Труды Кольского научного центра РАН. Серия: Технические науки. 2022. Т. 13, № 2. С. 23–30.
Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2022. Vol. 13, No. 2. P. 23–30.

Научная статья
УДК 004.853
doi:10.37614/2949-1215.2022.13.2.002

ИЗВЛЕЧЕНИЕ ОТНОШЕНИЙ ИЗ NER-РАЗМЕЧЕННЫХ ПРЕДЛОЖЕНИЙ ДЛЯ ОБУЧЕНИЯ ОНТОЛОГИИ*

Павел Андреевич Ломов[1]✉, Марина Леонидовна Никонорова[2], Максим Геннадьевич Шишаев[3]
[1, 2, 3]Институт информатики и математического моделирования имени В. А. Путилова Кольского научного центра Российской академии наук, Апатиты, Россия
[1]lomov@iimm.ru✉, https://orcid.org/0000-0002-0924-0188
[2]malozemova@iimm.ru, https://orcid.org/0000-0002-4358-2683
[3]shishaev@iimm.ru, https://orcid.org/0000-0001-7070-7878

Аннотация

Данная работа является продолжением исследования по обучению и использованию языковой нейросетевой модели для решения задачи обучения онтологий. Рассматривается проблема анализа естественно-языковых текстов по предметной области и извлечения из них возможных отношений между понятиями. Представлен краткий обзор существующих подходов к извлечению отношений из неструктурированных текстовых данных. Предложена процедура извлечения отношений между понятиями на основе анализа дерева синтаксического анализа предложения. Рассмотрены результаты оценки эффективности извлечения отношений в результате применения данной процедуры.

Ключевые слова:

извлечение отношений, обучение онтологии, NLP

Благодарности:

исследование выполнено в рамках государственного задания Института информатики и математического моделирования имени В. А. Путилова Кольского научного центра Российской академии наук от Министерства науки и высшего образования Российской Федерации, тема научно-исследовательской работы «Методология создания информационно-аналитических систем поддержки управления региональным развитием, основанных на формирующем искусственном интеллекте и больших данных» (регистрационный номер 122022800551-0).

Для цитирования:

Ломов П. А., Никонорова М. Л., Шишаев М. Г. Извлечение отношений из NER-размеченных предложений для обучения онтологии // Труды Кольского научного центра РАН. Серия: Технические науки. 2022. Т. 13, № 2. С. 23–30. doi:10.37614/2949-1215.2022.13.2.002

Original article

EXTRACTING RELATIONS FROM NER-TAGGED SENTENCES FOR ONTOLOGY LEARNING

Pavel A. Lomov[1]✉, Marina L. Nikonorova[2], Maxim G. Shishaev[3]
[1, 2, 3]Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences, Apatity, Russia
[1]lomov@iimm.ru✉, https://orcid.org/0000-0002-0924-0188
[2]malozemova@iimm.ru, https://orcid.org/0000-0002-4358-2683
[3]shishaev@iimm.ru, https://orcid.org/0000-0001-7070-7878

Abstract

This paper is a continuation of the research on training and using a neural-network language model for ontology learning. The paper deals with the problem of analyzing domain natural language texts and extracting possible relations between concepts from them. A brief overview of existing approaches to solving this problem is presented. A procedure for extracting relations between concepts based on the analysis of the sentence syntax dependency tree is proposed. The results of evaluating the efficiency of extracting relations using this procedure are considered.

Keywords:

relation extraction, ontology learning, NLP

* Адаптированный перевод статьи: Lomov P. A. Extracting Relations from NER-tagged Sentences for Ontology Learning / P. A. Lomov, M. L. Malozemova, M. G. Shishaev // Artificial Intelligence Trends in Systems. Lecture Notes in Networks and Systems / ed. R. Silhavy. Cham: Springer International Publishing, 2022. pp. 337–344.

