

УДК 33

DOI 10.34670/AR.2019.90.8.046

## Диалоговая система на основе динамического представления графов знаний

**Романов Алексей Андреевич**

Преподаватель,  
Санкт-Петербургский национальный исследовательский  
университет информационных технологий, механики и оптики,  
197101, Российская Федерация, Санкт-Петербург, просп. Кронверкский, 49;  
e-mail: romanov@itmo.ru

### Аннотация

Стремительный рост объемов информации и развитие методов ее обработки порождают развитие интерфейсов взаимодействия с данными. Одним из таких примеров является диалоговая система, основанная на интеллектуальной обработке данных. Диалоговые системы все больше используются в повседневной жизни, от голосовых ассистентов в мобильных телефонах, до целевых систем заказа билетов в кино, или помощников в банковских приложениях. Важным аспектом построения таких систем являются данные, а как известно, данные необходимо представить в некоторой форме, понятной для машины. В статье представлен обзор подходов и описание метода построения диалоговых систем на основе динамического представления графов знаний.

### Для цитирования в научных исследованиях

Романов А.А. Диалоговая система на основе динамического представления графов знаний // Экономика: вчера, сегодня, завтра. 2019. Том 9. № 8А. С. 453-461. DOI 10.34670/AR.2019.90.8.046

### Ключевые слова

Граф знаний, диалоговая система, векторные пространства.

## Введение

Современные образовательные подходы заключаются в применении онлайн или смешанных образовательных технологий [Кравченко, 2014; Нагаева, 2016]. Этому процессу способствует стремительное развитие информационно-коммуникационных технологий (ИКТ), которые обеспечивают как доступ к учебным материалам, так и различные формы контроля [Батьковский, 2016]. Кроме того, все больше высших учебных заведений реализуют свои онлайн-курсы в общем открытом пространстве, на платформах вроде Coursera, edX и «Открытое образование», чему способствует национальная программа «Цифровая экономика Российской Федерации».

Развитие и интеграция смешанных и онлайн-курсов в учебный процесс приводит к его изменению, лекция заменяются заранее записанными видеоматериалами, практические работы – различными типами онлайн упражнений, лабораторные – виртуальными установками. Роль преподавателя также меняется, от непосредственного носителя знаний, к наставнику, обеспечивающего учебный процесс. Появляются и новые профессии, такие как администраторы, программисты, разработчики онлайн-курсов, видео и аудио режиссеры, кураторы и специалисты поддержки курсов. При этом все они должны обладать соответствующими знаниями из самой дисциплины. Поэтому зачастую, все указанные роли, это преподаватели [Елизарьева, 2016].

Рост числа сотрудников, обеспечивающих учебный процесс, неминуемо приводит к увеличению затрат на поддержку такого процесса и исследованию вопросов, связанных с его оптимизацией. Важным аспектом здесь является работа с данными, это данные самих курсов, лекции, задания, форуме поддержки курса, запросы по почте. Решением может служить обобщение всей указанной информации в рамках единой информационной образовательной среды, но не в основе классических баз данных или документов, а в связанном виде, доступном для последующей обработки с помощью автоматизированных средств. Такой формой может служить семантическая сеть, которая в отличие от баз данных не только хранит данные, но и позволяет делать автоматические умозаключения, получать новые, ранее неизвестные связи между объектами. Такие возможности важны, при построении интеллектуальных систем, а в частности диалоговых систем.

Диалоговые системы переживают новый виток развития и находят все больше сфер применения, это и голосовые ассистенты, и интерфейсы для заказа билетов в кино, надоедливые онлайн-консультанты в интернет-магазинах, а также службы поддержки во многих банковских приложениях. Однако их обобщает единая цель, автоматизация рутинных вопросов. В перспективе диалоговая система позволяют экономить деньги, за счет автоматизации таких повторяющихся задач.

В образовательном процессе нового поколения, таких задач также достаточно. Так многие вопросы пользователей по материалам, срокам сдачи, техническим вопросам повторяются, а значит, их можно формализовать и автоматизировать процесс. Неоспорим преимуществом диалоговых систем, является их доступность для пользователя в любое время. Кроме того, подходы к построению таких систем на основе семантических сетей, позволяют накапливать опыт, анализировать разговоры и тем самым постоянно улучшать работу системы.

