и других обсуждениях во время саммита. В разделе 3 даны некоторые предложения относительно того, «Почему» ГЗ в последнее время стали популярными, а также «Откуда» возникли ГЗ. Раздел 4 посвящен вопросам «Как?», «Кто?», «Где?» и «Когда?» путём изучения примеров методов и инструментов, используемых во многих сферах деятельности систем ГЗ. Разделы 5, 6 и 7 посвящены вопросу «Куда?». В разделе 5 перечислены стандарты и усилия по стандартизации, относящиеся к ГЗ, а в разделе 6 перечислены некоторые проблемы ГЗ, которые были выявлены во время саммита. В разделе 7 обсуждаются перспективы ГЗ. Коммюнике завершается заключением и благодарностью.

Онтологический саммит 2020 охватил очень много вопросов. В этом коммюнике изложены только основные моменты 32 сессий, проведённых за 9 месяцев. Много из того, что обсуждалось, не рассматривается в этой статье. Планируется более полно и подробно представить результаты саммита в серии статей, которые будут опубликованы отдельно.

2 Что такое граф знаний?

Начнём с рассмотрения вопроса о том, что такое ГЗ. К сожалению, существует огромное количество научных работ, а также веб-сайтов и компаний, предлагающих множество различных определений. Чтобы сформулировать чёткое определение, которое поможет в обсуждении понятия ГЗ, были рассмотрены определения из следующей литературы [Krötzsch and Thost, 2016; Paulheim, 2017; Blumauer, 2014; Färber, Ell, Menne, Rettinger and Bartscherer, 2018; Pujara, Miao, Getoor, and Cohen, 2013; Rohrseitz, 2019; Aijal, 2019; Bergman, 2019; Aasman, 2019]. У них были следующие общие моменты:

1. ГЗ представляет собой взаимоотношения. Все определения указывают эту особенность, но делают это по-разному.
2. ГЗ использует методы для извлечения знаний из одного или нескольких источников. Типы источников различаются в этих определениях.
3. Организация - это граф, хотя точное значение слова «граф» варьируется от одного определения к другому.
4. У ГЗ должна быть схема, но не во всех определениях это упоминается. В тех, в которых говорится об этом, указывается, что схема определяет классы и отношения.
5. ГЗ поддерживает различные графические вычисления, интерфейсы поиска и запросов. Поддерживаемые операции и производительность будут различаться, и производительность будет зависеть от того, как решаются компромиссы между масштабируемостью, производительностью и ремонтопригодностью, а также от других технических проблем.

Из приведённых характеристик очевидно, что ГЗ - это не просто ещё один способ представления фактов. Он затрагивает программную архитектуру, которая включает в себя активные возможности для извлечения и обработки фактов. Янс Аасман [Aasman, 2019] охарактеризовал действия ГЗ следующим образом:

- генерация:
 - сбор: загрузка, извлечение из Интернета, извлечение каталогов, онтология, ...
 - обработка: отображение схемы, разрешение сущностей, очистка, ...
- хранение;
- приложения: запросы, анализ графиков, рекомендации, поиск, ответы на вопросы, ...
- статистические методы и методы машинного обучения используются для всего вышеперечисленного.

Другой пример определения дал Никола Рорзейтц: «ГЗ - это набор точек данных, связанных отношениями, которые описывают предметную область, например, бизнес, организацию или область исследования. ... ГЗ - это вторичные или производные наборы данных: они получают путём анализа и фильтрации исходных данных. ... ГЗ также иногда называют семантическими сетями. Семантика подчеркивает тот факт, что значение кодируется вместе с соответствующими данными. Это осуществляется с помощью таксономий и онтологий ...» [Rohrseitz, 2019].

Рорзейтц продолжил описание того, как строятся и используются ГЗ. Это описание аналогично ГЗ, охарактеризованному Янсом Аасманом выше.

Принципы ГЗ, которые являются его основной отличительной чертой, могут быть причиной путаницы в отношении того, что же такое ГЗ. Это также побудило многих охарактеризовать ГЗ как «ничего нового» и как просто ещё одно модное слово. К сожалению, отчасти в этом виноват термин «граф знаний». Он имеет тенденцию предполагать, что ГЗ - это не более чем особый вид графа или сеть. Соответственно, было бы лучше использовать менее запутанный термин, такой как «архитектура ГЗ» (АГЗ), определение которого приводится ниже. Рисунок 2 - один из примеров АГЗ.

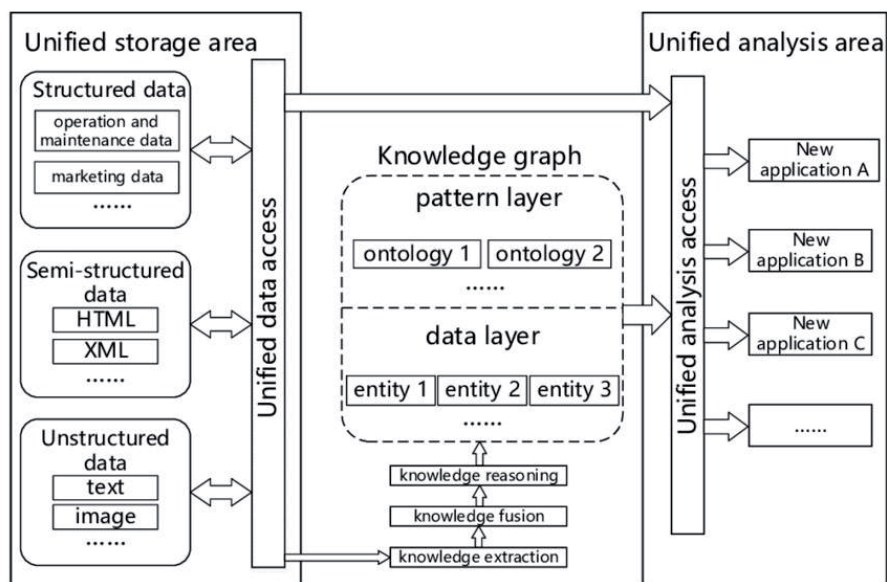


Рисунок 2 - Архитектура графа знаний из [Yuan, Zhang, Dai, Peng and Zhao, 2018]

Архитектура системы ГЗ (СГЗ) аналогична архитектуре хранилища данных (ХД). Технологии ХД использовались для интеграции и согласования данных, чтобы аналитики и пользователи могли надёжно извлекать смысл из своих больших корпоративных наборов данных.

Но, будучи хорошо зарекомендовавшим себя, подход ХД сопряжён со значительными первоначальными и текущими затратами, а также с серьезными рисками. Кроме того, из-за сложности данных ХД не обрабатывает важные области корпоративных данных. Однако ГЗ также может иметь значительные затраты и сложность. ХД используют старые технологии, которым не хватает той гибкости, что есть у ГЗ, что делает их слишком медленными для удовлетворения постоянно меняющихся требований больших данных. СГЗ предлагают более современный, гибкий и динамичный подход к обмену данными и интеграции; и, как представлено в разделе 4, в СГЗ используется множество различных методов и технологий.

