УДК 004.6+ 004.4

Эволюция понятия и жизненного цикла графов знаний

Апанович З.В.

(Институт систем информатики СО РАН, Новосибирский государственный университет)

В данной работе рассматривается эволюция понятия «граф знаний» с момента возникновения и до текущего момента. Также рассматривается вопрос о том, как эволюция систем, позиционирующих себя как графы знаний, повлияла на определение и жизненный цикл графов знаний.

Ключевые слова: *граф знаний, качество, покрытие, корректность, свежесть, происхождение.*

1. Введение

Первые использования термина «граф знаний» слабо связаны с современной практикой его применения. Этот термин появился еще в 1974 году в контексте определения интерактивного процесса обучения между обучаемым и обучающим. Вершинам графа знаний соответствовали единицы знаний, которые должен изучить обучающийся, а ребра между вершинами соответствовали отношению порядка изучения единиц знаний. [14].

В 1980-х годах исследователи из Нидерландов использовали термин «граф знаний» для формального описания их системы, основанной на извлечении знаний из медицинских и социологических текстов, и собирались постепенно увеличивать количество знаний в этом графе вплоть до построения экспертной системы. [24, 8].

Очередное «громкое» появление термина «граф знаний» под лозунгом «вещи, а не строки» (“things not strings”) было инициировано компанией Google в маркетинговых целях https://www.blog.google/products/search/introducing-knowledge-graph-things-not/. Лозунг подчеркивал полезность устранения неоднозначностей по отношению к сущностям, хранящимся в графе знаний. Смысл этого высказывания состоял в том, что текстовые строки часто неоднозначны, в то время как графы знаний состоят из сущностей, к которым применяется процедура устранения неоднозначностей, так что проще различать сущности, имеющие одинаковое название, но соответствующие разным объектам реального мира. С этого момента термин «граф знаний» стал использоваться по отношению к разным

продуктам, которые отличаются по таким характеристикам как архитектура, цель функционирования, и используемая технология, что затрудняет ответ на вопрос, почему они все позиционируются как «графы знаний».

Работа структурирована следующим образом. В разделах с 1 по 6 приводятся разные определения графов знаний и требования к графам знаний, а также примеры систем, позиционирующих себя как графы знаний, соответствующие этим определениям. На основании рассмотренных примеров предлагается схема жизненного цикла, соответствующего современным графам знаний.

2. Определения, связанные с понятием графа RDF

Поскольку современное использование графов знаний возникло в контексте направления Semantic Web, встречается много публикаций, в которых графом знаний определяют просто как *RDF-граф,* то есть множество триплет в виде (субъект, предикат, объект) [8, 13].

Например, предложение «Человек по имени Сергей Бондарчук снял фильм «Война и мир», снимался в главной роли в фильме «Судьба человека» и был женат на Инне Макаровой и Ирине Скобцевой» можно представить в виде множества триплет, показанных ниже.

dbr:Sergei_Bondarchuk rdf:type dbo:Person.
dbr:Inna Makarova rdf:type dbo:Person.
dbr:Irina Skobtseva. rdf:type dbo:Person.

dbr:Fate_of_a Man rdf:type dbo:Film.

dbr:War_and_Peace rdf:type schema:CreativeWork.

dbr:Sergei_Bondarchuk dbo:spouse dbr:Inna Makarova.

dbr:Sergei_Bondarchuk dbo:spouse dbr:Irina Skobtseva.
dbr:Sergei_Bondarchuk dbo:starring dbr:Fate_of_a Man.
dbr:Sergei_Bondarchuk dbo:director dbr:War_and_Peace.

Такой последовательности триплет соответствует граф, в котором вершинами являются субъекты и объекты триплет, а ребра помечены предикатами этих триплет. Например, показанное выше множество триплет можно изобразить при помощи графа, показанного на Рис. 1.

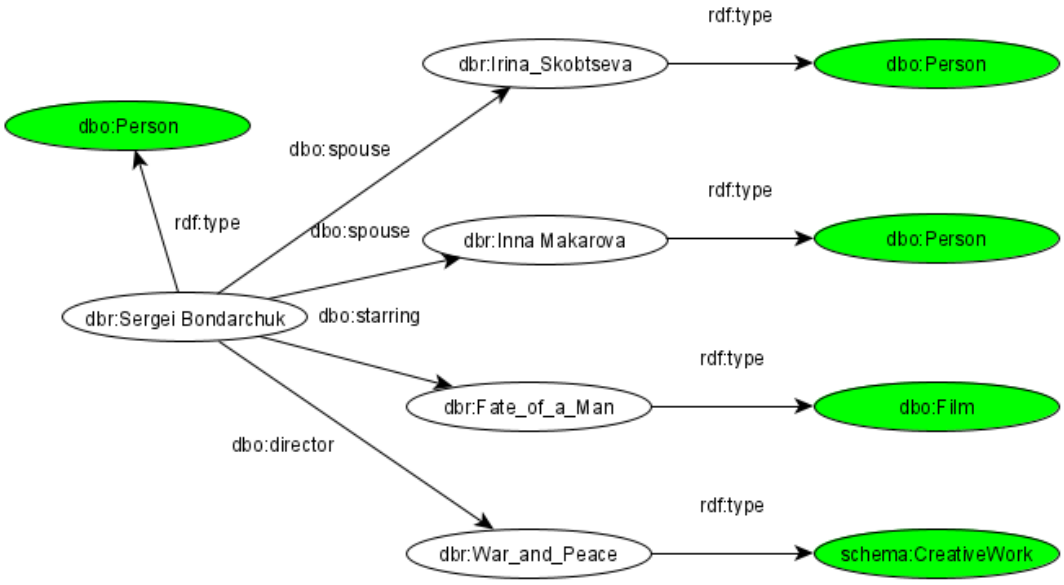

Рис 1. Пример RDF-графа.

