Научная статья

УДК 902.694
DOI 10.25205/1818-7900-2023-21-3-46-55

Методы создания семантически ориентированных интеллектуальных помощников

Артем Сергеевич Трегубов[1], Иван Сергеевич Немцев[2], Анна Александровна Котельникова[3], Дарья Андреевна Доможакова[4]

Новосибирский государственный университет
Новосибирск, Россия

[1]a.tregubov@g.nsu.ru
[2]i.nemtsev@g.nsu.ru
[3]a.kotelnikova@g.nsu.ru
[4]d.domozhakova@g.nsu.ru

Аннотация

В настоящий момент для создания интеллектуальных помощников используются различные технологии, основанные как на исследованиях в области нейронных сетей, средств анализа текстов на естественном языке, так и на использовании средств семантического моделирования. Каждый из этих подходов позволяет качественно решать отдельно взятые задачи. В рамках данной работы разрабатывается интеллектуальный помощник, объединяющий в себе все указанные подходы. Цель работы – создание интеллектуального помощника, выполняющего функции виртуального консультанта по процессам работы организации. Разработка основана на использовании семантической модели организации и бизнес-процессов. Для распознавания пользовательских намерений мы используем гомоморфные и генерализованные пользовательские намерения. Система позволяет осуществлять декомпозицию пользовательских задач и формировать последовательность их выполнения на основе семантических моделей пользователя и предметной области.

Ключевые слова

интеллектуальный помощник, онтология, машинное обучение, обработка текстов на естественном языке, распознавание намерений

Для цитирования

Трегубов А. С., Немцев И. С., Котельникова А. А., Доможакова Д. А. Методы создания семантически ориентированных интеллектуальных помощников // Вестник НГУ. Серия: Информационные технологии. 2023. Т. 21, № 3. С. 46–55. DOI 10.25205/1818-7900-2023-21-3-46-55

Methods for Developing Semantically Oriented Virtual Assistants

Artem S. Tregrubov[1], Ivan S. Nemtsev[2], Anna A. Kotelnikova[3], Darya A. Domozhakova[4]

Novosibirsk State University
Novosibirsk, Russian Federation

[1]a.tregubov@g.nsu.ru
[2]i.nemtsev@g.nsu.ru
[3]a.kotelnikova@g.nsu.ru
[4]d.domozhakova@g.nsu.ru

Abstract

Nowadays, various technologies are used to create intelligent assistants, based both on research in the field of neural networks, natural language text analysis tools, and on the use of semantic modeling tools. Each of these approaches allows you to qualitatively solve certain problems. As part of this work, an intelligent assistant is being developed that combines all these approaches. The purpose of the work is to create an intelligent assistant that performs as a virtual consultant on the organization's work processes. The development is based on the use of the semantic model of the organization and business processes. To recognize user intents, we use homomorphic and generalized user intents. The system allows decomposing user tasks and creating a consistency of their execution based on the user semantic models and the subject area.

Keywords

Intelligent assistant, ontology, machine learning, natural language processing, intent recognition

For citation

Tregubov A. S., Nemtsev I. S., Kotelnikova A. A., Domozhakova D. A. Methods for creating semantically oriented intelligent assistants. *Vestnik NSU. Series: Information technologies*. 2023, vol. 21, no. 3. P. 46–55. DOI 10.25205/1818 7900-2023-21-3-46-55

Введение

При попадании в незнакомую среду люди часто испытывают психологический дискомфорт и трудности, связанные с неопределенностью и неизвестностью. Новая предметная область кажется для них непонятной, а для некоторых даже враждебной. Решением данной проблемы могли бы выступать виртуальные ассистенты, содержащие знания о данной предметной области. Для формализации знаний хорошо подходят средства семантического и онтологического моделирования, с помощью которых можно задать правила и принципы устройства такой предметной области.

Задача создания интеллектуального помощника является достаточно сложной, объединяющей в себе достижения в различных областях научного знания. В частности, для создания интеллектуальных помощников используются исследования в области нейронных сетей, средств анализа текстов на естественном языке, средств семантического моделирования и так далее, что подчеркивает актуальность выбранной темы.

В рамках данной работы предполагается создание интеллектуального помощника, выполняющего функции электронного консультанта по процессам работы организации. Предполагается создание универсальной системы, способной настраиваться под любую предметную область.