Эти заключения приводят к следующему предложению по определению ГЗ и АГЗ:

- ГЗ - это набор утверждений в форме ориентированного мультиграфа с метками узлов и рёбер, позволяющего использовать несколько разнородных рёбер для одних и тех же узлов. Набор определений, определяющих значение меток ГЗ, называется его схемой.
- АГЗ обеспечивает комбинацию масштабируемых технологий, спецификаций и культур данных для представления плотно взаимосвязанных утверждений, разумным образом полученных из структурированных или неструктурированных источников в разных доменах, который читается как человеком, так и машиной.
- АГЗ вместе с набором ГЗ является СГЗ.

Ориентированный мультиграф с метками узлов и ребер - это набор из 8 элементов ($V, E, s, t, \Sigma_V, \Sigma_E, \ell_V, \ell_E$), такие что:

- 1) V - набор узлов, а E - набор рёбер;
- 2) $s: E \rightarrow V$ и $t: E \rightarrow V$ - это функции, которые определяют исходный и целевой узлы рёбер;
- 3) Σ_V - это набор меток узлов, а Σ_E - набор меток рёбер;
- 4) $\ell_V: V \rightarrow \Sigma_V$ и $\ell_E: E \rightarrow \Sigma_E$ - это функции, которые определяют метки узлов и рёбер.

Поскольку ГЗ и АГЗ являются инженерными артефактами, с ними связаны процессы разработки, тестирования, утверждения, управления, а также процессы всего жизненного цикла. По аналогии с DataOps [Liebmann, 2020], набор практик, объединяющий все эти процессы с системами ГЗ, можно назвать KnowOps. Несмотря на название «граф знаний», не существует требования, чтобы утверждения были реализованы в виде графа. СГЗ могут сами по себе быть источниками для всеобъемлющего ГЗ, который объединяет исходные ГЗ.

Хотя онтологии не указаны в определении ГЗ, они могут играть важную роль, даже если эта роль варьируется. Некоторые ГЗ включают онтологию как часть структуры, и в этом случае понятия ГЗ и онтологий по существу эквивалентны. В других случаях ГЗ и онтология не связаны, и возможно, что у одного ГЗ имеется более одной связанной онтологии, так что онтология играет роль представления реляционной базы данных.

3 Зачем использовать графы знаний?

Рассмотрим вопрос, почему ГЗ и СГЗ стали популярными. Он является частью более обширного вопроса, провокационно заданного Мэтью Уэстом [West, 2020] о том, зачем вообще нужна информация. В этом более широком контексте Янс Аасман предложил несколько причин, по которым ГЗ в последнее время стали настолько популярными [Aasman, 2019].

В бизнесе информация используется для поддержки решений. Если информация, необходимая для принятия решения, отсутствует или неточна, риск ошибки увеличивается. Итак, для поддержки решения информация должна соответствовать назначению, а это значит, что управление информацией - это процесс управления качеством, в котором информация является продуктом. Но как узнать, каковы требования к информации? Оказывается, опрос людей об их требованиях даёт ненадёжные результаты. Лучше всего записать процессы до уровня,

на котором ключевые решения объяснены. Затем можно задокументировать требования к информации для этих решений.

Информация имеет множество свойств, но только некоторые из них имеют решающее значение для её использования при принятии решений. Одно из самых сложных свойств - совместимость. Если данные совместимы, то, когда они поступают из разных источников, их можно просто объединить и сразу же использовать. Согласованные данные используют одну и ту же модель данных и контрольные данные (или, если хотите, ГЗ одной и той же онтологии). Однако, если источники не согласуются ни по отдельности, ни друг с другом, то необходимо не только извлечь информацию из источников, но и устранить несоответствия. Следовательно, для этого необходимо разработать набор инструментов. Другими словами, для информации необходима программная архитектура.

Учитывая, что нужна система сбора знаний, возникает естественный вопрос, что сделало ГЗ такими популярными. Хотя о причинах можно только догадываться, вот возможные объяснения.

- Графические базы данных теперь признаны лучшей технологией для хранения сложных семантических данных.
- Люди больше не боятся таксономий, хотя онтологии по-прежнему пугают.
- Извлечение объектов и NLP теперь стали ценными наряду с SPACY, BERT, IBM Natural Language Understanding и многими другими инструментами.
- Машинное обучение и расширенная аналитика теперь доступны в облаке [Aasman, 2019].

Обратите внимание, что в этом обосновании нет возможности рассуждать/делать выводы. Действительно, существуют успешные СГЗ, которые либо имеют минимальную схему, либо не уделяют ей значительного внимания. Тем не менее, существует общее мнение о полезности онтологий для СГЗ.

4 Методы и инструменты графа знаний

В этом разделе приведены примеры техник и инструментов, которые используются и разрабатываются для ГЗ. В разделе 4.1 описаны различные формы рассуждений и математические приёмы из теории вероятностей и теории категорий для ГЗ. В разделе 4.2 описана Открытая сеть знаний (Open Knowledge Network), программа NSF для ГЗ, которая обосновывает некоторые проекты в следующих подразделах. Одной из важных задач для ГЗ является пространственное и временное обоснование, и в разделе 4.3 представлены два проекта, направленных на её решение. Остальная часть этого раздела посвящена проектам из многих областей, в которых применялись методы ГЗ. Раздел 4.4 посвящён извлечению ГЗ из научных публикаций, раздел 4.5 посвящён ГЗ при проектировании и производстве продукции, в разделе 4.6 описаны два применения ГЗ для решения государственных задач, а в разделе 4.7 предложено использовать ГЗ для нового типа динамически интерактивного учебника. Для получения дополнительных сведений о каждом проекте см. ссылку на соответствующий слайд или видеопрезентацию в цитируемой ссылке.

4.1 Методы

Несмотря на различные определения понятия ГЗ, существует общая цель: использовать ГЗ для получения важных идей и открытий на основе данных. Анируддха Прабху определяет “идеи” как важные закономерности, тенденции и согласованную информацию, полученную из ГЗ, особенно в тех случаях, когда такие особенности не очевидны из простых задач исследования данных [Prabhu, 2020].

Использование аргументов для получения информации и выводов о данных - это метод, широко известный и используемый в сообществе семантической паутины. Но используя методы (как визуальные, так и аналитические), известные в сетевой науке, можно идентифицировать ранее невидимые закономерности и тенденции и использовать эти идеи для создания или проверки гипотез и помощи в научных открытиях.