Каждая триплета интуитивно представляет *утверждение*. Если граф знаний был построен правильно (со 100% точностью), и при этом данные собирались из достоверных источников данных, то эти «утверждения» можно было бы рассматривать как *факты*.

Такие факты часто представляют с помощью положительных унарных и бинарных логических предикатов логики первого порядка. Например, триплета
dbr:Sergei_Bondarchuk rdf:type dbo:Person
может быть представлена при помощи предиката
Person(Sergei_Bondarchuk), а триплета
dbr:Sergei_Bondarchuk dbo:director dbr:War_and_Peace
при помощи предиката
director (Sergei_Bondarchuk, "War_and_Peace").

Эти же самые факты можно представить при помощи «векторных вложений» "embeddings", которые создают вектора для каждой сущности и каждого отношения. Вектора кодируют латентные свойства сущностей и отношений и обладают свойством, что сходные сущности и сходные отношения будут представлены похожими векторами.

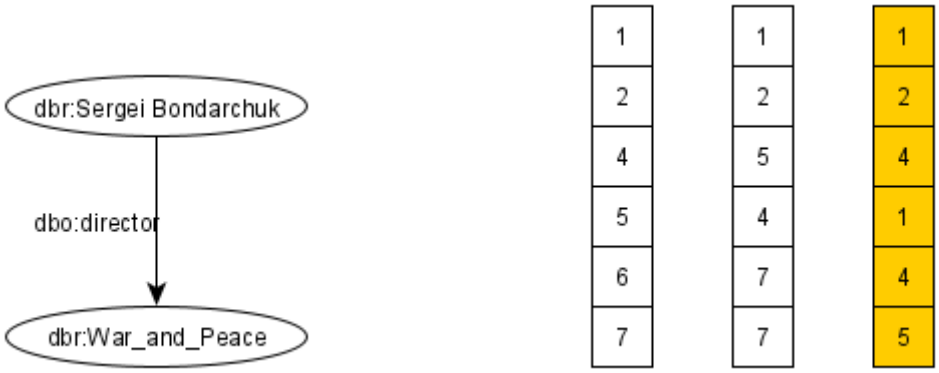

Рис 2. Пример того, что делают методы вложения с одной триплетой графа знаний.

Некоторые определения графов знаний на основе RDF подчеркивают более продвинутый характер графов знаний по сравнению с обычными Связанными данными [25]:

Граф знаний – это структурированный набор данных, собранный из разнородных источников данных, совместимый с моделью данных RDF и имеющий (OWL) онтологию в качестве своей схемы. Граф знаний не обязательно связан с внешними графами знаний; однако сущности в графе знаний обычно имеют информацию о типе, определенном в его онтологии, которая полезна для предоставления контекстной информации о таких сущностях. Графы знаний должны быть надежными, качественными, доступными и ориентированными на информационные услуги для конечного пользователя.

В этом определении в качестве характерной особенности графов знаний подчеркивается необходимость интеграции данных из множественных источников и повышенные требования к качеству контента.

Еще одно весьма популярное определение графов знаний приведено в работе [26]:
Граф знаний
1. в основном описывает сущности реального мира и их взаимосвязи, организованные в виде графа.
2. определяет возможные классы и отношения сущностей в схеме.
3. позволяет потенциально связывать произвольные объекты друг с другом.
4. графы знаний должны покрывать, по крайней мере, большую часть предметных областей, которые существуют в мире, и не должны ограничиваться только одной предметной областью.

Определению [26] удовлетворяют практически все междоменные графы знаний, предназначенные для поиска знаний в Интернете. Среди них DBpedia [17], YAGO [30], Wikidata [33], BabelNet [21], Cyc [18], NELL [6], CaLiGraph [11], VoldemortKG [32].

При этом следует обратить внимание, что все графы знаний из этого списка также обладают значительными объемами.

В таблице 1 показаны некоторые количественные характеристики междоменных графов знаний [12].

Таблица 1 Характеристики некоторых меж-доменных графов знаний [12].

	Экземпляров	**Триплет**	**Классов**	**Отношений**
DBpedia	**5 044 223**	**854 294 312**	**760**	**1355**
YAGO	**6 349 870**	**479 392 870**	**819292**	**77**
Wikidata	**52 252 549**	**732 420 508**	**2 356 259**	**6 236**
BabelNet	**7 735 436**	**178 982`397**	**6 044 564**	**22**
Cyc	**122 441**	**2 229 266**	**116 821**	**148**
NELL	**5 120 588**	**60 594 443**	**1 187**	**440**
CaLiGraph	**7 315918**	**517 099 124**	**755 963**	**271**
Voldemort	**55 861**	**693 428**	**621**	**294**

Значительным объемом контента обладают и наиболее известные графы знаний, представляющие ИТ-индустрию. В таблице 2 показаны некоторые количественные характеристики контента промышленных графов знаний [23].