Цель работы – создание интеллектуального помощника с использованием средств семантического моделирования для описания бизнес-процессов и структуры организации.

Для достижения данной цели были поставлены следующие задачи.

1. Разработать способ анализа входящих сообщений для распознавания намерений пользователя.

2. Разработать механизм управления диалогом с пользователем.
3. Разработать семантическую модель структуры организации.
4. Разработать семантическую модель бизнес-процессов организации.
5. Разработать прототип интеллектуального помощника.

Разрабатываемые семантические модели и методы работы с намерениями являются ключевыми компонентами создаваемого в рамках данной работы интеллектуального помощника.

1. Анализ существующих подходов и решений

Распознавание пользовательских намерений играет ключевую роль в вопросе создания интеллектуальных помощников. Для качественного решения данного вопроса требуется прежде всего решение задачи обработки текста на естественном языке [1]. Продвижение в данном вопросе позволит создать голосовой интерфейс для взаимодействия человека и компьютера. Но стоит отметить, что это не является ключевой задачей исследования.

Далее для работы интеллектуального помощника было необходимо выполнить исследование существующих подходов и решений.

В результате данного исследования было обнаружено, что наиболее близким к данной задаче решением является решение от консорциума W3C под названием «The Organizational Ontology» [2]. Данное решение представляет собой онтологию, предназначенную для публикации информации об организациях и организационных структурах, включая правительственные организации.

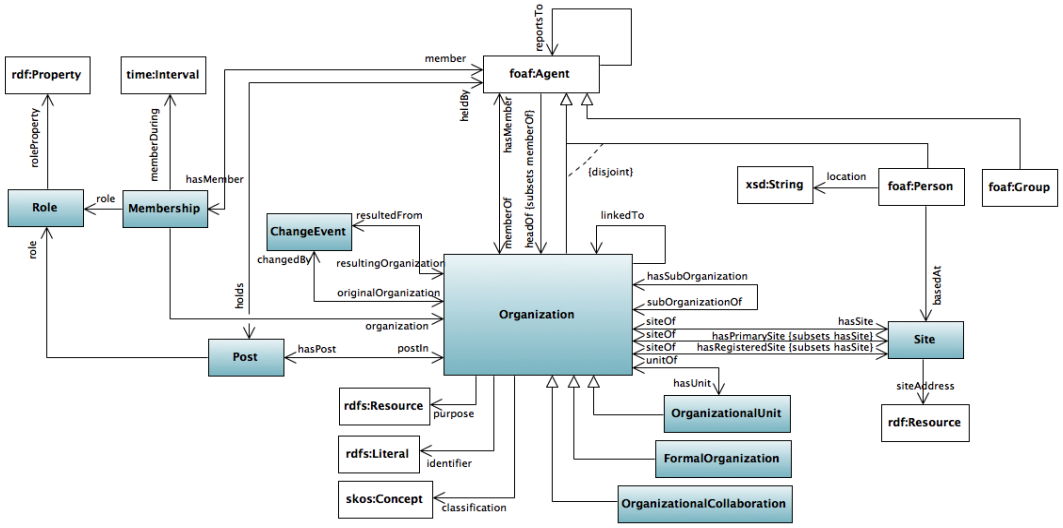

Рис. 1. Схема онтологии The Org от W3C
Fig. 1. Ontology scheme The Org by W3C

Решение, представляющее собой компонент отечественной системы по управлению бизнес-процессами ELMA BPM под названием «Моделирование оргструктуры». Данный инструмент позволяет описать как модель структуры организации, так и ее бизнес процессы. Данное решение является наиболее интересным. Но, к сожалению, не позволяет интегрировать его для интеллектуального помощника, а также не позволяет использовать механизмы логического вывода, позволяющие значительно упростить процесс описания модели.