Глобальные показатели используются для того, чтобы получить представление обо всей структуре сети и сравнить две или более сетей друг с другом. Локальные показатели используются для проверки структуры отдельной сети и выявления важных тенденций в этой сети. Алгоритмы обнаружения сообщества используются для математической идентификации групп узлов в сети, обычно в зависимости от того, как эти узлы связаны друг с другом. Наконец, Прабху изучил (как визуально, так и математически) эволюцию сети на основе изменения конкретной характеристики данных (например, времени, давления или температуры), чтобы определить, как добавление или удаление узла (или набора узлов) влияет на общую структуру сети.

Другой подход к получению информации - использовать вероятностные ГЗ, представленные [Srihari, 2020]. Эти ГЗ включают статистические модели для реляционных данных. Тройки считаются неполными и шумными. Существует два основных типа моделей: модели со скрытыми характеристиками и марковские случайные поля (MRF). Первые можно обучить с помощью глубокого обучения. MRF могут быть получены из марковских логических представлений фактов в базе данных.

Еще один способ получения информации - использовать математическую теорию категорий и функторов. В статье «Составление графов знаний изнутри и снаружи» Спенсер Брейнер объяснил, как можно формально устранить некоторые ограничения графических представлений знаний, используя базовые методы теории категорий [Breiner, 2020]. Хотя теория категорий считается очень абстрактной даже среди математиков, на самом деле категории тесно связаны с ГЗ. Категория состоит из набора объектов и стрелок (направленных связей) между ними, что и подразумевается под ориентированным графом. Этот подход можно применить к практическим вопросам. Чтобы проиллюстрировать это, была показана проблема планирования открытых магазинов в исследовании операций с использованием теории категорий.

4.2 Программа Open Knowledge Network

Open Knowledge Network (OKN) - это программа Национального научного фонда США, в цели которой входят следующие разработки:

- развитая инфраструктура научных данных, которая является интероперабельной и имеет открытую архитектуру, что упрощает доступ и связывание разнородных продуктов данных;
- открытая семантическая информационная инфраструктура для открытия новых знаний из множества разрозненных источников знаний;
- непатентованная инфраструктура общих знаний с особым упором на общедоступные данные, например, данные правительства США, научные данные и другие аналогичные общедоступные наборы данных [Baru, 2020].

OKN приносит пользу множеству областей применения, включая научные и инженерные исследования. Если говорить более кратко, то это «Siri для науки». Спонсируемые проекты OKN включают общие темы:

- интеграция разнородных типов данных;
- размещение динамической информации;
- поддержка доступа и взносов в ГЗ со стороны разнородных сообществ пользователей;

- включение новой информации в ГЗ с использованием подходов машинного обучения и краудсорсинга.
Некоторые из описанных ниже проектов спонсируются программой OKN.

4.3 Временные и пространственные проекты

Одна из проблем, стоящих перед ГЗ, - это проблема представления времени и пространства. Даже очень мощные системы ИИ могут не справиться с временем. Если вы спросите Google: «Сколько лет Джо Байдену?» или «Сколько лет Митчу МакКоннеллу?», вы получите правильные ответы; но если вы спросите: «Кто старше, Джо Байден или Митч МакКоннелл?», всё, что вы получите, - это ссылки на статьи, в которых упоминаются оба политика. Проблема в том, что хотя ГЗ обычно включают временные характеристики сущностей, они рассматриваются как текстовые строки, не более, без какой-либо другой семантики. Более того, многие характеристики и связи, которые фактически зависят от времени, такие как пространственная протяжённость стран, считаются вневременными. Эта ситуация удивительна, поскольку темпоральное мышление широко развито в ИИ и управлении базами данных. Некоторые аспекты временных и пространственных мышлений были рассмотрены на Онтологическом саммите 2018 по контексту и онтологиям [Baclawski et al, 2018]. Кроме того, органы по стандартизации разрабатывают стандарты временного и пространственного представления и мышления, как описано в разделе 5. К сожалению, в рамках ГЗ время является второстепенным, если оно вообще включается. Пространственное мышление имеет аналогичные проблемы, хотя потребность в пространственном мышлении встречается реже, чем во временном мышлении. Исследовательское сообщество ГЗ должно исследовать все аспекты времени и пространства, от абстрактных до конкретных, от общих рассуждений до весьма конкретных применений. В долгосрочной перспективе преимущества эффективных, гибких временных и пространственных мышлений для систем ИИ будут большими [Davis, 2020]. Следующие два проекта пытаются разобраться в данной проблеме.

KnowWhereGraph, разработанный Кшиштофом Яновичем, - это проект, который выводит географическую информационную систему (ГИС) на новый уровень, предоставляя связанные технологии на основе открытых графов и семантического обогащения, выходящие далеко за рамки заранее определённых тем данных и разрозненных хранилищ [Janowicz, 2020]. Конечная цель - понять, как создавать значимые характеристики (независимые переменные) с помощью ГИС на базе ГЗ для последующих моделей, таких как прогнозирование цепочки поставок или картографирование состояния почвы, путём включения пространственно-временной семантики.

Шон Гордон является частью команды, которая создаёт прототип OKN для поддержки пространственных решений [Gordon, 2020]. На основе существующей работы членов группы были созданы четыре подгруппы по изучению конкретных примеров, которые работают над анализом потребностей многосторонних организаций, сосредоточенных на трёх основных экологических темах (качество воды, лесные пожары, биоразнообразие) в различных регионах запада США; была создана одна подгруппа по изучению конкретных примеров, которая работает над профессиональным комплексом знаний в области географической информации и технологий (GIS&T BoK). Каждая из четырёх подгрупп по изучению конкретных случаев использовала интервью и/или семинары, чтобы сотрудники выявляли проблемы и вопросы, которые необходимо знать (NTKC/Qs). Этот подход помог приоритезировать: а) схему ГЗ, которая будет поддерживать принятие решений по каждой теме; б) ресурсы поддержки пространственных решений для добавления в ГЗ; в) конкретные варианты использования.

4.4 Научные публикации

Основной продукт науки - публикации. Существует около 30 000 журналов, и ежегодно публикуется около двух миллионов статей. Усилия по извлечению знаний из научных записей появились ещё до появления Всемирной паутины [Baclawski, Futrelle, Fridman, and Pescitelli, 1993; Baclawski et al, 1993]. Иоланда Гил описывает семь онтологий, которые предоставляют необходимые возможности, но ещё предстоит проделать большую работу для более полного сбора научных данных. Неужели мы далеки от того дня, когда каждая научная статья будет правильно связана с гипотезами, моделями, программным обеспечением, происхождением, рабочими процессами и другими ключевыми научными объектами в Интернете, как показано на рисунке 3? Смогут ли инструменты ИИ обеспечить доступ к этой информации для получения новых результатов? Смогут ли системы ИИ в конечном итоге автономно писать научные статьи в будущем? [Gil, 2020].