Таблица 2 Характеристики известных промышленных графов знаний [23]

	Модель данных	Размер графа	Этап разработки
Microsoft	Типы сущностей, отношений и атрибутов графа определены в онтологии	2 миллиарда первичных сущностей, 22 миллиардов фактов	Активно используется в продуктах
Google	Строго типизированные	1 миллиард сущностей, 70 миллиардов	Активно используется в

	сущности, отношения с выводом области определения и области значений	утверждений	продуктах
Facebook	Все атрибуты и отношения структурированы и строго типизированы, опционально проиндексированы для эффективного извлечения, поиска и обхода	50 миллионов сущностей, 500 миллионов утверждений	Активно используется в продуктах
eBay	Сущности и отношения хорошо структурированные и строго типизированные	Ожидается около 100 миллионов продуктов, 1 миллиард триплет	На этапе разработки и запуска
IBM	Сущности и отношения, с котороми ассоциирована информация о свидетельствах, подтверждающих извлеченные факты	Различные размеры Доказанныхдокументов > 100 миллионов Отношений >5 миллиардов Сущностей > 100 миллионов	Активно используется в продуктах и клиентами

Неудивительно, что перечисленные выше системы графов знаний сталкиваются с проблемой *управления графами большого объема*. Проблемы этого измерения графов знаний рассматривались во множестве публикаций в академическом и исследовательском сообществе (например, задача устранения неоднозначности). Тем не менее, в промышленных условиях возникают новые вызовы. Управление масштабом - основная проблема, которая затрагивает операции, напрямую связанные с производительностью и рабочей нагрузкой. Проблема масштаба также проявляется косвенно, так как влияет на другие операции, например, управление быстрыми инкрементными обновлениями для крупномасштабных графов знаний в IBM, или управление согласованностью на большом развивающемся графе знаний в Google [23].

3. Графы знаний, описывающие сущности одного класса

Многие исследователи лишь частично соглашаются с определением графа знаний из [6]. В частности, у многих вызывает возражение тезис о том, что «графы знаний не должны ограничиваться только одной предметной областью". Во-первых, в последнее время наблюдается тенденция по созданию графов знаний, которые собирают информацию обо всех сущностях, принадлежащих одному классу. Например, Amazon и EBay создают графы знаний обо всех продуктах в мире, Google и Apple создали графы знаний обо всех локациях в мире, Центральный банк Италии создал граф знаний обо всех итальянских компаниях, ClaimsKG [31] извлекает утверждения из веб-страниц, занимающихся проверкой фактов, таких как политифакт, и связывает их с другими графами знаний, такими как DBpedia, что также позволяет находить связанные претензии, а EventKG [9] извлекает информацию о месте и времени всех сущностей, относящихся к классу Event (событие).

Во-вторых, многие приложения, созданные для определенных областей, таких как биология, могут считаться графами знаний, если они удовлетворяют другим требованиям графа. К таковым относятся, например, работы, посвященные автоматическому построению графов знаний из текстовых медицинских знаний и медицинских записей [28, 15, 27].

4. Граф знаний как однозначный граф с атрибуцией происхождения

В работе [19], приводится следующее определение графов знаний.

Граф знаний - это «граф, состоящий из множества утверждений (ребер, помеченных отношениями), которые выражаются между сущностями (вершинами графа), где смысл графа закодирован в его структуре, отношения и сущности однозначно идентифицированы, ограниченное множество отношений используется для меток ребер, и граф кодирует происхождение, особенно обоснование и атрибуцию этих утверждений».

В этом определении рассматривается несколько существенных аспектов современного понимания графа знаний.

4.1. Однозначная идентификация сущностей и ребер

Во-первых, авторы подчеркивают, что для того, чтобы утверждения графа знаний были недвусмысленными, они должны состоять из однозначных единиц. То есть, все объекты в

графе знаний, включая типы и отношения, должны быть идентифицированы при помощи глобальных идентификаторов с однозначным обозначением. Одним из примеров такого идентификатора является универсальный идентификатор ресурса (URI), используемый в RDF. К сожалению, требование однозначного идентификатора для каждой сущности связано с давно известной, но до сих пор не решенной проблемой *идентификации сущностей.*

В простейшей форме задача заключается в присвоении уникального нормализованного идентификатора и типа высказыванию или упоминанию объекта. Многие сущности, извлеченные из источников данных автоматически, имеют очень похожие поверхностные формы, например, люди с одинаковыми или похожими именами или фильмы, песни и книги с одинаковыми или похожими названиями. Без правильной привязки и устранения неоднозначности сущности будут неправильно ассоциироваться с неверными фактами и приводить к неверным выводам при последующей обработке.

В случаях, когда управление идентификацией должно выполняться с разнородной базой участников и в масштабе, проблема становится намного сложнее. Например, только в одной Википедии имеется двести Уиллов Смитов, а результат поиска движка Bing для актера Уилла Смита составляется из сто восемь тысяч фактов, взятых с сорока одного сайта. [23]. Кроме того, эффективная система идентификации сущностей также должна расти органично на основе постоянно меняющихся входных данных. Например, компании могут слиться или разделиться, а новые научные открытия могут разбить существующий объект на несколько частей. Если одна компания приобретает другую компанию, изменяется ли идентичность приобретающей компании?

Что касается «ограниченного набора типов отношений», то в контексте системы знаний открытого мира это требование надо понимать как множество основных отношений, которые истинны *независимо от контекста.* Примеры отношений, зависящие от контекста, часто встречаются в DBpedia. Например, в англоязычной DBpedia можно найти следующие две триплеты, касающиеся жен советского режиссера Сергея Бондарчука.

dbr:Sergei_Bondarchuk dbo:spouse dbr:Inna Makarova.

dbr:Sergei_Bondarchuk dbo:spouse dbr:Irina Skobtseva.