Acknowledgments:

the study was carried out within the framework of the Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences state assignment of the Ministry of Science and Higher Education of the Russian Federation, research topic “Methodology for creating information and analytical systems to support the management of regional development based on formative artificial intelligence and big data” (registration number of the research topic 122022800551-0).

For citation:

Lomov P. A., Nikonorova M. L., Shishaev M. G. Extracting relations from NER-tagged sentences for ontology learning // Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2022. Vol. 13, No. 2. P. 23–30. doi:10.37614/2949-1215.2022.13.2.002

Введение

На сегодняшний день задача извлечения отношений из неструктурированных текстовых данных является довольно актуальной при решении различного рода прикладных задач обработки естественного языка (Natural Language Processing, NLP). В частности, задача извлечения отношений предполагает обнаружение в текстовых данных отношений между сущностями. Они могут быть представлены в виде троек <субъект, отношение, объект>, которые могут быть извлечены с помощью подходов, предполагающих использование лексико-синтаксических шаблонов, автоматических или созданных вручную правил (эвристик), а также с помощью методов машинного обучения (нейронные сети).

Извлечение отношений является одним из ключевых этапов в процессе разработки онтологий. Под онтологией понимается концептуальная модель предметной области, разделяемая некоторой группой агентов (люди, организации, информационные системы и т. д.) [1]. Разработка онтологий рассматривается в рамках процесса обучения онтологий (ontology learning), который включает в себя такие шаги, как извлечение таксономических и нетаксономических отношений из текста. Извлечение таксономических отношений позволяет построить основную иерархию обнаруженных в текстах предметной области концептов, а извлечение нетаксономических отношений позволяет отразить предметные связи между ними.

В данной работе рассматривается проблема извлечения отношений из текстов по предметной области с целью дальнейшего их добавления в онтологию. При этом извлечение производится из тех предложений, в которых с помощью ранее предложенной технологии [2, 3] были обнаружены возможные понятия онтологии. Поэтому здесь рассматривается последующее использование полученного набора предложений с метками, указывающими положение найденных понятий.

Предыдущая работа

В прошлой работе [2] была предложена технология, предполагающая анализ онтологии для формирования исходного списка ее понятий, сбор и анализ текстов, относящихся к предметной области онтологии, с формированием в результате обучающего набора размеченных предложений. Метка, присваиваемая предложению, в этом случае содержала границы обнаруженного в нем понятия и его категорию.

Далее данный набор применялся для обучения нейросетевой языковой модели, ориентированной на решение задачи извлечения именованных сущностей. Модель впоследствии применялась для извлечения из текстов новых понятий — кандидатов на добавление в онтологию.

В работе [3] в технологию были добавлены шаги «уточнения» понятий и аугментации полученного набора. Шаг уточнения предполагал включение в состав понятия некоторых синтаксически связанных с ним слов предложения. Например, уточнение понятия «аэропорт» до «Международный аэропорт Шереметьево» в предложении «Международный аэропорт Шереметьево является крупнейшим в России». Это позволяло точнее скорректировать границы понятия в предложениях обучающего набора и тем самым правильно задать контекст его употребления, что положительно сказалось при обучении модели.

Аугментация, в свою очередь, предполагала генерацию новых образцов — размеченных предложений путем замены некоторых комбинаций слов из контекста понятия на альтернативные, предложенные предобученной общеязыковой BERT-моделью для русского языка. Это также давало некоторое (незначительное по сравнению с уточнением понятий) улучшение результативности модели.

В данной работе мы рассмотрим проблему извлечения отношений из размеченных предложений, полученных в результате применения уже обученной модели на наборе текстов. Предполагается, что в таких предложениях уже существует понятие, релевантное исходной онтологии, и требуется найти связанное с ним некоторым отношением другое понятие.