## Диалоговые системы

Основной проблемой и сложностью создания диалоговых систем, являются данные, а точнее проблема представления данных в таком виде, чтобы диалоговая система непрерывно имела к ним доступ, и могла обрабатывать запросы на естественном языке от пользователя. Кроме того, важна и сама архитектура диалоговой системы, и понимание того, какие цели она решает. Классически, диалоговые системы делятся на два вида:

- 1) целевые — помогают пользователям достичь заранее определенной цели (например, заказать билет в кино) в закрытой области знаний. Такие системы обрабатывают сообщение на естественном языке и предпринимают некоторые действия в соответствии с правилами и текущим состоянием диалога. Хотя понимание языка обрабатывается статистическими моделями, большинство диалоговых систем все еще используют созданные вручную правила обработки. Это не только делает затратным и трудоемким развертывание реальной диалоговой системы, но также ограничивает ее использование в других сферах.
- 2) нецелевые — служат для имитации разговора с реальным собеседником и могут быть разделены на два вида:
  - порождающие или генеративные — это диалоговые системы на основе больших корпусов языка с использованием генеративных моделей, которые генерируют ответы автономно, слово за словом.
  - поисковые — основаны на поиске информации и предположении, что соответствующий ответ на запрос пользователя содержится в составленных наборах данных. Такие наборы собираются из переписок с пользователями на форумах, по почте и из других средств общения.

Целевые системы просты в реализации, так как заранее известны действия, которые пользователь будет выполнять, нет необходимости генерировать ответы на естественном языке, а достаточно лишь отображать действия, выполненные системой, например, факт бронирования билета в кино. В случае с нецелевыми системами задачу усложняется, ведь с одной стороны необходимо имитировать диалог, составлять предложения, с другой стороны необходимо подкреплять их фактами из существующих источников данных. В рамках образовательного процесса мы сталкиваемся с комбинацией различных подходов, так, с одной стороны, необходимо предоставить четкие знания, если студент запрашивает, к примеру некоторое определение для неясного термина. С другой стороны, вопрос мог быть уже задан ранее другим пользователем, например по почте или на форуме, что порождает необходимость поиска информации в ресурсах поддержки. В целом вопросы можно классифицировать следующим образом:

- 1) фактологические или закрытые вопросы, которые фиксируют уже свершившееся действие, указывают на наличие какого-то общемирового факта или знания. Обычно такие вопросы имеют стандартные ответы, могут быть оценены как правильные или неправильные и подразумевают короткий ответ. Например:
  - Сколько граней у додекаэдра? (Ответом является числовое значение)
  - Из каких фигур состоит развертка цилиндра? (Ответом является список)
  - Что такое биткойн? (Ответом является определение)
  - Почему для распознавания изображений используются нейронные сети? (Ответом является объяснение)

2) Интерпретационные или открытые вопросы, которые обычно не имеют определенных достоверных ответов.

- Что случится если черный кот дорогу перейдет? (Гипотетически)
- Как собрать компьютер? (Ожидаем получить последовательность действий)
- Почему на этот компьютер такая маленькая гарантия?

Построение диалоговых систем для работы с фактологическими вопросами проще, так как результат легко оценить качественными метриками, например точность (precision), полнота (recall), F-мера. А ответы на такие вопросы можно автоматически извлечь из существующих источников данных. С открытыми вопросами сложнее, как правило для оценки ответов приходится спрашивать мнения пользователя, на основе которого ответы ранжируются. Ответы на вопросы обычно получают с нуля, от экспертов или агрегируют, например из учебных материалов. Важным аспектом является агрегирование и создание наборов данных, на основе которых и будут порождаться ответы. Далее рассмотрим предложенный метод на основе графов-знаний и их векторных представлений.