Поскольку в DBpedia не указано, что брак Бондарчука с Макаровой продолжался с 1948 по 1956 год, а на Скобцевой он был женат 1959 года вплоть до своей смерти, можно предположить, что у него было две жены одновременно. Чтобы данные утверждения не зависели от контекста, можно расширить эти два утверждения информацией о начале и конце каждого брака. Но описание такой информации требует дополнительных механизмов

таких, как реификация утверждений, либо использование именованных графов, или использование модели графов свойств.

В отличие от DBpedia, информация о браках Сергея Бондарчука отображена более корректно в графе знаний Wikidata. В этом наборе данных у каждого отношения *dbo:spouse* имеется дата начала брака, и дата окончания брака. Поэтому набор данных Wikidata является более качественным источником данных, чем DBpedia.

4.2. Кодирование происхождения утверждений

С точки зрения [19], граф знаний должен быть источником *достоверного* знания, а не просто набором некоторых утверждений. Поэтому каждое важное утверждение графа знаний должно сопровождаться дополнительной информацией о происхождении этого утверждения.

Заметим, что Консорциум World Wide Web разработал стандарт описания происхождения данных в Интернете (PROV) вместе с его представлением в виде онтологии PROV-O [16] и нано-публикаций [10].

Пример. Существуют приложения, где достоверность знаний, представленных в графе знаний, является критичной. Например, в работе [20] описан граф знаний, используемый для поиска новых лекарств против различных разновидностей меланомы, а также новых способов использования известных лекарств. Для решения этой задачи нужны достоверные знания относительно взаимодействия лекарств, белков и заболеваний. При построении графа знаний отбираются только те факты, достоверность которых превышает заданный порог. Поэтому все важные утверждения этого графа знаний снабжены информацией о происхождении утверждений в формате нано-публикации. На Рис. 3 показан пример утверждения о взаимодействии между двумя белками в виде нанопубликации. Представлены три графа. Граф утверждений (утверждение NanoPub 501799_Assertion) утверждает, что взаимодействие (X) имеет тип *sio: DirectInteraction,* имеет целью белок SLC4A8 и участником взаимодействия является белок CA2. Опорный граф (NanoPub 501799 Supporting), утверждает, что граф утверждений был создан в результате эксперимента с понижением (один из многих используемых типов закодированных экспериментов, подкласс *prov: Activity*. Граф атрибуции (NanoPub 501799 Attribution), в свою очередь, заявляет, что это утверждение имеет в качестве первоисточника публикацию (Loiselle et al., 2004), и что взаимодействие было процитировано из BioGrid.

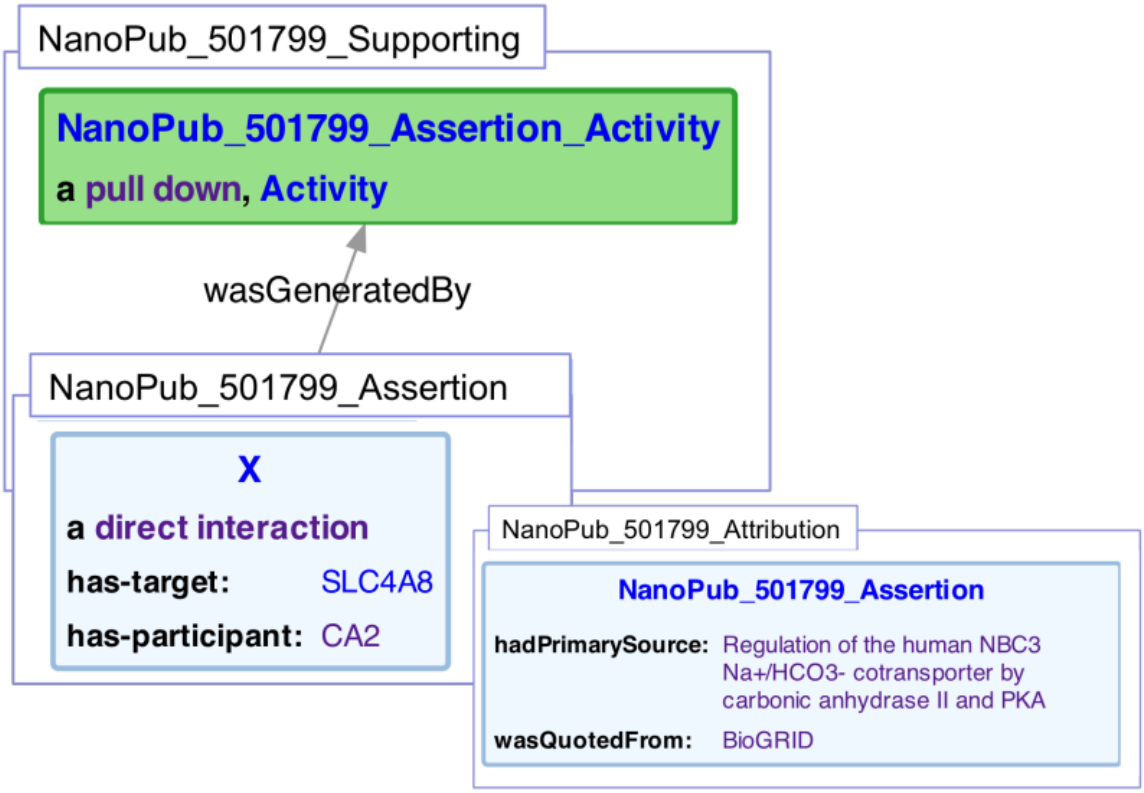

Рис. 3. Пример указания иформации о происхождении утверждения в формате нано-публикации [20].