Обзор существующих подходов к извлечению отношений

Извлечение отношений из текста — это подзадача NLP, ориентированная на выявление отношений между парами сущностей в неструктурированных текстовых данных. В ранних работах по извлечению отношений из текста использовались подходы на основе правил (например, паттерны Hearst [4, 5]), а также признаковые модели. Например, Mintz и др. [6] учитывали в модели лексические признаки, такие как: последовательность слов между двумя сущностями и их тэги частей речи; маркер, указывающий, какая сущность появляется первой в предложении; количество токенов слева от первой сущности и количество токенов справа от второй сущности, а также их тэги частей речи. Кроме того, авторы извлекали путь зависимости (dependency path) между двумя сущностями как синтаксический признак, а также их типы.

В работе [7] для извлечения отношений учитывали следующие признаки: главные токены двух сущностей, токены двух сущностей, токены между двумя сущностями, их теги частей речи, порядок следования двух сущностей, расстояние между ними и кластер Брауна [8] для каждого токена.

В недавних исследованиях задачу извлечения отношений решают с использованием нейросетевых моделей на основе архитектуры Transformer [9]. Данные работы можно условно разделить на две группы: первая группа ориентирована на извлечение отношений в рамках предложения [10–12], а вторая — на извлечение отношений в рамках документа [13, 14]. Например, в работе [10] представлен фреймворк CASREL на базе BERT-модели, который позволяет идентифицировать все возможные тройки токенов (субъект, отношение, объект) в предложении, где некоторые такие тройки могут включать одинаковые сущности. Первым этапом в предложении идентифицируются все возможные токены-субъекты. Вторым этапом, используя обученные для каждого типа отношения тэггеры и выявленные токены-субъекты, идентифицируются все возможные отношения и соответствующие токены-объекты.

В работе [11] на вход предобученной BERT-модели передается предложение, в котором позиции двух сущностей, между которыми необходимо определить отношение, помечены специальными токенами. BERT учитывает эти позиции, а также контекст предложения, благодаря чему более точно предсказывает отношение между сущностями. В работе [12] авторы генерируют обучающие наборы данных, содержащие операторы отношений, представляющие собой предложения, в которых сущности заменены токеном [BLANK]. Модель BERT принимает на вход пару таких операторов отношений, содержащих одинаковые маскированные сущности, а на выходе строит их схожие векторные представления отношений.

Помимо архитектуры Transformer в последнее время стало распространенной практикой использование архитектуры на базе графовых сверточных сетей (Graph Convolutional Network, GCN). GraphRel [15] позволяет извлекать не только отношения, но и сами сущности в два этапа. На первом этапе, применяя двунаправленные рекуррентные и GCN сети, извлекаются последовательные и локальные признаки слов, с учетом которых предсказываются отношения для каждой пары слов, а также их типы. На втором этапе для каждого предсказанного отношения строятся полные реляционные графы, к которым последовательно применяются GCN. В результате этого этапа извлекаются достаточные признаки слов, с учетом которых выполняется более точная классификация сущностей и отношений.

В качестве работ, рассматривающих извлечение отношений в рамках документа, можно привести работы [13, 14]. В работе [13] для этой цели также используют BERT-модель. Чтобы точнее описать каждое упоминание сущности в тексте, авторы предлагают маскировать сущности двумя специальными токенами, один из которых — тип сущности, а второй указывает на номер сущности в документе (например, «[LOC] Австралия [MASK_1]»). Такое представление входных данных обеспечивает более точное определение типов отношений между сущностями, а также позволяет решить проблему кореференции за счет связывания слов одинаковым токеном в разных предложениях.

В работе [14] для извлечения отношений в рамках документа также используют GCN. Авторы рассматривают каждый токен в документе как узел в графе, а для создания связей между ними используют отношения из дерева зависимостей предложения, ребра кореференции (соединяют токены документа, которые относятся к одной и той же сущности), ребра смежности слов и предложений (для сохранения последовательности информации) и ребра-петли (чтобы учесть информацию о самих узлах).

Другим примером решения задачи извлечения отношений является ее рассмотрение как вопросно-ответной задачи [16]. Первым этапом, используя шаблоны вопросов о типе сущности, из исходного предложения извлекают сущность-субъект отношения. Вторым этапом генерируется вопрос, используя шаблоны отношений, куда помещают эту сущность-субъект. В результате ответа на данный вопрос извлекается само отношение и сущность-объект отношения.