### Графы знаний

Современной формой представления и хранения знаний в сети являются графы знаний — структуры, состоящие из уникальных сущностей (узлов графа) и связей между ними (ребер графа) [5]. Сущностями могут быть как материальные вещи, так и абстрактные концепции. Например, “Университет ИТМО” как материальный концепт и “Университет” как собирательное описание всех университетов. Связи (ребра графа) описывают отношения между сущностями и их атрибуты (также называемые свойствами). Например, “дата основания” — это атрибут для сущности “Университет ИТМО” и может содержать значение “26.03.1900”, тогда как для абстрактного концепта “Университет” атрибут “дата основания” может содержать “некоторую условную дату” и создать высказывание, что каждый университет имеет дату основания. Таким образом, графы знаний позволяют моделировать как абстрактные логические высказывания и схемы, так и наполнять эти схемы конкретными объектами реального мира. Более того, графы знаний позволяют машинам производить рассуждения и выводить новые знания, ранее не описанные в графе. Формальная полуструктурированность и мощный логический аппарат отличают графы знаний от традиционных реляционных баз данных, которые являются структурированными, то есть обладают установленными связями и отношениями.

- Исследование графов знаний стоит на стыке многих областей компьютерных наук:
- информационный поиск (information retrieval), позволяющий ускорить наполнение графа из различных источников;
- обработка естественных языков (natural language processing) и семантические технологии (semantic technologies), которые позволяют описывать и использовать при анализе смысл хранящихся знаний;
- системы управления базами данных (data management), обеспечивающие эффективное хранение графов;
- машинное обучение (machine learning), необходимое для анализа содержащихся в исходных данных знаний и генерации новых знаний;
- искусственный интеллект в целом (artificial intelligence) как общее теоретико-методологическое обобщение, изучающее возможности машинного интеллекта.

Граф знаний является лишь структурой и данными, представленными в доступном для обработки машиной формате, однако это лишь объекты и связи, а любые методы обработки основаны на оперировании числами. Именно по этой причине необходимо представить граф знаний в виде точек пространства.

### Векторные пространства

Самым простым для понимания является представление графа знаний в векторном пространстве таким образом, чтобы тройки  $(h, r, t)$  примерно удовлетворяли соотношению  $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ , где жирным буквами обозначены векторные представления [Bordes et al, 2013]. Для троек, которых нет в графе знаний такое соотношение не должно выполняться. При этом  $\mathbf{h}, \mathbf{r}, \mathbf{t}$  задаются координатами в некотором пространстве размерности  $\mathbb{R}^k$ , где  $k$  — гиперпараметр модели. Поиск таких координат векторов достигается с помощью градиентного спуска при минимизации функции потерь:

$$\mathcal{L} = \sum_{(h,r,t) \in \mathcal{S}} \sum_{(h',r,t') \in \mathcal{S}'_{(h,r,t)}} [\gamma + d(\mathbf{h} + \mathbf{r}, \mathbf{t}) - d(\mathbf{h}' + \mathbf{r}, \mathbf{t}')]_+,$$

где  $d$  — некоторая мера схожести, для представления векторов используется норма  $r_1$  (манхэттенское расстояние) обозначаемая  $\|x\|_1$  или евклидова норма  $r_2$  обозначаемая  $\|x\|_2$ ,  $[x]_+$  —обозначает что рассматриваются только положительные слагаемые,  $\gamma > 0$  — гиперпараметр отступа.

Визуализация такого процесса на плоскости будет выглядеть следующим образом: от эпохи 1 до эпохи 3 расстояние между схожими сущностями уменьшилось (писатели), а сущность «Фэнтези» удалилась на большее расстоянии от них. При этом можно наблюдать некое несоответствие, вектор жанр от Клайва Льюиса не попадает в сущность Фэнтези. За это и отвечает гиперпараметр гамма функции потерь.