Для подтверждения своей точки зрения, авторы создали Каталог Графов Знаний (Knowledge Graphs Catalog, KGC) (http: //graphs.whyis.io), в который вошли тридцать семь различных систем, в той или иной мере отвечающим требованиям графа знаний. В частности, системами, удовлетворяющими всем критериям, перечисленным в определении, (структурированный смысл, отсутствие неоднозначностей, отслеживание происхождения, ограниченные отношения (не зависящие от контекста) оказались UniProt KB, Gene Ontology, BioPortal, и сам Knowledge Graph Catalog. Известны и другие графы знаний, хранящие информацию о происхождении. В частности, Facebook [23] использует информацию о происхождении данных для поддержки корректности своего графа знаний. С точки зрения Facebook *корректность* не означает, что граф знаний всегда знает «правильное» значение атрибута, но скорее что всегда можно *объяснить*, *почему* было сделано определенное утверждение. Поэтому он сохраняет *происхождение* всех данных, которые проходят через систему, от сбора данных до слоя сервисов.

5. Графы знаний как системы, позволяющие получать новые знания

В настоящее время имеется значительное количество работ, которые считают, что специфической особенностью графов знаний является не только *способ представления знаний*, но и *способ получения новых знаний.* Так в работе [7] говорится, что специфической особенностью графов знаний, помимо требований большого объема данных, и интеграции множественных источников данных, является использование некоторого механизма порождения новых знаний. Поэтому дается следующее определение графа знаний:

Граф знаний собирает и интегрирует информацию в онтологию и применяет ризонер, чтобы получать новые знания.

В настоящее время имеется множество различных механизмов порождения этого нового знания, основными из которых являются логические методы, основанные на применении правил вывода [29] и статистические методы, основанные на так называемых *векторных вложениях* (embeddings) графов знаний [22], а также различные комбинации этих двух методов.

В сжатой форме подобное определение графа знаний формулируется следующим образом [5]:

«Граф данных, предназначенный для создания новых знаний».

Эти два определения рассматривают не только представление контента графа знаний, но и действия по формированию новых знаний, то есть способы управления графом знаний, что существенно расширяет границы определения понятия графа знаний. Поэтому в [2] представлено три разных взгляда на понятие графа знаний:

- как инструмента представления знаний, где основное внимание уделяется тому, как граф знаний используется для представления некоторой формы знаний;

- системы управления знаниями: основное внимание уделяется управлению системой графа знаний, аналогично тому, как системы управления базами данных играют эту роль для баз данных;

- сервисам приложений знаний: основное внимание уделяется обеспечению уровня приложений поверх графа знаний.

6. Данные, участвующие в создании нового знания

Как только мы признаем, что важной функцией графа знаний является создание новых знаний, важность поддержки *неявных* знаний становится центральным местом для графа

знаний, особенно когда они являются компонентом приложений ИИ предприятия до такой степени, что интенсиональное знание следует рассматривать как часть самого графа знаний.

Например, в финансовых приложениях корпоративных графов знаний множество регулирующих правил и правил функционирования конкретной финансовой области являются существенными. Аналогично, в приложениях логистики знание о том, как взаимодействуют определенные шаги в цепочке поставок, часто более важно, чем обычные данные, лежащие в основе цепочки поставок. В работе [3] утверждается, что «в современных системах, основанных на KG, необходимо учитывать и надлежащим образом обрабатывать богатое представление знаний, чтобы сбалансировать повышенную сложность со многими другими свойствами, включая удобство использования, масштабируемость, производительность и надежность приложения KG». По этой причине [3] предлагает следующее определение графа знаний.

«Полуструктурированная модель данных, состоящая из трех компонентов:

1) базовый экстенсиональный компонент, то есть набор реляционных конструкций для схемы и данных (которые можно эффективно смоделировать в виде графов или их обобщений);

2) интенсиональный компонент, то есть набор правил вывода над конструкциями экстенсиональной компоненты;

3) производный экстенсиональный компонент, который может быть создан в результате применения правил вывода к базовому экстенсиональному компоненту (так называемым процесс «вывода»).

В качестве примера к этому определению рассмотрим пример обнаружения новых связей в графе собственности компаний [1] Центрального банка Италии. Графы собственности компании являются центральными объектами корпоративной экономики и имеют большое значение для центральных банков, финансовых органов и национальных статистических управлений, чтобы решить актуальные проблемы в разных сферах: банковский надзор, оценка кредитоспособности, противодействие отмыванию денег, обнаружение страхового мошенничества, экономические и статистические исследования и многое другое.

Как показано на рисунке 4, на таких графах ключевым понятием является *отношение владения*: узлами являются компании и персоны (черные или синие узлы), а ребра соответствуют отношению собственности (черные сплошные ссылки) и помечены долей акций компании *y*, которыми владеет компания или персона *x*.