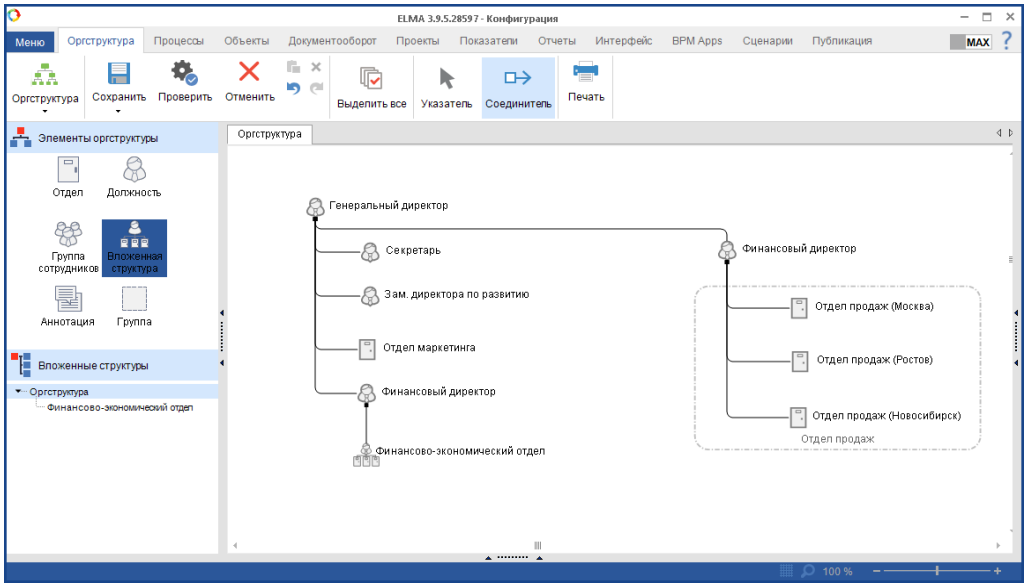

Рис. 2. Пример построения модели структуры организации
в программном решении ELMA
Fig. 2. Example of building an organization structure model
in the ELMA software solution

Рассмотренные аналоги имеют ряд существенных недостатков. Так, в большинстве описанных решений отсутствуют подобия таких понятий, как «поддерживаемые операции» и «бизнес-правила», которые позволили бы строить ответы на вопросы о правилах работы отдельных элементов организационной структуры. Некоторые аналоги делают упор на процессной части моделирования организаций, а продукт от W3C, кроме всего прочего, обладает рядом лишних параметров, представляющих собой накладные расходы. Перечисленные недостатки данных решений затрудняют или делают невозможным их использование в разрабатываемой системе.

На основе анализа было предложено решение, использующее лучшие практики существующих решений [3; 4], дополненное средствами семантического моделирования.

2. Предлагаемое решение

2.1. Модель структуры организации

В рамках работы была разработана модель структуры организации, обладающая универсальностью в отношении моделирования организаций и поддерживающая такие понятия, как «бизнес-правила» и «поддерживаемые операции».

Ключевым звеном онтологии является OWL-класс под названием «Workspace». Данный класс представляет собой единицу организационной структуры и позволяет отражать такие сущности моделируемой предметной области, как «Организация», «Отдел», «Должность», «Человек» и «Место». Для этого экземпляры класса «Workspace» связываются отношением «workspaceType» с экземпляром одного из приведенных в онтологии классов, отражающих тип сущности предметной области. В свою очередь, для того чтобы экземпляр класса «Workspace» не являлся одновременно, например, и Местом, и Человеком, была установлена кардинальность отношения «workspaceType».

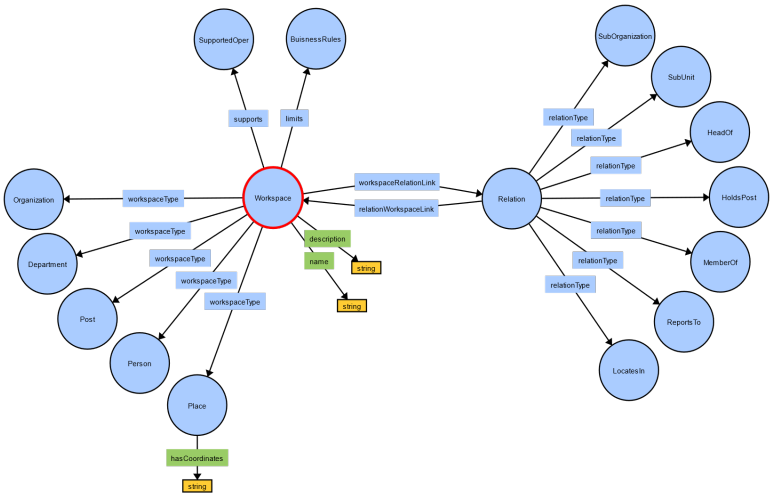

Рис. 3. Схема разработанной онтологии
Fig. 3. Developed ontology scheme

«Бизнес-правила» представляют собой набор правил-ограничений, описывающих порядок работы некоторой сущности предметной области, например, время работы того или иного элемента организационной структуры.