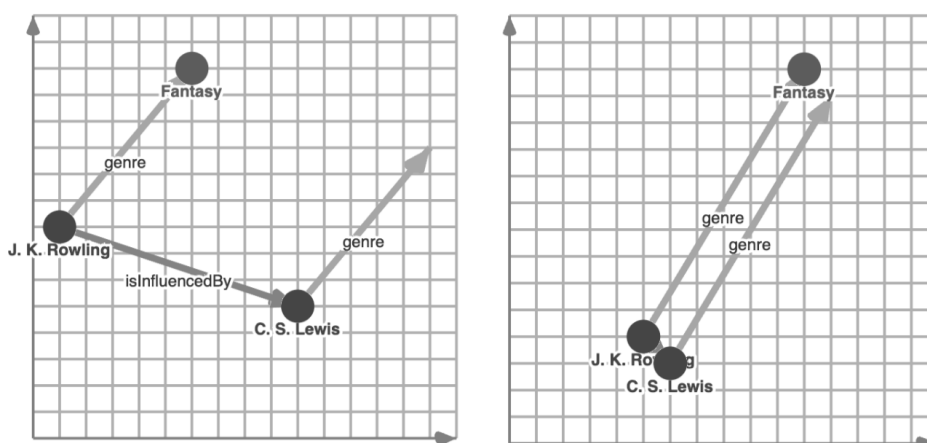


Рисунок 1 - Визуализация такого процесса на плоскости

В результате такого обучения модели, и отображения всего графа знаний в некотором пространстве, мы можем предсказать новые связи между сущностями.

Например, если мы хотим предсказать в каком жанре писал Ллойд Александер, используя полученные значения для вектора “жанр” мы с большей долей вероятности выберем жанр Фэнтези, так как он ближе, чем жанр Хоррор.

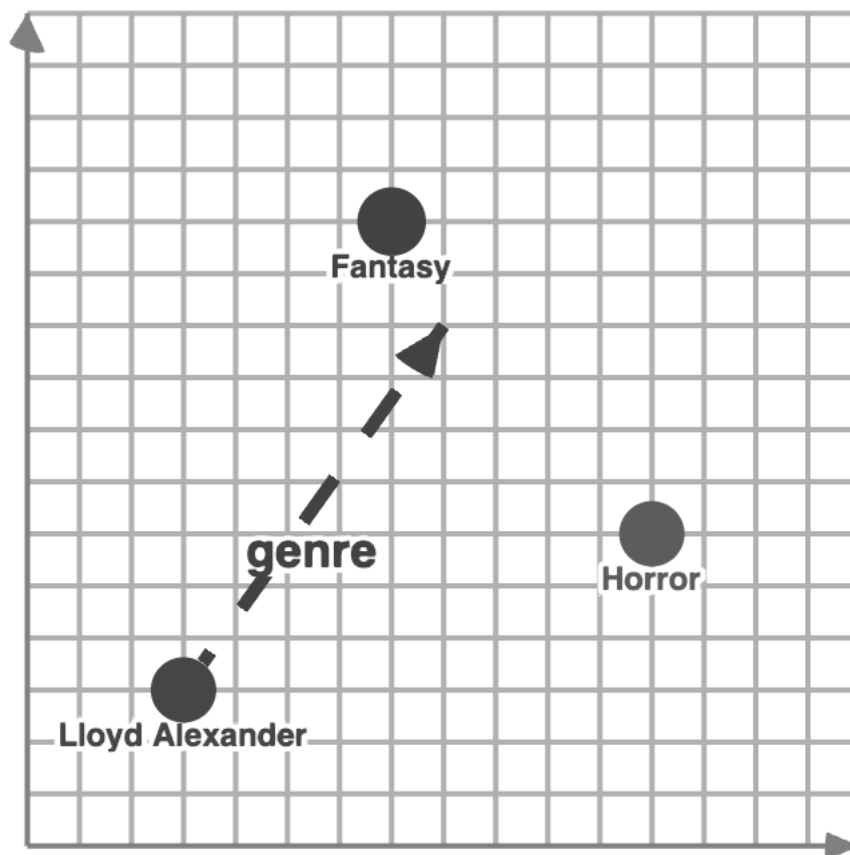


Рисунок 2 – Выбор жанра, в котором писал Ллойд Александер

### Метод

В основу диалоговой системы заложен ряд модулей. Первый — модуль кореференций, отвечающий за поиск ответов исходя из истории текущего диалога или истории прошлых диалогов, представленных в графе знаний. Модуль позволяет связать следующий диалог:

- Где расположен Университет ИТМО?
- В Санкт-Петербурге.
- А МГУ?