Capturing Scientific Knowledge

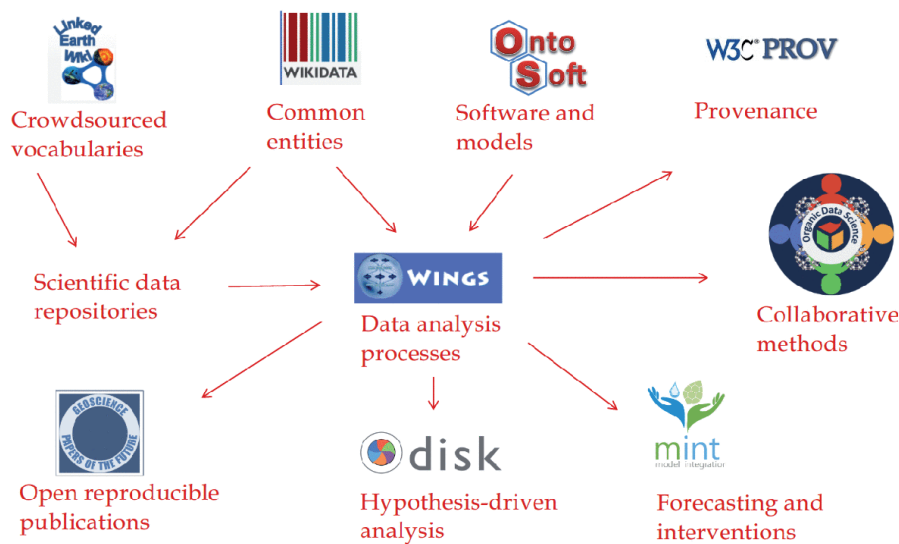


Рисунок 3 - Связи между объектами научного знания из (Gil, 2020)

4.5 Производство

Задача проекта *Manufacturing Open Knowledge Graph* (MOKN) - структурировать всемирную общедоступную информацию о проектировании и производстве продукции [Starly, 2020]. Большее влияние MOKN заключается в том, чтобы сделать доступной информацию о поставках критически важных компонентов, мгновенном сборе данных о конкретных производственных возможностях, местонахождении этих услуг и доступности ресурсов. Глобальный пандемический кризис является примером важности этих знаний, особенно для альтернативных источников и предварительной квалификации поставщиков, что может иметь последствия для здравоохранения и национальной безопасности. Доступность также расширяет возможности в сёлах и маленьких городках, зависящих от производственных услуг.

4.6 Правительство

В Великобритании Мэтью Уэст участвует в амбициозной попытке создать Цифрового двойника (Digital Twin) всей национальной инфраструктуры. Цель - создать распределённого Цифрового двойника согласованных данных, чтобы авторизованные пользователи могли де-

лать к нему запросы, и он мог ответить на такие вопросы, как: «Какие блоки башни имеют такой же тип облицовки, как башня Гренфелл?» Ландшафт управления информацией выдвигает информацию, необходимую для поддержки важнейших свойств данных и процесса управления качеством информации. Частью этой инфраструктуры является интеграционная архитектура, которая позволяет виртуализировать распределённый национальный Цифровой двойник, чтобы пользователи могли видеть его как единую базу данных с доступом только к тем данным, которые им разрешено просматривать [West, 2020]. Фактически, это система ГЗ, для которой базовые исходные данные извлекаются из большого набора систем ГЗ, каждая из которых предназначена для одного города или небольшого региона.

Проект *Rich Context*, описанный Пако Натаном, представляет собой АГЗ платформы административных исследований данных (*Administrative Data Research Facility - ADRF*), которая в настоящее время используется 50 федеральными, государственными и местными агентствами в США для выявления людей с конкретным опытом [Nathan, 2020]. ADRF был назван первым примером безопасного доступа к конфиденциальным данным в заключительном отчёте Комиссии по разработке политики на основе фактов.

4.7 Образование

Студенты колледжей сегодня сталкиваются с проблемой усвоения концепций в новых предметных областях и соотнесения этих концепций с несколькими дисциплинами, хотя их учебники можно охарактеризовать как «один размер подходит всем». В работе «Учебник сети открытых знаний» Винеи К. Чаудри представил интеллектуальные учебники (*Intelligent Textbooks - ITB*) с использованием ИИ и ГЗ для решения этих проблем. Учащиеся могут динамически взаимодействовать с содержанием учебника, повышая их способность понимать концепции, повышая вовлечённость, улучшая академическую успеваемость [Chaudhri, 2020].

5 Стандарты

Далее представлены некоторые стандарты, относящиеся к ГЗ. Что делает стандарты особенно полезными для ГЗ, так это то, что между многими разработанными ГЗ есть существенные различия. Стандарты могут помочь таким разрозненным ГЗ взаимодействовать друг с другом. Стандарты также служат целям развития ГЗ. Например, можно разработать стандарт для представления объектов и отношений для производственного ГЗ, который можно использовать во всем мире для разработки ГЗ в определённой области. Эти ГЗ могут быть легко интегрированы на более позднем этапе. Системы ГЗ различаются не только источниками своих знаний (например, Интернет, сенсорные данные в некоторых областях, данные коммерческих транзакций и т.д.), но и операциями по генерированию, обработке и использованию результатов. Например, поддерживает ли система ГЗ рассуждения? Если да, то что за рассуждения? Доступен ли для рассуждений целый ГЗ? Когда используются рассуждения или умозаключения, ожидается, что результат такого действия даст результаты, согласующиеся с ожидаемой интерпретацией. Такая интерпретация основана на различиях между объектами, участвующими в умозаключении, и выражается (обычно на естественном языке) через символы (также известные как метки), используемые в представлении.

ГЗ создается для удовлетворения определённых потребностей и использования в определённом контексте, хотя контекст может быть недостаточно понятен (или представлен). Следовательно, ГЗ обязательно будет иметь ограничения по охвату (т.е. поле деятельности) и полноте (уровню детализации), что будет препятствовать взаимодействию. Есть несколько способов справиться с этой проблемой. Один из вариантов - использовать онтологический анализ при создании ГЗ. С этим связано использование хорошо разработанной онтологии в

качестве схемы, основанной на таком анализе. Другой вариант - использовать применимые стандарты (например, инженерные, терминологические, логические и т.д.), которые являются предметом данного раздела.

В статье о роли стандартов в инновациях Аллен и Шрирам заявляют: «Стандарты - это задокументированные соглашения, содержащие технические рекомендации, обеспечивающие соответствие материалов, продуктов, процессов, представлений и услуг своему назначению» [Allen and Sriram, 2000]. Затем они обсуждают, как стандарты, введенные в нужное время, приведут к большим инновациям. Например, стандартизованная нотная грамота на протяжении сотен лет стимулирует творческие музыкальные композиции.