Необходимо отметить, что класс «SupportedOperations» является интерфейсом к семантической модели бизнес-процессов, которая также разрабатывается в рамках проекта по созданию интеллектуального помощника, призванного помочь людям в адаптации к новой для них организации.

Таким образом, разработанная модель представляет собой универсальную онтологию, позволяющую строить модели даже самых сложных организационных структур, а используемые технологии семантического моделирования дают возможность использования логики первого порядка для вывода новых фактов и проверки модели на противоречивость заданным правилам.

2.2. Модель описания бизнес-процессов

Но модель структуры организации была бы бессмысленна без наличия модели, позволяющей описывать бизнес-процессы, происходящие внутри системы.

Организации, взаимодействующие с людьми, как правило, имеют сложные бизнес-процессы, подразумевающие как участия в них большого числа сотрудников, так и возможные интеграции с другими компаниями.

Стоит отметить важность последнего пункта, бизнес-процесс нельзя представить в виде обычного конвейера, а наоборот, это сложная структура с возможностью условных переходов между состояниями [5].

Разработанная модель, так же как и предыдущая, представляет собой онтологию, описанную с помощью модели представления данных RDF и языка описания онтологий OWL.

Предлагаемая модель позволяет описать бизнес-процессы любой организации. При этом учитывается возможность параллельного выполнения автоматизации некоторых действий, возможного ветвления и т. д. Стоит отметить ключевые особенности данной системы и описать ключевые понятия.

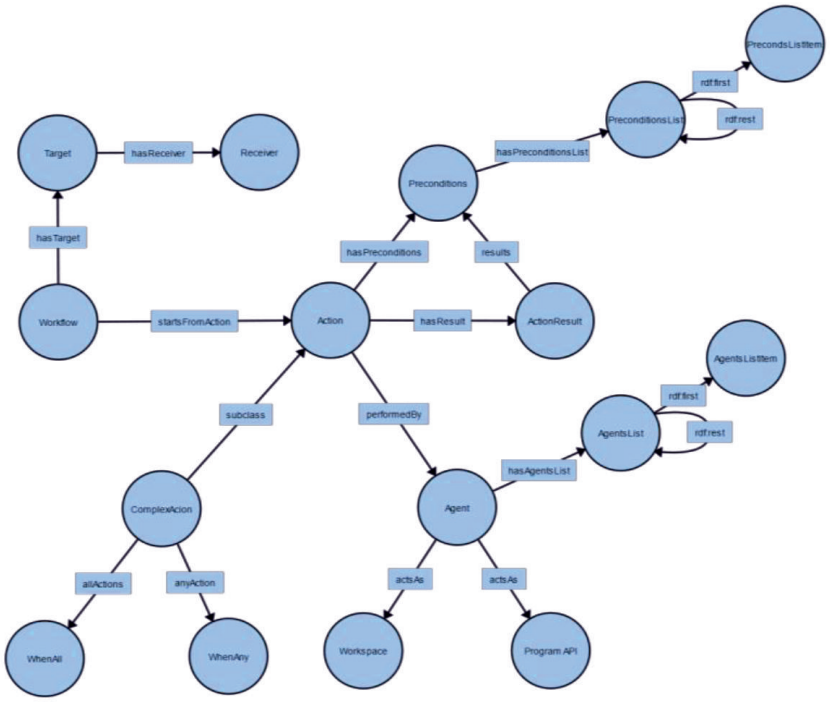

Рис. 4. Схема представления бизнес-процессов
Fig. 4. Business process representation diagram

2.3. Модель пользователя

В последнее время тяжело представить работающую информационную систему, которая бы не использовала в своей работе информационную модель представления знаний о пользователе, поскольку любая информационная система разрабатывается в конечном счете для людей, которые будут пользоваться данной системой.