В работе [17] используют прототипы для извлечения отношений. Прототипами называют представления (embeddings) в пространстве признаков, которые выделяют наиболее важную семантику отношений между сущностями в предложениях. Прототипы представляют собой центры кластеров, представляющих множество предложений, выражающих одно и то же отношение. По словам авторов, такой подход позволяет выделить значимые, интерпретируемые прототипы для окончательной классификации отношений.

Процедура извлечения отношений между понятиями на основе дерева синтаксического анализа предложения

В данной работе извлечение отношений для последующего их добавления в онтологию производится в рамках отдельных предложений из текстов предметной области. При этом рассматриваются только те предложения, в которых модель, обученная с применением технологии, описанной в предыдущей работе [3], обнаружила понятия, соответствующие тематике пополняемой онтологии.

Таким образом, предполагается, что слова, представляющие отношение, и понятия, которое оно связывает, находятся в одном предложении. В основе алгоритма лежит гипотеза о том, что отношение между понятиями может быть выражено: 1) через сказуемое, представленное глаголом, который является корнем синтаксического дерева предложения; 2) синтаксические отношения между словами внутри именных групп. Например, в именной группе «*пресс-секретарь полиции Сассекса*» можно выделить отношения между парами слов: «*пресс-секретарь*» — «*полиции*» и «*полиции*» — «*Сассекса*».

С учетом данной гипотезы анализ предложений с целью обнаружения в них отношений производится следующим образом:

1. Для каждого предложения с помощью предобученной языковой модели из библиотеки spaCy [18] формируется дерево синтаксического анализа.

2. Далее производится обход дерева, начиная с его вершины, и формирование N-грамм путем комбинации слова с текущего уровня дерева и связанных с ним слов с дочерних уровней. Таким образом представлялись различные варианты разбивки именных групп предложения на фрагменты, соответствующие возможным понятиям. На данном этапе исследования рассматривались униграммы и биграммы, то есть комбинации из одного и двух слов. Например, именная группа «*пресс-секретарь полиции Сассекса*» представлялась в виде N-грамм: «*пресс-секретарь*», «*пресс-секретарь полиции*», «*полиции*», «*полиции Сассекса*».

3. Среди сформированных N-грамм осуществляется поиск вероятных понятий предметной области — участников отношений. Для их выявления предварительно на большом наборе текстов по тематике предметной области обучается Word2Vec-модель [19]. С помощью нее оценивается близость векторов полученных N-грамм и векторов исходного списка понятий онтологии. Если среднее расстояние между вектором сформированной N-граммы и векторами слов онтологии не превышает порогового значения, то N-грамма рассматривается как понятие предметной области.

Например, если биграмма «*пресс-секретарь полиции*» имеет близость выше пороговой, а униграмма «*пресс-секретарь*» — нет, то это значит, что в текстах предметной области понятие «*пресс-секретарь*» встречается реже, чем «*пресс-секретарь полиции*», и поэтому выделять отдельное понятие «*пресс-секретарь*» не нужно.

Пороговое значение вычисляется для каждого предложения отдельно как среднее значение близости всех сформированных на его основе N-грамм к понятиям онтологии. Таким образом, на данном шаге определяются возможные связанные между собой понятия, представленные словами внутри одной именной группы.

4. Если в предложении обнаруживается более одной N-граммы, определенной на предыдущем шаге как понятие предметной области, то предполагается, что между каждой их парой существует отношение. При этом, если N-граммы не накладываются друг на друга и связаны в синтаксическом дереве через другие слова предложения, то упорядоченный по порядку следования в предложении набор промежуточных слов рассматривается как некоторый контекст предполагаемого отношения, указывающий на его смысл.