Лиза Карнахан более подробно остановилась на стандартах и процессе их создания в своём выступлении «Стандартный процесс ИТ» [Carnahan, 2020]. В США стандарты разрабатываются организациями по разработке стандартов (SDO). SDO - это любая организация, которая разрабатывает и утверждает задокументированные стандарты с использованием различных методов для достижения консенсуса среди своих участников. Существуют сотни SDO. Такие организации могут быть: аккредитованными (например, аккредитованными ANSI INCITS или NIST ITL); на основе международных договоров (например, ITU-T, ICAO); на основе международного частного сектора (например, ISO/IEC или IEEE); международный консорциум (например, OMG, OASIS, IETF, W3C); или государственное учреждение (например, DoD, DHS, NIST).

Одна из SDO - это ISO, крупнейший в мире разработчик добровольных международных стандартов. Барри Смит рассказал о своём опыте улучшения взаимодействия ГЗ в статье «От BFO до IOF и до ISO/IEC 21838» [Smith, 2020]. Этот стандарт называется «Информационные технологии - онтологии верхнего уровня» и включает в себя в качестве одной из своих частей Базовую формальную онтологию (BFO). Онтологии были чрезвычайно успешными в биомедицинской области в течение 20 лет, где геновая онтология (GO), представляющая знания о генах и генных продуктах и первая версия которой была создана в 1998 году, с самого начала называлась «направленным ациклическим графом». Фундаментальная онтология GO - это BFO. С ростом влияния данных из проектов генома человека и других модельных организмов потребность в аннотации данных в мире биомедицинской информатики значительно расширилась, и это привело к созданию новых онтологий, например, для белков, типов клеток, болезней и другие. Это развитие онтологий продолжается и по сей день с новой онтологией COVID-19. О влиянии BFO в немедицинских областях свидетельствует также принятие стандарта ISO/IEC 21838 в таких областях, как цифровое производство, в частности, благодаря созданию *Industrial Ontology Foundry* (IFO). Под эгидой этой организации продолжается работа по увязке BFO с текущими разработками по стандарту STEP (стандарт обмена данными модели продукта) (ISO 10303) и стандарту MTConnect для данных заводских устройств [«Industrial Ontology Foundry», 2020].

Другой SDO - это Группа управления объектами (OMG), наиболее известная благодаря унифицированному языку моделирования (UML) и архитектуре, управляемой моделями (MDA). Элиза Кендалл предоставила обновлённую информацию о стандартах OMG и деятельности, имеющих отношение к онтологиям и ГЗ [Kendall, 2020]. Специальная группа по онтологической платформе (OPSIG) является активной рабочей группой более 15 лет. К настоящему времени опубликованы следующие стандарты платформы: Метамодель определения онтологии (ODM); Язык распределённой онтологии, моделей и спецификаций (DOL); Языки, страны и коды (LCC); и преобразование MOF в RDF - MOF2RDF. Группа OMG также опубликовала несколько онтологий для конкретных предметных областей, в том числе бизнес-онтологию финансовой индустрии (FIBO), глобальный идентификатор финансовых инструментов (FIGI) и словарь политик обмена информацией (IEPPV). Онтология ро-

ботизированных сервисов в настоящее время готовится совместно с сообществом IEEE Robotics.

Другие стандарты, относящиеся к ГЗ, включают Модель и нотацию управления делами (CMMN), Модель и нотацию решения (DMN), Словарь даты и времени (DTV), Представление производственных правил (PRR) и Семантику бизнес-словарей и правил (SBVR). Третьи находятся в стадии подготовки. Однако единого золотого стандарта для ГЗ не будет. Появятся несколько стандартов, которые необходимо будет разумно объединить.

Общая логика (CL) - это стандарт ISO/IEC для логики первого порядка (ISO/IEC 24707:2007). В статье «Графы знаний и логика» Джон Сова сделал обзор CL и связанных с ним стандартов логики [Sowa, 2020]. Стандарт CL включает спецификации для трех диалектов: формата обмена общей логикой (CLIF), формата обмена концептуальными графами (CGIF) и нотации на основе XML для общей логики (XCL). Диалект CLIP сочетает в себе лучшие черты двух диалектов, CLIF и CGIF. Основные цели проектирования CLIP следующие:

- возможность немедленного чтения любым, кто разбирается в исчислении предикатов;
- читается так же, как Turtle для подмножеств RDF и OWL;
- читаема, как любая нотация правил «если-то»;
- служить линеаризацией для широкого диапазона логик графов, включая диаграммы CG, EG, KG, RDF, OWL и UML;
- параметр запроса: выберите (список имен), где (любое предложение CLIP);
- поддержка сопоставлений между логикой и естественными языками (NL).

Упомянутый выше стандарт DOL является стандартом OMG для интеграции и взаимодействия между распределёнными онтологиями, моделями и спецификациями (OMS). DOL формально определяется логикой и математикой. Другими словами, DOL может интегрировать гетерогенные OMS, связывая логики, которые их определяют.

Индустрия финансовых услуг - это крайне большая, зрелая и требовательная к обработке данных отрасль, которая оказывает влияние практически на всех. Майкл Беннетт представил обзор ГЗ в финансовом секторе [Bennett, 2020]. Хотя большинство исторических стандартов в индустрии финансовых услуг касается требований к обмену сообщениями или форматам данных, существуют также отраслевые стандарты формальной семантики. Бизнес-онтология финансовой индустрии (FIBO) была задумана, чтобы обеспечить общий язык для этих стандартов обмена сообщениями, в то время как недавняя инициатива ISO TC68 (Технический комитет ISO, занимающийся финансовыми услугами) направлена на дополнение существующего стандарта обмена сообщениями XML ISO 20022 формальной семантикой. FIBO возникла из-за необходимости унифицировать термины в отрасли как общий язык для повторного использования данных в отчётности, управлении рисками и соблюдении нормативных требований. Эта потребность возникла из-за осознания того, что концепции были хорошо поняты несмотря на то, что было трудно достичь согласия по общим условиям.

Хотя финансовая индустрия является специфической областью, она даёт важные уроки, которые актуальны для онтологий и ГЗ в целом. Например, одно различие состоит в том, предоставлять ли глубокую иерархию фундаментально примитивных терминов, основанную на онтологии верхнего уровня (TLO), или нет. Обычно они не нужны для приложений OWL и были удалены из стандарта OMG FIBO. Другое различие состоит в том, представляет ли онтология реальных «творцов истины» (утверждения, которые порождают значение класса вещей) или данные о вещах. Например, *быть* банком значит обладать определёнными правоспособностью и возможностями, тогда как *знать, что что-то является* банком, - значит исследовать доступные данные для некой подходящей «подписи данных», что такие возможности существуют, в данном случае в форме лицензии банка. Поэтому онтологии могут

быть основополагающими для использования в качестве ориентира или могут быть ориентированы на приложения; и они могут быть основаны на предмете изучения или данных о предмете. Эти различия могут быть отклонены разработчиками как несущественные, но если их не учитывать, то в результате совместимость может быть серьёзно затруднена.