Для формирования релевантных ответов и оптимальных предложений необходимо узнать о пользователе как можно больше информации за короткое время взаимодействия [6]. В связи с этим возникает потребность в самостоятельном извлечении недостающих знаний в ходе диалога. Это становится возможным с применением методов машинного обучения, которые активно развиваются в последние годы [7]. Для дополнения уровня знаний модели пользователя корректными и непротиворечивыми данными используются семантические технологии.

Модуль, реализующий гибридный подход, включает несколько обученных нейронных сетей, выявляющих требуемые характеристики для семантической модели пользователя [8]. Архитектура позволяет свободно дополнять набор используемых нейронных сетей, что расширяет спектр его применения.

В соответствии с целью работы были сформулированы требования, архитектура системы и составлена семантическая модель пользователя. На диаграмме также показаны такие элементы, более подробно описанные выше, как сущности Workspace или BusinessProcess.

2.4. Модель диалога

Распознавания голосовых команд часто недостаточно, поскольку пользователь может не указать некоторые параметры, требуемые помощнику. Например, человек сказал: «Кто мо-

жет выдать справку о том, что я студент?» Но для получения справки необходимо указать организацию, для которой она выдается.

Для уточнения или получения обратной связи необходимо иметь возможность управлять процессом диалога и вести двунаправленный диалог. Предлагается моделировать диалог в виде конечного автомата.

В качестве примера рассмотрим задачу «заказ еды». Во время диалога система заполняет следующие слоты (аргументы для пользовательских намерений): адрес (slot1), телефон (slot2), заведение для заказа (slot3), название еды (slot4), тип еды (slot5), диапазон цен (slot6). Некоторые из них являются обязательными.

Таким образом, мы видим, что необязательные слоты могут помочь в ситуации, когда пользователь затрудняется ответить или это для него не важно. В итоге оптимизируется количество раундов разговора.

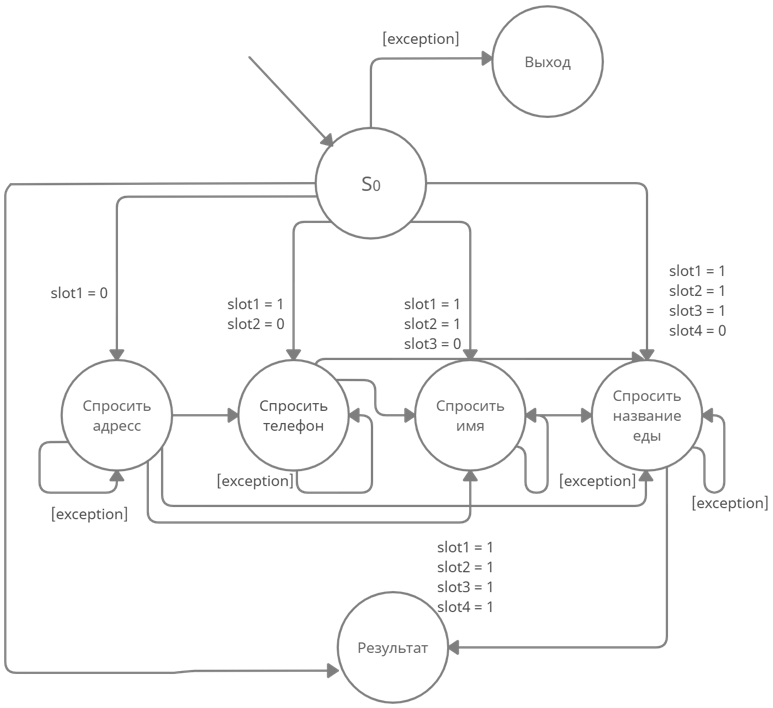

Рис. 5. Пример модели диалога, построенного в процессе работы системы
Fig. 5. Example of a dialog model built during the operation of the system

2.5. Методология работы с пользовательскими намерениями

В рамках данной работы мы рассматриваем намерение как сущность, привязанную к семантической модели организации. Это означает, что намерения могут ссылаться на какой-то отдел, должность, поддерживаемую операцию, бизнес-правило и т. д.

Для порождения пользовательских намерений мы используем средства логического вывода для семантических моделей. Для этого задается некоторый набор правил, с помощью которого удается получить список пользовательских намерений. Намерения, созданные с помощью такого механизма, будем называть генерализованными.

Пример

Пусть имеем следующий шаблон непараметризованного интента:

template Intent **<*T*>** {},

где Intent = *ОформитьДокументыНа* + **<*T*>**.