5. К набору отношений, полученных на основе одного предложения, применяются следующие корректирующие процедуры:

- Если два отношения имеют в качестве одного понятия N-грамму, в которой главным словом является одинаковый глагол, то они объединяются в одно отношение по этому понятию. N-грамма с глаголом переходит в контекст данного отношения. Например, отношения (*компания «Роснефть», купила, context:{})* и *(В 2011, купила, context:{})* будут объединены в одно отношение с новым контекстом *(компания «Роснефть», В 2011, context:{купила})*.
- Если отношение включает N-граммы, в которых главным словом является одинаковое существительное, то данное отношение интерпретируется как таксономическое (is-kind-of). При этом предполагаемым родителем назначается N-грамма с меньшим количеством слов. Например, *(нефтяная компания, компания, context:{is_kind_of, parent: компания, child: нефтяная компания})*.

Таким образом, в результате применения такой процедуры формируется набор отношений в виде пары N-грамм и иногда его контекста. Далее полученный набор представляется разработчику онтологии для принятия окончательного решения о добавлении обнаруженных отношений в онтологию.

Оценка эффективности процедуры извлечения отношений

Для оценки эффективности предлагаемой процедуры извлечения отношений было проведено два эксперимента, для которых были вручную сформированы два тестовых набора, содержащих предложения и эталонные результаты извлечения из них отношений. Первый набор использовался для оценки извлечения отношений произвольного типа. Он включает в себя 500 образцов, где образец представляет собой пару *(предложение, отношение)*. Второй тестовый набор был сформирован для оценки извлечения только таксономических отношений (is-kind-of), поскольку они имеют особую важность в контексте обучения онтологий — позволяют построить иерархию понятий предметной области. Его размер составил 75 тыс. образцов, аналогичных по структуре первому тестовому набору.

Таким образом, результаты работы процедуры сравнивались с эталонными результатами тестовых наборов и вычислялись оценки полноты и точности. Были получены следующие результаты.

Эксперимент 1. Оценка извлечения отношений произвольного типа из предложений первого тестового набора: точность = 0,016; полнота = 0,052.

Эксперимент 2. Оценка извлечения таксономических отношений из предложений второго тестового набора: точность = 0,128; полнота = 0,207.

Низкие оценки точности, полученные в результате первого эксперимента, говорят о наличии в результирующем наборе большого количества нерелевантных тематике онтологии понятий в извлеченных отношениях. Это вызвано тем, что используемая для их фильтрации на следующем шаге Word2Vec-модель пропускает значительное количество N-грамм, не являющихся понятиями предметной области.

Повышение ее результативности, вероятно, требует более тонкой настройки ее метапараметров при обучении. В первую очередь к таковым относится размер словаря, который определяет минимальное количество обнаружений некоторой N-граммы в рассматриваемых текстах для того, чтобы она учитывалась моделью. То есть, если N-грамма отсутствует в словаре, то ее близость к исходным понятиям онтологии не оценивается и она не попадает в формируемые отношения. Таким образом, увеличение значения данного метапараметра может способствовать сокращению количества не относящихся к тематике онтологии N-грамм и, соответственно, содержащих их отношений.

Результаты обнаружения таксономических отношений оказались более высокими, что указывает на целесообразность использования эвристик для уточнения важных для формирования базовой структуры онтологии типов отношений (is-kind-of, part-of, depend-on).

Таким образом, предложенная процедура извлечения отношений может быть использована в качестве дополнительного шага в разработанной ранее технологии наполнения онтологии, а также для формирования иерархии понятий в процессе обучения онтологий.

Заключение

В данной работе рассматривается проблема извлечения отношений из текстов по тематике предметной области с целью их последующего добавления в существующую онтологию. В качестве решения предложена процедура, которую предполагается использовать для расширения предложенной ранее технологии применения нейросетевых моделей для обучения онтологий. Представленная процедура ориентирована на поиск отношений в предложениях, в которых на предыдущих этапах технологии были обнаружены понятия предметной области — кандидаты на включение в онтологию.