В «Стандартах и онтологиях» Михаэль Грюнингер обсудил преимущества и недостатки стандартизации онтологий [Gruninger, 2020]. Проблема со стандартами де-факто состоит в том, что онтологии будут приняты просто потому, что они популярны и широко используются, даже если они не были должным образом разработаны с достаточной оценкой и анализом. Риск этого подхода заключается в том, что могут использоваться онтологии, содержащие онтологические ошибки, непреднамеренные модели и пропущенные модели, или они могут включать неявные онтологические обязательства, предотвращающие повторное использование. Следовательно, нам нужны стандарты, которые позволяют оценивать и сравнивать онтологии. Прежде всего, это стандарты для языков представления онтологий с формальной семантикой, таких как *Common Logic* (ISO 24707) и *OWL* (W3C). Во-вторых, стандарты для спецификации сопоставлений между онтологиями и между логиками, ярким примером которых является язык распределённых онтологий (DOL) от Группы управления объектами. Наконец, существуют стандартизированные аксиоматизации онтологий, в частности ISO 18629 (язык спецификации процессов) и ISO 21838 (онтологии верхнего уровня).

В области стандартов ещё предстоит сделать большую работу. Недавно Международная ассоциация по прикладным онтологиям (IAOA) учредила Технический комитет по промышленности и стандартам (ISTC). Этот комитет преследует две основные цели:

- содействовать использованию прикладной онтологии в инициативах по стандартизации,
- облегчить взаимодействие между людьми в промышленности и в прикладных онтологических исследованиях.

Деятельность комитета ISTC включает распространение информации об инициативах с целью созыва экспертов, заинтересованных в разработке онтологически обоснованных стандартов. Комитет также организует виртуальные и физические встречи и мероприятия, чтобы обсудить, как понять и применять онтологические подходы и методологии, как в целом, так и для систем ГЗ в частности.

6 Проблемы

Методы построения ГЗ ведут от сырых, беспорядочных и разрозненных данных, которые трудно запросить, анализировать и визуализировать, к более точному, организованному, улучшенному и связанному продукту, который легче визуализировать, запрашивать и анализировать. Проблемы существуют на каждом этапе этого процесса, включая рекурсии как часть жизненного цикла. В этом разделе перечислены некоторые из этих проблем. Первый столбец в таблице - это действие АГЗ. Эти действия помечены как «Шаг ГЗ», потому что они обычно представляют собой шаги в цепочке действий, как показано на рисунке 4.

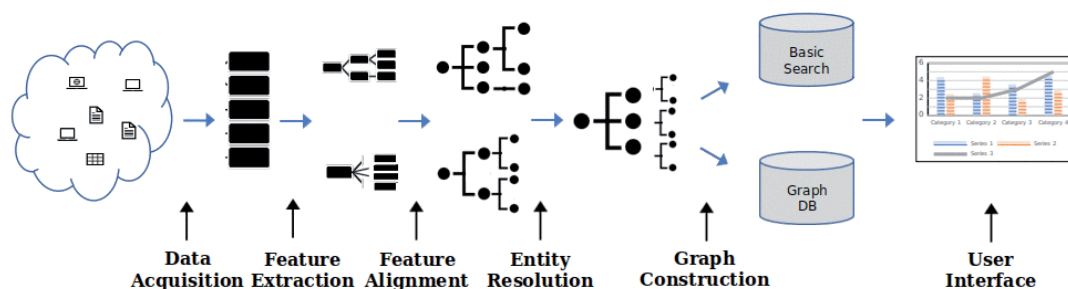


Рисунок 4 - Цепь построения ГЗ

Для каждого шага ГЗ может присутствовать много проблем. Наиболее важные из них перечислены во втором столбце таблицы. В следующем столбце описан контекст проблемы. В последнем столбце приведены ссылки. Более подробно об этих проблемах будет опубликовано в отдельной статье.

Таблица – Проблемы построения ГЗ

Шаг ГЗ	Проблема	Контекст	Заметки	Ссылки
Определение объёма	Выявление лучших доступных источников в огромном пространстве возможностей	Понимание требований к использованию и знаниям	Определение фактов о кандидатах, которые следует включить в ГЗ	[Pujara, Miao, Getoor, and Cohen, 2013]
Сбор данных и популяции	Объём, разнообразие, скорость и достоверность данных. Может быть слишком мало структурированных данных для заполнения графика	Объём необходимых данных может быть неизвестен. Метаданные домена и междоменные метаданные со временем развиваются.	Мы не будем знать все типы и отношения, необходимые для моделирования. Необходимо изучить окрестности сущностей и сравнить соседние сущности и значения.	[Dong, 2020]
Выделение признаков	Нужны зрелые процессы для поиска типов и создания векторных функций, используемых в модели машинного обучения. Ограниченные обучающие метки для больших объёмов данных. Часто скрытые шаблоны, например, в заголовках, несут в себе ключевые связи, атрибуты, даты и т. д.	Помимо технических проблем, есть ли необходимость в проверке выделения, чтобы понять, согласуются ли извлечённые функции с человеческим мнением?	Активное обучение, слабое обучение, дистанционное наблюдение вместе с полуконтролируемым обучением, трансферным обучением и мета-обучением - всё это методы для работы с ограниченными данными обучения. Менталистическая терминология функций может ввести в заблуждение тех, кто не занимается компьютерными науками. Важная информация также находится в изображениях, что затрудняет извлечение функций.	[Joshi, 2019; Dong et al, 2020; Wang, Xu, Li, Dong, Gao, 2020]
Выравнивание функций	Гетерогенные данные и большие пространства данных затрудняют согласование многих записей	Две функции одинаковы? Означают ли «родился» и «дата рождения» одно и то же?	Изучить избыточность данных. Находятся ли значения одного и того же атрибута в одном пространстве для встраивания?	[Pham, Alse, Knoblock, and Szekely, 2016; Taheriyani, Knoblock, Szekely, and Ambite, 2016]
Разрешение сущности	Проблемы с шумными данными и масштабирования. Используются альтернативные текстовые формулировки.	Не все данные заслуживают доверия и разнородны. Данные и большие пространства данных требуют разрешения многих записей	Это большой вызов. Используются статистические методы и методы машинного обучения, но понимаем ли мы диапазон возможных ошибок, которые могут возникнуть в извлечённых фактах?	[Zhu et al, 2020]