Тогда можно сгенерировать следующие примеры конкретных интентов по заданному шаблону:

1) Intent_1 = *ОформитьДокументыНаОтпуск*;
2) Intent_2 = *ОформитьДокументыНаПринятиеНаРаботу;*
3) Intent_3 = *ОформитьДокументыНаПолучениеСтипендии*;
4) Intent_4 = *ОформитьДокументыНаУвольнение*;

и т. д.

Таким образом, генерализированные намерения – это результат шаблонизации пользовательских намерений на основе семантической модели. Подход шаблонизации дает возможность работы с разными типами интентов, требуя от них наличия лишь некоторых определенных свойств. Однако данная система содержит в себе семантическую модель бизнес-процессов, которые зачастую могут быть параметризованы, это также необходимо учитывать, поэтому используется подход параметризации интентов [9]. Таким образом, мы можем создать множество намерений, которые зависят от разных параметров, задаваемых пользователем.

Пример

Пользователю необходимо заказать такси. Тогда имеем следующий *параметризованный* интент:

Заказать_такси (*время*, *место_отправления*, *место_прибытия*, *класс_автомобиля*).

Стоит отметить, что возможны случаи, когда пользователем определены не все параметры интента или же значения определенных параметров заданы по умолчанию. Тогда можно считать такие намерения полиморфными.

Полиморфным классом будем называть множество интентов, описывающих одно и то же намерение пользователя, но имеющих различную степень детализации. Интенты, принадлежащие одному классу полиморфности, называются полиморфными.

В данном случае имеет место аналогия с полиморфизмом в объектно-ориентированном программировании, когда функция умеет работать с различными типами аргументов, как будто это один тип, однако поведение каждого типа уникально в зависимости от его реализации. В рассматриваемом же случае свойство полиморфизма означает, что процесс распознавания и последующей обработки интента зависит от того, к какому классу полиморфности он принадлежит.

2.6. Распознавание пользовательских намерений

Распознавание пользовательских намерений можно выполнить различными методами [10, 11], например, с помощью *SoftTripletLoss* функции потерь для нейронных сетей, обучаемых с помощью метрических способов. Архитектура *SoftTripletLoss* позволяет снизить критерии качества для обучающей выборки.

Часто за основу для обучения таких нейронных сетей выбирают массив данных *SNLI*, представленный группой ученых, занимающихся обработкой текстов на естественном языке из Стэнфордского университета.

Однако в целом задача распознавания пользовательских намерений [12] не является ключевой в данном исследовании, поскольку основной акцент сделан именно на методах семантического моделирования и способе порождения пользовательских намерений.

Заключение

В рамках работы предложены методы создания интеллектуальных помощников на основе семантического моделирования с использованием средств логического вывода для порождения новых пользовательских намерений.

Для достижения цели исследования по созданию интеллектуального помощника с использованием средств семантического моделирования для описания бизнес-процессов и структуры организации были поставлены следующие задачи.

1. Разработать способ анализа входящих сообщений для распознавания намерений пользователя.
2. Разработать механизм управления диалогом с пользователем.
3. Разработать семантическую модель структуры организации.
4. Разработать семантическую модель бизнес-процессов организации.
5. Разработать прототип интеллектуального помощника.

Разрабатываемые семантические модели и методы работы с намерениями являются ключевыми компонентами создаваемого в рамках данной работы интеллектуального помощника.