В рамках данной процедуры производится анализ синтаксического дерева предложения с целью выявления именных групп и рассмотрения их возможных интерпретаций в виде пар понятий предметной области, связанных между собой отношением. При этом определение слов внутри именной группы, представляющих вероятные понятия предметной области, производится путем оценки близости различных комбинаций слов (N-грамм) к исходному набору понятий онтологии в векторном пространстве. Последнее представляется отдельной Word2Vec-моделью, обученной на наборе текстов, релевантных предметной области онтологии.

Полученные результаты экспериментов говорят о необходимости в первую очередь улучшения фильтрации формируемых N-грамм. Для этого предполагается проанализировать влияние на получаемый результат различных значений метапараметров Word2Vec-модели, а также рассмотреть возможность ее замены и/или дополнения бинарным классификатором для получения более точной оценки близости при выполнении фильтрации.

Список источников

1. Gruber T. R. A translation approach to portable ontology specifications // Knowledge Acquisition. 1993. Vol. 5, № 2. P. 199–220.
2. Lomov P., Malozemova M., Shishaev M. Training and application of neural-network language model for ontology population // Software engineering perspectives in intelligent systems / ed. Silhavy R., Silhavy P., Prokopova Z. Cham: Springer International Publishing, 2020. P. 919–926.
3. Lomov P., Malozemova M., Shishaev M. Data Augmentation in Training Neural-Network Language Model for Ontology Population // Data Science and Intelligent Systems / ed. Silhavy R., Silhavy P., Prokopova Z. Cham.: Springer International Publishing, 2021. P. 669–679.
4. Hearst M. A. Automated Discovery of WordNet Relations // WordNet: An Electronic Lexical Database. MIT Press. Cambridge, 1998. P. 26.
5. Garcia M., Gamallo P. A Weakly-Supervised Rule-Based Approach for Relation Extraction. 2011. P. 10.
6. Mintz M. et al. Distant supervision for relation extraction without labeled data // Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Suntec, Singapore: Association for Computational Linguistics, 2009. P. 1003–1011.
7. Ren X. et al. CoType: Joint Extraction of Typed Entities and Relations with Knowledge Bases // Proceedings of the 26th International Conference on World Wide Web. Perth Australia: International World Wide Web Conferences Steering Committee, 2017. P. 1015–1024.
8. Implementation of the Brown word clustering algorithm [Электронный ресурс]. URL: https://github.com/percyliang/brown-cluster (дата обращения: 23.12.2021).
9. Devlin J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // arXiv:1810.04805 [cs]. 2018.

10. Wei Z. et al. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020. P. 1476–1488.
11. Wu S., He Y. Enriching Pre-trained Language Model with Entity Information for Relation Classification // arXiv:1905.08284 [cs]. 2019.
12. Baldini Soares L. et al. Matching the Blanks: Distributional Similarity for Relation Learning // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. С. 2895–2905.
13. Han X., Wang L. A Novel Document-Level Relation Extraction Method Based on BERT and Entity Information // IEEE Access. 2020. Vol. 8. P. 96912–96919.
14. Sahu S. K. et al. Inter-sentence Relation Extraction with Document-level Graph Convolutional Neural Network // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. P. 4309–4316.
15. Fu T.-J., Li P.-H., Ma W.-Y. GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. P. 1409–1418.
16. Li X. et al. Entity-Relation Extraction as Multi-Turn Question Answering // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019. P. 1340–1350.
17. Ding N. et al. Prototypical Representation Learning for Relation Extraction. 2021. P. 16.
18. Russian spaCy Models Documentation [Электронный ресурс]. URL: https://spacy.io/models/ru#ru_core_news_sm (дата обращения: 15.06.2021).
19. Gensim: Word2Vec Model [Электронный ресурс]. URL: https://radimrehurek.com/gensim/auto_examples/tutorials/run_word2vec.html#word2vec-model (дата обращения: 15.02.2022).