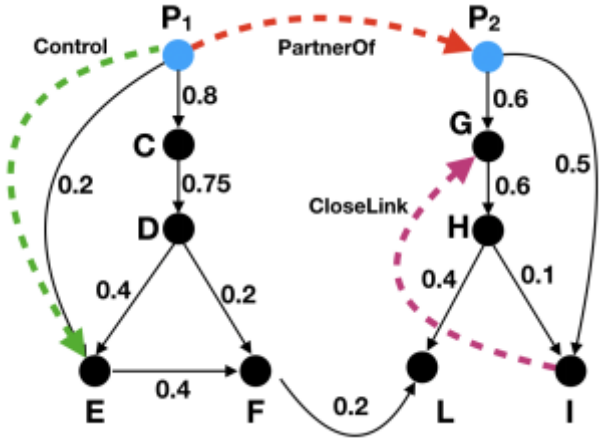

Рис. 4. Пример фрагмента из графа владения компаниями [34]. Ребра, показанные черным цветом, присутствуют изначально в качестве базового контента графа знаний. Ребра зеленого, красного и сиреневого цвета появляются в результате применения комбинированного метода вывода, использующего как логические правила, так и векторные вложения вершин [1].

1) При помощи графов компаний можно решить такую проблему как контроль над компанией. Рассмотрим граф на рисунке 4: Персона P1 управляет компаниями C, D (через C), E (поскольку она контролирует D, которому принадлежит 40% акций E, а также персона P1 напрямую владеет двадцатью процентами акций), и F (через E и D). Точно так же персона P2 контролирует всех своих потомков по графу компаний, кроме L. По-видимому, P1 также не контролирует L. То есть, ребро (P1, E) не принадлежит исходному графу владения компаниями, но эту информацию можно обнаружить при помощи процедуры обнаружения знаний.
2) Второе особенно важное приложение графов компании состоит в оценке риска предоставить конкретную ссуду компании *x*, которая обеспечена под залог, выданный другой «близко связанной» компанией *y*. Например, на рисунке 4, в результате процедуры создания нового знания появляется ребро (I, G), имеющее тип *CloseLink*, означающее, что компании G и I тесно связаны, поскольку персона P2 владеет более чем двадцатью процентами акций обеих компаний.
3) Помимо финансовых отношений, личные или семейные связи позволяют более широкое использование таких графов компаний: обнаружение семейного бизнеса или изучение реального распределения контроля. В примере на рисунке 4 знание того, что персоны P1 и P2 имеют личные связи - например, женаты - позволяет сделать вывод, что на самом деле P1 и P2 вместе управляют компанией L. Скорее всего, они

действуют как единый центр интересов: L на самом деле семейный бизнес, с контролем в руках одной семьи, а P1 и P2 вместе контролируют 60% этой компании.

Таким образом, в исходном графе компаний имеются только ребра, изображающие отношение владения компанией. Ребра, соответствующие отношениям *PartnerOf, Control, CloseLink* «создаются» в результате применения комбинированной процедуры созданий нового знания, комбинирующего применение правил логического вывода к фрагментам исходного графа, полученным на основе векторных вложений вершин графа. В соответствии с определением, данным выше, элементами графа знаний должны быть исходный граф знаний плюс правила вывода, использованные для обнаружения неизвестных связей между вершинами исходного графа, а также эти новые ребра, полученные в результате процедуры «вывода».

Представители фирмы IBM [23] идут еще дальше в понимании того, что должно храниться в графе знаний. Они считают, что *доказательства (или свидетельства*) должны быть примитивами по отношению к системе. Основное звено между реальным миром (которое разработчики часто пытаются моделировать) и структурами данных, содержащими *извлеченное* знание, – это «свидетельства» знания. “Свидетельствами могут быть необработанные документы, базы данных, словари или файлы изображений, текста и видео, из которых знания получены. Когда дело доходит до точных и полезных контекстных запросов во время процесса обнаружения, метаданные и другая связанная информация часто играют важную роль в выводе знаний. Таким образом, критически важно не потерять связь между отношениями, хранящимися в графе, и тем, откуда берутся эти отношения”.

Такое требование вполне сочетается со следующим определением графа знаний [4].

Графы знаний можно представить как сеть всех видов вещей, которые относятся к конкретной предметной области или организации. Они не ограничиваются абстрактными понятиями и отношениями, но могут также содержать экземпляры таких вещей, как документы и наборы данных ".

7. Заключение

В контексте больших промышленных графов знаний таких как графы знаний Microsoft, Google, Facebook, IBM наиболее важными критериями полезности графов знаний являются *покрытие, корректность и свежесть* контента.

Под покрытием понимается наличие в графе всей необходимой информации для предоставления наилучшего сервиса для пользователя. Поэтому ответ на этот вопрос всегда остается отрицательным, а разработчики графов знаний всегда стремятся расширить

множество источников знаний, чтобы увеличить качество покрытия. Заметим, что эта задача является дополнительной по отношению к задаче «завершения графа», которая решается для фиксированного множества источников данных.

Проблема *корректности* связана с вопросом, верна ли информация, предоставляемая пользователю? Действительно ли два источника информации относятся к одному и тому же факту, и что делать, если они противоречат друг другу? Ответы на эти вопросы остаются огромной областью исследования и инвестиций.

Понятие *свежести* контента связано с ответом на вопрос: «Обновлен ли контент»? Возможно, это было правильные когда-то, но устаревшие данные. Свежесть может быть разной для сущностей, которые меняется постоянно (цена акций) и редко меняющихся сущностей (столица страны или название фирмы), с множеством различных вариантов свежести между этими крайними точками. Например, граф знаний Facebook предназначен *для постоянного изменения*. Поэтому граф знаний Facebook - это не единое представление в базе данных, которое обновляется при появлении новой информации. Граф знаний Facebook строится с нуля, из исходников, каждый день, а система сборки идемпотентна и создает полный граф в конце сборки [23].