Финальное построение графа. Этот шаг может включать новые ссылки и конфиденциальные сведения о фактах и связях	Рабочее решение может не масштабироваться до большего количества данных.	По мере добавления новых данных вводятся разные словари и разные шаблоны могут кодировать один и тот же атрибут.	Построение графа можно рассматривать как инкрементный процесс с окончательной сборкой, который может включать в себя проверку семантических отношений из ведущей онтологии.	[Madison, Barnhill, Napier, and Godin, 2015; Deprizio, 2020]
Пользовательские интерфейсы	Насколько гибки интерфейсы для пользователей? Можно ли легко визуализировать ГЗ при отображении отношений и связывания сущностей?	Предоставляется ли объяснение?	Разрешить клиентам указывать информацию и указывать, какое требование менее важно в рамках смягчения или уточнения запроса.	[He et al, 2019]
Рассуждение	Насколько хорошо АГЗ поддерживает временные и пространственные мышления?		См. раздел 4.3	[Davis, 2020]

7 Будущее графов знаний

Некоторые возможности будущего развития и использования ГЗ, в первую очередь в промышленности, но также и для исследовательского сообщества ГЗ:

- эффективное определение «граф знаний» будет широко распространено;
- разработчики ГЗ поймут необходимость хорошо продуманной схемы и то, как онтологии или, по крайней мере, онтологический анализ могут в этом помочь;
- разработчики ГЗ будут использовать лингвистический анализ, чтобы помочь преодолеть двусмысленность использования терминов (и идентификаторов) естественного языка;
- разработчики ГЗ будут включать формальные различия для предполагаемых интерпретаций терминов и фраз естественного языка, используемых для обозначений в ГЗ, вместо неудачной практики полагаться на предполагаемые общие интерпретации семантики таких терминов и фраз;
- ГЗ будут использоваться при создании и работе программно-интенсивных систем (например, для представления пользовательских интерфейсов);
- архитекторы информационных систем будут лучше использовать ГЗ и их инфраструктуру для поддержки более динамичных информационных систем;
- архитектуры будут разработаны, чтобы помочь предприятиям и их обширным информационным системам в переходе к использованию ГЗ;
- ГЗ окажут значительное влияние на управление данными и знаниями в целом.

8 Заключение

Графы знаний - эффективные инструменты для информационных систем и очень популярная тема, несмотря на отсутствие общего определения того, что же такое ГЗ. В этом коммюнике рассматривается понятие ГЗ и достигнут некоторый прогресс в уточнении краткого практического определения того, что такое ГЗ, которое не только сопоставимо с основными опубликованными определениями, но и проясняет источники путаницы вокруг этого понятия. Здесь представлены исторические тенденции, которые сошлись на ГЗ, и предложены некоторые из причин, почему ГЗ стали настолько популярными. Было описано несколько при-

меров методов, используемых АГЗ и разработанными СГЗ. В настоящее время существуют или разрабатываются стандарты, относящиеся к ГЗ. Несмотря на то, что ГЗ добились успеха, вопросы и проблемы ещё остаются.

9 Благодарность

В документе указаны некоторые коммерческие программные системы. Их упоминание не подразумевает рекомендации или одобрения Национальным институтом стандартов и технологий (NIST) или организациями авторов или сторонников этого Коммюнике; это также не означает, что указанные продукты обязательно являются лучшими из имеющихся для этой цели. Кроме того, любые мнения, выводы или рекомендации, выраженные в этом материале, принадлежат авторам и не обязательно отражают точку зрения NIST или любых других организаций, поддерживаемых правительством США или другими корпорациями.

Мы хотим поблагодарить сообщество онтологов за поддержку, особенно приглашённых докладчиков и участников, которые внесли свой вклад в онтологический саммит. Приглашенных докладчиков было 22: Янс Аасман, Андреас Блюмауэр, Барри Смит, Эрнест Дэвис, Анируд Прабху, Саргур Шрихари, Спенсер Брейнер, Пако Натан, Винеи К. Чаудри, Кшиштоф Янович, Бинил Старли, Шон Гордон, Майкл Ушольд, Йоланда Гил, Мэтью Уэст, Чайтанья Бару, Лиза Карнахан, Элиза Кендалл, Майкл Грюнингер, Джон Ф. Сова, Майкл Беннетт и Элиза Стиклс. Полный список сессий, докладчиков и ссылки на слайды презентаций и видеозаписи доступны на <http://bit.ly/33NydJ9>.

Неполный список других участников включает: Кингсли Иден, Джанет Сингер, Дуг Фоксвог, Джек Ходжес-младший, Алекс Шкотин, Сджир Нейссен, Пол Тайсон, Майкл ДеБеллис, Эдвард Баркмейер, Азамат Абдуллаев, Амит Шет, Паскаль Хитцлер, Алессандро Олтрамари, Джек Парк, Джордж Херлберт, Рассел Райнш, Мария Евтимова и Брюс Баргмайер. Мы особенно благодарим Тодда Шнайдера за предложение темы ГЗ для онтологического саммита 2020.

References

- Aasman, J. (2019) Why Knowledge Graphs Hit the Hype Cycle and What they have in common. Retrieved on December 1, 2020 from <http://bit.ly/34jSlmJ>.
- Aijal, J. (2019) What is a knowledge graph and how does one work? Retrieved on December 1, 2020 from <http://bit.ly/2IwjVTu> and <https://thenextweb.com/podium/2019/06/11/what-is-a-knowledgegraph-and-how-does-one-work/>.
- Allen, R.H. & Sriram, D. (2000) The Role of Standards in Innovation, Special Issue on “Innovation: The Key to Progress in Technology and Society”, Journal Technological Forecasting and Social Change.
- Baclawski, K., Bennett, M., Berg-Cross, G., Casanave, C., Fritzsche, D., Ring, J., Schneider, T., Sharma, R., Singer, J., Sowa, J., Sriram, R.D., Westerinen, A. & Whitten, D. (2018) Ontology Summit 2018 Communiqué: Contexts in Context, J. Applied Ontology, IOS Press.
- Baclawski, K., Futrelle, R., Fridman, N. & Pescitelli, M. (1993) Database techniques for biological materials & methods. In First Int. Conf. Intell. Sys. Molecular Biology 21-28.
- Baclawski, K., Futrelle, R., Hafner, C., Pescitelli, M., Fridman, N., Li, B. & Zou, C. (1993) Data/knowledge bases for biological papers and techniques. In Proc. Sympos. Adv. Data Management for the Scientist and Engineer 23-28.
- Baru, C. (2020) The Open Knowledge Network. Retrieved on December 1, 2020 from <https://go.aws/31rSjbe>.
- Bennett, M. (2020) Standards for KGs in the Financial Sector. Retrieved on December 1, 2020 from <https://go.aws/2YXCdXw>.
- Bergman, M. (2019) A common sense view of knowledge graphs. Retrieved on December 1, 2020 from <http://bit.ly/307PEBs> and <http://bit.ly/2RAbE6X>.
- Blumauer, A. (2014) From Taxonomies over Ontologies to Knowledge Graphs. Retrieved on August 1, 2020 from <https://blog.semantic-web.at/2014/07/15/from-taxonomies-over-ontologies-to-knowledgegraphs>.
- Breiner, S. (2020) Composing Knowledge Graphs, inside and out. Retrieved on December 1, 2020 from <https://go.aws/2QfatbQ>.
- Carnahan, L. (2020) The IT Standard Process. Retrieved on December 1, 2020 from <https://go.aws/3gPkyYK>.
- Chaudhri, V. (2020) Textbook Open Knowledge Network. Chaudhri Retrieved on December 1, 2020 from <http://bit.ly/310xXpd>.
- Davis, E. (2020) Time and Space in Knowledge Graphs. Retrieved on December 1, 2020 from <https://go.aws/2SxOVZ9>.
- Deprizio, J. (2020) Comparative Analysis of Database Spatial Technologies (CADST). Dissertation. George Mason University.
- Dong, L. (2020) Knowledge Graph and Machine Learning: A Natural Synergy, Presentation at Stanford Seminar on KGs. Stanford University.