References

1. Gruber T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 1993, vol. 5, no. 2, pp. 199–220.
2. Lomov P., Malozemova M., Shishaev M. Training and application of neural-network language model for ontology population. Software engineering perspectives in intelligent systems. In: Silhavy R., Silhavy P., Prokopova Z. (eds). S*oftware Engineering Perspectives in Intelligent Systems. CoMeSySo 2020 Advances in Intelligent Systems and Computing*, 2020, vol. 1295, Springer, Cham, pp. 919–926.
3. Lomov P., Malozemova M., Shishaev M. Data Augmentation in Training Neural-Network Language Model for Ontology Population. In: Silhavy R., Silhavy P., Prokopova Z. (eds). *Data Science and Intelligent Systems. CoMeSySo 2021. Lecture Notes in Networks and Systems*, 2021, vol. 231, Springer, Cham, pp. 669–679.
4. Hearst M. A. Automated Discovery of WordNet Relations. WordNet: An Electronic Lexical Database. MIT Press. Cambridge, 1998, 26 p.
5. Garcia M., Gamallo P. A Weakly-Supervised Rule-Based Approach for Relation Extraction, 2011, 10 p.
6. Mintz M., Bills S., Snow R., Jurafsky D. Distant supervision for relation extraction without labeled data. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. Suntec, Singapore: Association for Computational Linguistics, 2009, pp. 1003–1011.
7. Ren X., Wu Z., He W. Qu M. CoType: Joint Extraction of Typed Entities and Relations with Knowledge Bases. *Proceedings of the 26th International Conference on World Wide Web*. Perth Australia: International World Wide Web Conferences Steering Committee, 2017, pp. 1015–1024.
8. Implementation of the Brown word clustering algorithm. Available at: https://github.com/percyliang/brown-cluster (accessed 23.12.2021).
9. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs], 2018.

10. Wei Z., Su J., Wang Yue, Tian Yu. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, 2020, pp. 1476–1488.
11. Wu S., He Y. Enriching Pre-trained Language Model with Entity Information for Relation Classification. arXiv:1905.08284 [cs], 2019.
12. Baldini Soares L., FitzGerald N., Ling J., Kwiatkowski T. Matching the Blanks: Distributional Similarity for Relation Learning. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019, pp. 2895–2905.
13. Han X., Wang L. A Novel Document-Level Relation Extraction Method Based on BERT and Entity Information. *IEEE Access*, 2020, vol, 8, pp. 96912–96919.
14. Sahu S. K., Christopoulou F., Miwa M., Ananiadou S. Inter-sentence Relation Extraction with Document-level Graph Convolutional Neural Network. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019, pp. 4309–4316.
15. Fu T.-J., Li P.-H., Ma W.-Y. GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019, pp. 1409–1418.
16. Li X., Yin F., Sun Z., Li X. Entity-Relation Extraction as Multi-Turn Question Answering. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, 2019, pp. 1340–1350.
17. Ding N., Wang X., Fu Yao, Xu G. Prototypical Representation Learning for Relation Extraction, 2021, 16 p.
18. Russian spaCy Models Documentation. Available at: https://spacy.io/models/ru#ru_core_news_sm (accessed 15.06.2021).
19. Gensim: Word2Vec Model. Available at: https://radimrehurek.com/gensim/auto_examples/tutorials/run_word2vec.html#word2vec-model (accessed 15.02.2022).

Информация об авторах

П. А. Ломов — кандидат технических наук, старший научный сотрудник;
М. Л. Никонорова — инженер-исследователь;
М. Г. Шишаев — доктор технических наук, главный научный сотрудник.

Information about the authors

P. A. Lomov — Candidate of Science (Tech.), Senior Research Fellow;
M. L. Nikonorova — Research Engineer;
M. G. Shishaev — Doctor of Science (Tech.), Chief Research Fellow.

Статья поступила в редакцию 15.10.2022; одобрена после рецензирования 01.11.2022; принята к публикации 08.11.2022.
The article was submitted 15.10.2022; approved after reviewing 01.11.2022; accepted for publication 08.11.2022.