Таким образом, рассмотрев все требования к представлениям знаний в современных графах знаний, можно представить обобщенную картину жизненного цикла современного графа знаний, как процесс постоянного расширения внешний источников данных, алгоритмов, позволяющих создавать новые знания из этого расширяющегося множества источников данных, и, наконец, постоянно расширяющееся множество приложений, создаваемых над графами знаний. См. рисунок 5.

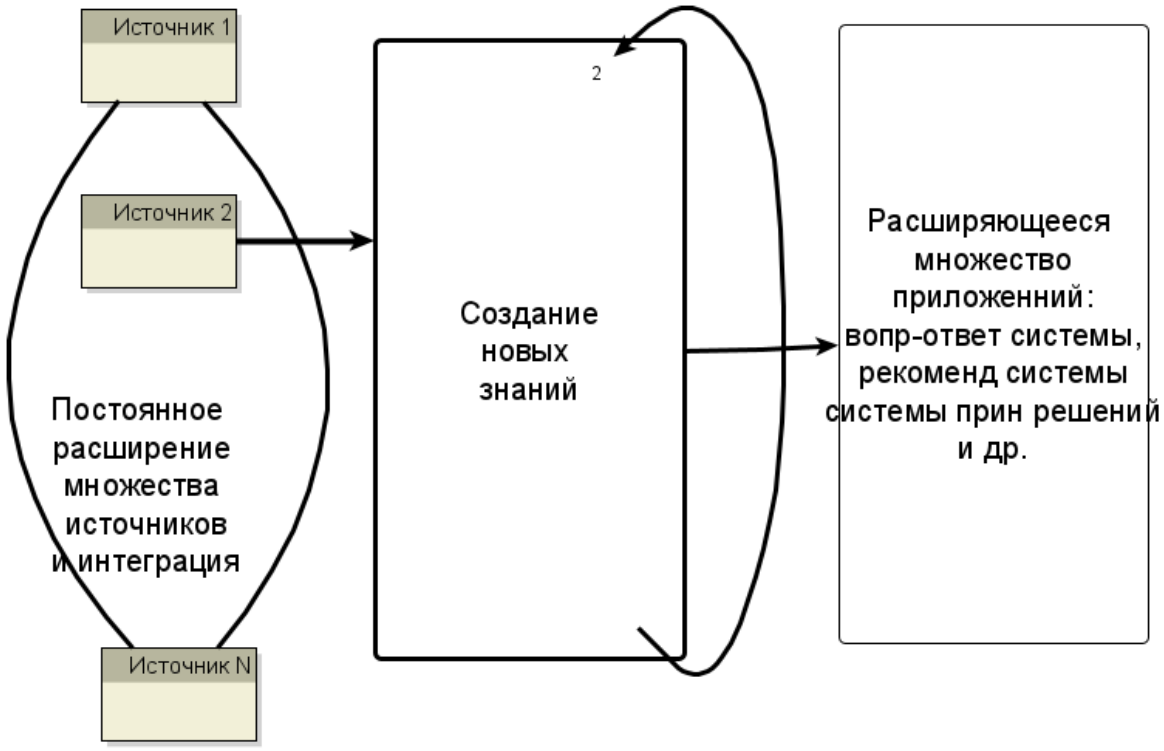

Рис. 5 Жизненный цикл графов знаний.

Если еще несколько лет назад основным приложением графов знаний считался семантический поиск, то сейчас в этот список входят вопросно-ответные системы, рекомендательные системы, системы принятия решений и многие другие. Более того, именно возможность обнаружения новых знаний и создания на их основе новых приложений является основным поводом для разработки корпоративных графов знаний.

Список литературы

1. Atzeni P., Bellomarini L., Iezzi M., Sallinger E., Vlad A. Weaving Enterprise Knowledge Graphs: The Case of Company Ownership Graphs https://openproceedings.org/2020/conf/edbt/paper_334.pdf
2. Bellomarini L., Sallinger E., Vahdati S. Chapter 2 Knowledge Graphs: The Layered Perspective/ / Janev V., Graux D., Jabeen H., Sallinger E. (eds) Knowledge Graphs and Big Data Processing. Lecture Notes in Computer Science. 2020. vol 12072. Springer, Cham. https://doi.org/10.1007/978-3-030-53199-7_2
3. Bellomarini, L., Fakhoury, D., Gottlob, G., Sallinger, E.: Knowledge graphs and enterprise AI: the promise of an enabling technology// 2019. IEEE 35th International Conference on Data Engineering (ICDE), P. 26–37.
4. Blumauer, A.: From taxonomies over ontologies to knowledge graphs (2014) https://semantic-web.com/2014/07/15/from-taxonomies-over-ontologies-to-knowledge-graphs/