И тем самым понять смысл последнего вопроса, с пропущенным ключевым словом «где расположен». Если модуль не дал никакого ответа, то стоит поискать результат по именованным сущностям. В указанном примере таковыми, конечно, являются: «Университет ИТМО», «Санкт-Петербург» и «МГУ». Извлечения именованных сущностей не достаточно, необходимо их связать с графом знаний, как локальным (основанном на учебных материалах или обращениях в поддержку), но и внешних, так как в локальном попросту может не оказаться таких сущностей, а связи построить необходимо. Для этого используется граф знаний Wikidata, содержащий около 50 млн. сущностей с синонимичными связями на сотнях языков. Кроме того,

если необходимо, всегда можно подключить и другие внешние источники по специализированным тематикам. Так из нашего первого вопроса, может быть определена сущность «Университет ИТМО» со следующим идентификатором <https://www.wikidata.org/wiki/Q1342013> в графе знаний Wikidata.

По идентификатору можно узнать все существующие связи в графе знаний, однако возникает необходимость связать существующие связи, с той, что мы хотим получить из контекста вопроса. А именно связать строку «где расположен» со связью «headquarters location» («расположение штаб-квартиры»), именно так она представлена в графе знаний Wikidata. За поиск таких связей отвечает соответствующий модуль поиска связей. Он использует все имеющиеся связи и их описание в графе знаний <https://www.wikidata.org/wiki/Property:P159>, а затем ранжирует их в соответствии с компонентами запроса и синонимами. В результате чего возможно перейти по связи «расположение штаб-квартиры» и получить ответ «Санкт-Петербург». Но любой граф знаний не может быть полным и иметь все возможные связи, зачастую они будут отсутствовать. В таком случае должна быть построена некоторая вероятностная модель, связывающая субъект и отношение. В этом и помогают рассмотренные векторные представления. Ведь в них все сущности типа «Университет» и все сущности «Город» будут расположены в своих сконцентрированных областях, тогда если хотя бы у одной пары будет связь, то на основе суммарного значения  $\mathbf{h} + \mathbf{r}$  можно определить наиболее похожие объекты за счет kNN кластеризации.

Рассмотренный метод показал свою эффективность, однако встречаются ситуации, результат которых можно улучшить, если учитывать некоторую метаинформацию, например время и программа обучения. А ответ на вопрос диалоговой системой может зависеть от семестра, или курса. В рассматриваемом методе используется временной подход, что позволяет отслеживать изменения в связях вопросно-ответных конструкций обращений поддержки. Для достижения такого подхода применяется динамическая архитектура на основе Skip-gram, метода, где на основе текущего слова предугадывается его окружение, но в отличие от классического подхода, используется также байесовская модель (Barkan, 2017) и временные ряды.

## Заключение

В результате разработан метод на основе динамического представления графов знаний, что повышает общую точность предсказаний при построении связей и автоматического вывода знаний. Подход может быть расширен для других метапризнаков.

## Библиография

1. Кравченко Г. В. Использование модели смешанного обучения в системе высшего образования // Известия Алтайского государственного университета. – 2014. – Т. 1. – №. 2 (82).
2. Нагаева И. А. Смешанное обучение в современном образовательном процессе: необходимость и возможности // Отечественная и зарубежная педагогика. – 2016. – №. 6 (33).
3. Батьковский А. М. Развитие информационно-коммуникационных технологий // Инновационная наука: прошлое, настоящее, будущее. Сборник статей Международной научно-практической конференции. – 2016. – №. 5. – С. 66-68.
4. Елизарьева Ю. А. Современный преподаватель в процессе «Моокизации» образования // Гуманитарная информатика. – 2016. – №. 10.
5. Лещева И. А. Практический подход к созданию и наполнению баз знаний онтологического типа: проблемы и вызовы // Сборник научных трудов XXI Российской конференции «Инжиниринг предприятий и управление знаниями». – РЭУ им. ГВ Плеханова, 2018. – С. 194-201.