Список литературы

1. **Шолле Ф.** Глубокое обучение на Python. СПб.: Питер, 2018. С. 346–348.
2. **Kuratov Y., Arkhipov M.** Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language. 2019. arXiv preprint arXiv:1905.07213.
3. **Hamilton W. L., Rex Ying, and Leskovec J.** Representation learning on graphs: Methods and application // IEEE Data Eng. Bull., 2017a.
4. **Sahisnu Mazumder, Nianzu Ma, and Bing Liu.** Towards a Continuous Knowledge Learning Engine for Chatbots. 2018. arXiv:1802.06024.
5. **Zhanming Jie, Wei Lu.** Dependency-guided lstm-crf for named entity recognition // Proceedings of EMNLP. 2019.
6. **Kipf Th. N., Welling M.** Semi-supervised classification with graph convolutional networks // Proceedings of ICLR. 2017.
7. **Gubichev A., Bedathur S., Seufert S., Weikum G.** Fast and accurate estimation of shortest paths in large graphs // Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM ’10. New York, NY, USA, 2010. P. 499–508.
8. **Fellbaum C.** (to appear) Future Challenges for the Princeton WordNet. In: Special Issue on Linking, Integrating and Extending Wordnets. Linguistic Issues in LanguageTechnology, eds. Francis Bond, Christiane Fellbaum and Ewa Rudnicka.
9. **Шолле, Ф.** Глубокое обучение на Python. СПб.: Питер, 2018. С. 173–191.
10. **Guo Z., Zhang Y., Lu W.** Attention Guided Graph Convolutional Networks for Relation Extraction // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July – 2 August 2019. P. 241–251.
11. BERT large model (uncased) for Sentence Embeddings in Russian language. URL: https://huggingface.co/sberbank-ai/sbert_large_nlu_ru
12. **Corso G., Cavalleri L., Beaini D., Lio P., Velicˇkovic´ P.** Principal neighborhood aggregation for graph nets // Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 13260–13271.

References

1. **Chollet, F.** Deep learning with Python. St. Petersburg: Piter, 2018, pp. 346–348.

2. **Kuratov, Y., Arkhipov, M.** Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language, 2019, arXiv preprint arXiv:1905.07213.
3. **Hamilton, W. L., Rex Ying, and Leskovec, J.** Representation learning on graphs: Methods and applications. *IEEE Data Eng. Bull.*, 2017a.
4. **Sahisnu Mazumder, Nianzu Ma, and Bing Liu.** Towards a Continuous Knowledge Learning Engine for Chatbots, 2018, arXiv:1802.06024.
5. **Zhanming Jie and Wei Lu.** Dependency-guided lstm-crf for named entity recognition. Proceedings of EMNLP, 2019.
6. **Kipf, Th. N., Welling, M.** Semi-supervised classification with graph convolutional networks. *Proceedings of ICLR*, 2017.
7. **Gubichev, A., Bedathur, S., Seufert, S., Weikum, G.** Fast and accurate estimation of shortest paths in large graphs. In Proceedings of the 19th ACM Inter- national Conference on Information and Knowledge Management, CIKM ’10, New York, NY, USA, 2010, pp. 499–508.
8. **Fellbaum, C.** (to appear) Future Challenges for the Princeton WordNet. In: Special Issue on Linking, Integrating and Extending Wordnets. Linguistic Issues in LanguageTechnology, eds. Francis Bond, Christiane Fellbaum and Ewa Rudnicka.
9. **Chollet, F.** Deep learning with Python. St. Petersburg: Piter, 2018, pp. 173–191.
10. **Guo, Z., Zhang, Y., Lu, W.** Attention Guided Graph Convolutional Networks for Relation Extraction. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 28 July – 2 August 2019; pp. 241–251.
11. BERT large model (uncased) for Sentence Embeddings in Russian language, Access mode: https://huggingface.co/sberbank-ai/sbert_large_nlu_ru
12. **Corso, G., Cavalleri, L., Beaini, D., Lio, P., Velic̆ković, P.** Principal neighborhood aggregation for graph nets. *Advances in Neural Information Processing Systems*, 2020, vol. 33, pp. 13260–13271.

Информация об авторах

Трегубов Артем Сергеевич, аспирант Новосибирского государственного университета

Немцев Иван Сергеевич, магистрант Новосибирского государственного университета

Котельникова Анна Александровна, бакалавр Новосибирского государственного университета

Доможакова Дарья Андреевна, бакалавр Новосибирского государственного университета

Information about the Authors

Artem S. Tregrubov, Graduate Student, Novosibirsk State University (Novosibirsk, Russian Federation)

Ivan S. Nemtsev, Master’s Student, Novosibirsk State University (Novosibirsk, Russian Federation)

Anna A. Kotelnikova, Bachelor, Novosibirsk State University (Novosibirsk, Russian Federation)

Darya A. Domozhakova, Bachelor, Novosibirsk State University (Novosibirsk, Russian Federation)

Статья поступила в редакцию 10.07.2023;
одобрена после рецензирования 20.08.2023; принята к публикации 20.08.2023

The article was submitted 10.07.2023;
approved after reviewing 20.08.2023; accepted for publication 20.08.2023