5. Bonatti, P.A., Decker, S., Polleres, A., Presutti, V.: Knowledge graphs: new directions for knowledge representation on the semantic web (Dagstuhl Seminar 18371)// Dagstuhl Rep. 2019. Vol. 8, № 9. P. 29–111.
6. Carlson A., Betteridge J.,. Wang R. C, Hruschka E. R Jr, Mitchell T. M. Coupled semi-supervised learning for information extraction// *Proceedings of the third ACM international conference on Web search and data mining* 2010. P. 101–110.
7. Ehrlinger L., Woß W. Towards a definition of knowledge graphs//SEMANTiCS (Posters, Demos, SuCCESS), 2016. № 48.Ernst P., Siu A., Weikum G., KnowLife: a versatile approach for constructing a large knowledge graph for biomedical sciences, *BMC bioinformatics vol.* 2015. Vol. **16**, № 1. P. 157.
8. Färber M., Bartscherer F., Menne C., Rettinger A., Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago// *Semantic Web.* 2016, P. 1–53.
9. Gottschalk S, Demidova E. EventKG: A Multilingual Event-Centric Temporal Knowledge Graph https://arxiv.org/pdf/1804.04526.pdf
10. Groth P., Gibson A., Velterop J., "The anatomy of a nanopublication," *Information Services & Use*, 2010. vol. 30, N. 1-2, P. 51–56.
11. Heist N., Paulheim H.. Entity extraction from wikipedia list pages// *Extended Semantic Web Conference*, 2020.
12. Heist N., Hertling S., Ringler D., Paulheim H. Knowledge Graphs on the Web – an Overview https://arxiv.org/pdf/2003.00719.pdf 2020
13. Huang Z., Yang J., Harmelen F. van, Hu Q., Constructing disease-centric knowledge graphs: a case study for depression (short version), in: *Conference on Artificial Intelligence in Medicine in Europe*, Springer, 2017, P. 48–52.
14. James P.. Knowledge Graphs. // Linguistic Instruments in Knowledge Engineering, Elsevier Science Publishers B.V.,.1992. P. 97-117.
15. Lamurias A., Ferreira J.D., Clarke L.A., Couto F.M., Generating a Tolerogenic cell Therapy Knowledge graph from literature, *Frontiers in immunology* 2017. № 8, P. 1656.
16. Lebo T., Sahoo S., McGuinness D., Belhajjame K., Cheney J., Corsar D., Garijo D.,. Soiland-Reyes S, Zednik S., Zhao J., "PROV-O: The prov ontology," *W3C recommendation*, 2013.
17. Lehmann J., Isele R., Jakob M., Jentzsch A., Kontokostas D., Mendes P. N., Hellmann S., Morsey M., van Kleef P., Auer S., Bizer C.. DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia// *Semantic Web Journal*. 2013. Vol. 6. № 2.
18. Lenat D. B.. CYC: A large-scale investment in knowledge infrastructure *Communications of the ACM*, 1995. Vol. 38 №1. P. 33–38.

19. McCusker J. P., Erickson J. S., Chastain K., Rashid S., Weerawarana R., Bax M., McGuinness D. L. What is a knowledge graph http://www.semantic-web-journal.net/system/files/swj1954.pdf, 2018
20. McCusker J.P., Dumontier M., Yan R., He S., Dordick J.S., McGuinness D.LFinding melanoma drugs through a probabilistic knowledge graph.// PeerJ Computer Science . (2017) 3:e106 https://doi.org/10.7717/peerj-cs.106
21. Navigli R., Ponzetto S. P. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network// *Artificial Intelligence*. 2012. № 193 P. 217–250.
22. Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E. A review of relational machine learning for knowledge graphs.//IEEE . 2016. vol. 104, 1. № . P. 11-33.
23. Noy N., Gao Y., Jain A., Narayanan A., Patterson A., Taylor J.. Industry-scale knowledge graphs: Lessons and challenges//*Communications of the ACM* . 2019. Vol. 62. № 8. P. 36–43.
24. Nurdiati S., Hoede C. 25 Years Development of Knowledge Graph Theory: The Results and the Challenge, September 2008.
25. Pan J. Z., Vetere G., Gomez-Perez J. M., Wu H. Editors. Exploiting Linked Data and Knowledge Graphs in Large Organizations Springer, 2017.
26. Paulheim H.. Knowledge graph refinement: A survey of approaches and evaluation methods// *Semantic Web*. 2017, Vol. 8. №3 . P. 489–508,
27. Rotmensch M., Halpern Y., Tlimat A., Horng S., Sontag D., Learning a health knowledge graph from electronic medical records// *Scientific reports.* 2017. Vol. **7**, № 1. P. 5994.
28. Shi L., Li S., Yang X., Pan J. Qi, G., Zhou B., Semantic health knowledge graph: Semantic integration of heterogeneous medical knowledge and services, *BioMed Research International* 2017.
29. Stepanova D., Gad-Elrab M. H., Thinh V. Ho Rule Induction and Reasoning over Knowledge Graphs https://people.mpi-inf.mpg.de/~dstepano/conferences/RW2018/paper/RW2018paper.pdf
30. Suchanek F. M., Kasneci G., Weikum G.. YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia. // *16th international conference on World Wide Web*. 2007. P. 697–706,
31. Tchechmedjiev A., Fafalios P.,. Boland K, Gasquet M., Zloch M., ¨Zapilko B., Dietze S., Todorov K.. ClaimsKG: A knowledge graph of fact-checked claims. In International Semantic Web Conference, Springer, 2019. P. 309–324.
32. Tonon A., Felder V., Difallah D. E., Cudre-Mauroux P. Voldemortkg: ´Mapping schema. org and web entities to linked open data// *International Semantic Web Conference*, Springer, 2016. P. 220–228.
33. Vrandecic D., Krotzsch M.. Wikidata: a Free Collaborative Knowledge Base.// *Communications of the ACM* 2014. Vol. 57, № 10. P. 78–85.