- Dong, X., He, X., Kan, A., Li, X., Liang, Y., Ma, J., Xu, Y., Zhang, C., Zhao, T., Saldana, G., Deshpande, S., Manduca, A., Ren, J., Singh, S., Xiao, F., Chang, H.-S., Karamanolakis, G., Mao, Y., Wang, Y., Faloutsos, C., McCallum, A. & Han, J. (2020) AutoKnow: Self-driving knowledge collection for products of thousands of types, SigKDD 2020.
- Färber, M., Ell, B., Menne, C., Rettinger, A., & Bartscherer, F. (2018) Linked Data Quality of DBPedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web Journal* 9 (1), 77-129.
- Gil, Y. (2020) Seven Ontologies for Publishing the Scientific Record on the Web. Retrieved on December 1, 2020 from <https://go.aws/2yHUuO4>.
- Gordon, S. (2020) Prototyping an Open Knowledge Network for Spatial Decision Support. Retrieved on December 1, 2020 from <http://bit.ly/2Kquljr>.
- Grüninger, M. (2020) Standards and Ontologies. Retrieved on December 1, 2020 from <https://go.aws/2AykuMf>.
- He, X., Zhang, R., Rizvi, R., Vasilakes, J., Yang, X., Guo, Y., He, Z., Prosperi, M., Huo, J., Alpert, J. & Bian, J. (2019) ALOHA: developing an interactive graph-based visualization for dietary supplement knowledge graph through user-centered design. *BMC medical informatics and decision making* 19.4
- "Industrial Ontology Foundry" (2020) Retrieved December 1, 2020 from <https://www.industrialontologies.org/>
- Janowicz, K. (2020) KnowWhereGraph: Enriching and Linking Cross-Domain Knowledge Graphs using Spatially-Explicit AI Technologies to Address Pressing Challenges at the Human-Environment Nexus. Retrieved on December 1, 2020 from <https://go.aws/2xmMSQd>.
- Joshi, P. (2019) Learn How to Perform Feature Extraction from Graphs using DeepWalk. Retrieved December 1, 2020 from <https://www.analyticsvidhya.com/blog/2019/11/graph-feature-extractiondeepwalk/>.
- Kendall, E. (2020) The Object Management Group. Retrieved on December 1, 2020 from <https://go.aws/3fISLc0>.
- Krötzsch, M. & Thost, V. (2016) Ontologies for knowledge graphs: Breaking the rules. In *International Semantic Web Conference*. Springer, Cham.
- Liebmman, L. (2020) 3 reasons why DataOps is essential for big data success. In *IBM Big Data & Analytics Hub*. Retrieved October 28, 2020 from <https://www.ibmbigdatahub.com/blog/3-reasons-whydataops-essential-big-data-success>.
- Madison, M., Barnhill, M., Napier, C. & Godin, J. (2015) NoSQL database technologies. *Journal of International Technology and Information Management* 24.1.
- Nathan, P. (2020) Rich Context Knowledge Graphs. Retrieved on December 1, 2020 from <https://go.aws/2TwytYO>.
- Paulheim, H. (2017) Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods. *Semantic Web Journal* 8(3):489-508.
- Pham, M., Alse, S., Knoblock, C. & Szekely, P. (2016) Semantic labeling: a domain-independent approach. In *International Semantic Web Conference*. Springer, Cham.
- Prabhu, A. (2020) Insights from Knowledge Graphs. Retrieved on December 1, 2020 from <https://go.aws/3a9Niax>.
- Pujara, J., Miao, H., Getoor, L., and Cohen, W. (2013) Knowledge Graph Identification. In *Proceedings of the 12th International Semantic Web Conference - Part I, ISWC'13*, pages 542-557, New York, NY, USA.
- Rohrseitz, N. (2019) Knowledge Graphs and Machine Learning: A powerful combination for the semiautomatic generation of insights. Retrieved on December 1, 2020 from <https://towardsdatascience.com/knowledge-graphs-and-machine-learning-3939b504c7bc> and <http://bit.ly/2ZWVmqa>.
- "Semantic Network" (2020) Retrieved November 2, 2020 from <https://bit.ly/36qXdct>.
- Smith, B. (2020) From BFO to IOF to ISO/IEC 21838. Retrieved on December 1, 2020 from <https://go.aws/2zY2Otx>.
- Sowa, J. (2020) Knowledge Graphs and Logic. Retrieved on December 1, 2020 from <https://go.aws/2LkvpN>.
- Srihari, S. (2020) Probabilistic Knowledge Graphs. Retrieved on December 1, 2020 from <http://bit.ly/36zrva9>.
- Starly, B. (2020) Building an Open Knowledge Network Graph in Product Design and Manufacturing. Retrieved on December 1, 2020 from <https://go.aws/2Xna2Ay>.
- Taheriyani, M., Knoblock, C., Szekely, P. & Ambite, J. (2016) Leveraging Linked Data to Discover Semantic Relations Within Data Sources. In *International Semantic Web Conference*. Springer, Cham.
- Wang, Y., Xu, Y., Li, X., Dong, X., Gao, J. (2020) Automatic validation of textual attribute values in eCommerce Catalog by learning with limited labeled data, In *KDD'20*.
- West, M. (2020) The Digital Twin Project in the UK. Retrieved on December 1, 2020 from <https://go.aws/2HdGBYr>.
- Yuan, W., Zhang, K., Dai, Q., Peng, C. & Zhao, K. (2018) Construction and Application of Knowledge Graph in Full-service Unified Data Center of Electric Power System. In *IOP Conf. Ser.: Mater. Sci. Eng.* 452 032065.
- Zhu, Q., Wei, H., Sisman, B., Zheng, D., Faloutsos, C., Dong, X. & Han, J. (2020) Collective multitype entity alignment between knowledge graphs. In *WebConf 2020*.