6. Bordes A. et al. Translating embeddings for modeling multi-relational data //Advances in neural information processing systems. – 2013. – С. 2787-2795.
7. Школьник И.С. Форсайт технология оценки технологического цикла инфокоммуникаций // Экономика: вчера, сегодня, завтра. 2018. Том 8. № 12А. С. 140-146.
8. Школьник И.С. Институциональные условия диффузии инноваций в области телекоммуникаций: на примере Китая // Инновации и инвестиции. 2019. № 1. С. 25-28.
9. Школьник И.С. Ограничения внедрения концепции открытых инноваций в российской экономике. В книге: Новое в науке и образовании Тезисы докладов Международной ежегодной научно-практической конференции Еврейского университета. Ответственный редактор Ю.Н. Кондракова. 2019. С. 46-47.
10. Школьник И.С. Адаптация парадигмы открытых инноваций в деятельности телекоммуникационных компаний // Экономика: вчера, сегодня, завтра. 2018. Том 8. № 10А. С. 333-341.

## Dialog system based on dynamic representation of knowledge graphs

**Aleksei A. Romanov**

Lecturer,

Saint Petersburg National Research University  
of Information Technologies, Mechanics and Optics,  
197101, 49, Kronverksky av., Saint Petersburg, Russian Federation;  
e-mail: romanov@itmo.ru

### Abstract

The rapid growth of information volumes and the development of methods of its processing give rise to the development of interfaces of interaction with data. One such example is a dialog system based on intelligent data processing. Dialog systems are increasingly used in everyday life, from voice assistants in mobile phones, to targeted movie ticket systems, or assistants in banking applications. An important aspect of building such systems is data, and as you know, the data must be presented in some form that is understandable for the machine. The article presents an overview of approaches and a description of the method of constructing dialog systems based on the dynamic representation of knowledge graphs.

### For citation

Romanov A.A. (2019) Dialogovaya sistema na osnove dinamicheskogo predstavleniya grafov znaniy [Dialog system based on dynamic representation of knowledge graphs]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 9 (8A), pp. 453-461. DOI 10.34670/AR.2019.90.8.046

### Keywords

Knowledge graph, dialog system, vector spaces.

### References

1. Kravchenko, G. V. (2014) the Use of models of blended learning in higher education //news of Altai state University. Vol. 1. - №. 2 (82).
2. Nagaeva I. A. (2016) Mixed learning in the modern educational process: the need and opportunities // Domestic and foreign pedagogy. №. 6 (33).
3. Batkovskij M. A. (2016) the Development of information and communication technologies //Innovative science: past, present, and future. Collection of articles of the International scientific and practical conference. №. 5. Pp. 66-68.



4. Elizarieva Y. A. (2016) Modern teacher in the process of "Booksale" education //Humanitarian Informatics. № 10.
5. Leshcheva I. A. (2018) Practical approach to creation and filling of knowledge bases of ontological type: problems and challenges // Collection of scientific papers of the XXI Russian conference "engineering of enterprises and knowledge management". - REU im. GV Plekhanov, Pp. 194-201.
6. A. Bordes et al. (2013) Translating embeddings for modeling multi-relational data //Advances in neural information processing systems. Pp. 2787-2795.
7. Shkol'nik I.S. (2018) Adaptatsiya paradigmy otkrytykh innovatsiy v deyatel'nosti telekommunikatsionnykh kompaniy [Adaptation of the paradigm of open innovation in the activities of telecommunications companies]. Ekonomika: vchera, segodnya, zavtra [Economics: Yesterday, Today and Tomorrow], 8 (10A), pp. 333-341.
8. Shkol'nik I.S. (2018) Forsayt tekhnologiya otsenki tekhnologicheskogo tsikla infokommu-nikatsiy [Foresight technology assessment of the technological cycle of infocommunications]. Ekonomika: vchera, segodnya, zavtra [Economics: Yesterday, Today and Tomorrow], 8 (12A), pp. 140-146.
9. Shkolnik I. S. (2019) Institutional conditions of diffusion of innovations in the field of telecommunications: the case of China. Innovations and investments. No. 1. pp. 25-28.
10. Shkolnik I. S. (2019) Limitations of implementation of the concept of open innovations in the Russian economy. In the book: New in science and education Abstracts of the international annual scientific-practical conference of the Hebrew University. Responsible editor Yu. N. Every. 2019. 46-47.