

Научная статья

DOI: 10.15593/2224-9397/2022.4.04

УДК 004.891.3

**Е.В. Долгова¹, О.В. Комягина², С.Н. Костарев¹, Д.С. Курушин¹,
О.В. Соболева¹, Н.А. Татарникова³, Р.А. Файзрахманов¹**

¹Пермский национальный исследовательский политехнический университет,
Пермь, Россия

²АО МЦ «Философия красоты и здоровья», Пермь, Россия

³Пермский государственный аграрно-технологический университет,
Пермь, Россия

О ВОЗМОЖНОСТИ АВТОМАТИЧЕСКОГО ИЗВЛЕЧЕНИЯ СИСТЕМЫ ПРАВИЛ ИЗ СЛАБОФОРМАЛИЗОВАННОГО ТЕКСТА

Практика создания систем искусственного интеллекта (ИИ) показывает, что экспертные системы могут эффективно использоваться в медицинской сфере. Основными проблемами создания и применения экспертных систем (ЭС) на практике являются: объяснение полученных результатов в терминах, понятных пользователю; организация диалога с пользователем при помощи интуитивно понятного интерфейса; актуализация знаний. Причем актуализация знаний является важнейшей проблемой. Пополнение знаний в экспертной системе в ручном режиме является трудоемкой задачей и в значительной мере зависит от подбора экспертов и от квалификации инженера по знаниям. **Цель исследования:** варианты автоматического заполнения экспертной системы на основе анализа находящихся в свободном доступе знаний позволяют решить проблемы актуализации. В особенности это актуально для динамически меняющейся предметной области. Поэтому важное значение приобретает обработка текстов медицинского характера, изначально написанных на естественном языке. Целью обработки являются извлечение знаний и формализация их на основе выбранной модели знаний, после чего результат агрегируется в экспертную систему классической структуры. **Методы и результаты:** в статье рассмотрен механизм выделения системы правил из слабоформализованных документов и формирование на ее основе правил экспертной медицинской системы. Для извлечения правил использованы методы корпусной лингвистики, морфологический анализ на основе открытых библиотек для языка Python, графовые методы оценки связей в системе правил. **Практическая значимость:** результат будет использован для оценки результатов тестирования на антигены SARS-CoV-2 при разработке прототипа-демонстратора технологии экспресс-тестирования на антигены широкого спектра вирусов в рамках проекта «Приоритет 2030», подпроект №4 «Человекоцентричные технологии и сервисы».

Ключевые слова: экспертная система, система правил, автоматическое выделение правил, медицина, механизм извлечения связей.

**E.V. Dolgova¹, O.V. Komyagina², S.N. Kostarev¹, D.S. Kurushin¹,
O.V. Soboleva¹, N.A. Tatarnikova³, R.A. Fayzrakhmanov¹**

¹Perm National Research Polytechnic University, Perm, Russian Federation

²АО МС "Philosophy of Beauty and Health", Perm, Russian Federation

³Perm State Agrarian and Technological University, Perm, Russian Federation

ON AUTOMATIC RULE SYSTEM EXTRACTION FROM A WEEKLY FORMALIZED TEXT

The practice of creating artificial intelligence (AI) systems shows that expert systems can be effectively used in the medical field. The main problems of creating and applying expert systems (ES) in practice are: explaining the results obtained in terms understandable to the user; organizing a dialogue with the user using an intuitive interface; updating knowledge. Moreover, updating knowledge is the most important problem. Updating knowledge in the expert system in manual mode is a time-consuming task and largely depends on the selection of experts and on the qualification of a knowledge engineer. **Purpose:** The options for automatic filling of the expert system based on the analysis of freely available knowledge allow solving the problems of updating. This is especially true for a dynamically changing subject area. Therefore, the processing of medical texts originally written in natural language becomes important. The purpose of processing is to extract both knowledge and formalize them based on the selected knowledge model, after which the result is aggregated into an expert system of a classical structure. **Methods and results:** The paper considers the mechanism of the allocation of a system of rules from poorly structured documents and the formation of rules of an expert medical system on its basis. To extract the rules, the methods of corpus linguistics, morphological analysis based on open libraries for Python, graph methods for evaluating relationships in the rule system were used. **Practical relevance:** The result will be used to evaluate the results of testing for SARS-CoV-2 antigens during the development of a prototype-demonstrator of rapid testing technology for antigens of a wide range of viruses within the framework of the "Prioriet 2030" project, sub-project No. 4 "Human-centric technologies and services".

Keywords: expert system, rule system, automatic rule extraction, medicine, relation extraction procedure.

Введение

Медицинская диагностика вообще и диагностика новых инфекций в частности связаны с принятием решений на основе часто обновляемой информации, которая зачастую требует нетривиальной логической и семантической обработки и глубинного анализа [1]. Вместе с тем поток такой информации с течением времени значительно возрастает, и она неизбежно начинает включать в себя неоднозначные по семантике, нечеткие (fuzzy) элементы. В этой ситуации становится перспективным использование в диагностике систем ИИ, в частности, сложных экспертных систем (ЭС) динамического характера с глубинным анализом [1–5]. Однако создание таких ЭС неизбежно сталкивается с проблемой заполнения знаниями. Поскольку существующие медицинские глобальные онтологии, хотя и отлично формализованы, но довольно

инертны и редко обновляются. Ручное заполнение базы знаний не может быть реальной альтернативой, поскольку медленно и связано с огромными затратами труда. Создание же базы знаний на основе нейросетей, хотя и является перспективным, сталкивается с той же самой проблемой – получением обучающей выборки.

Таким образом, актуальной становится разработка методологий, моделей и алгоритмов, которые позволяют извлекать информацию из слабоформализованного текста на естественном языке, поскольку такие тексты по всем актуальным медицинским направлениям, как правило, имеются в достаточном количестве, и проблема состоит исключительно в формализации содержащихся в них знаний.

1. Проблемы применения экспертных систем (ЭС) в медицине

Практика создания систем искусственного интеллекта (ИИ) показывает, что экспертные системы могут эффективно использоваться в медицинской сфере. Среди них следует выделить самообучающиеся экспертные системы, рассмотренные, в частности, в [1].

Такие системы должны включать в себя возможности тонкой настройки и позволять системе накапливать опыт в процессе функционирования. В настоящее время разработан широкий спектр различных ЭС, в той или иной мере удовлетворяющих этим требованиям. В частности:

1. ABEL: система для диагностики кислотных и электролитных заболеваний [2].

2. AI/RHEUM: консультирующая система по ревматологии, используется для диагностики ревматологических заболеваний. AI/RHEUM содержит такие данные, как клинические признаки, симптомы, результаты лабораторных анализов, рентгенологические наблюдения, результаты биопсии тканей и гипотезы промежуточного диагноза [3].

3. BLUE FOX: система для диагностики и лечения депрессивных состояний [2].

4. CASNET/GLACOMA: система для диагностики и лечения глаукомы [2].

5. CADUCEOS: система для диагностики внутренних заболеваний общего профиля [2].

6. MYCIN: ранняя экспертная система с обратной связью, которая использовала искусственный интеллект для выявления тяжелых инфекций,

таких как бактериемия и менингит. Также применялась для рекомендации антибиотиков с дозировкой, адаптированной к массе тела пациента [4].

7. ONCOCIN: экспертная система для онкологии [2].

8. PIP: система в области диагностики заболеваний почек [2].

9. SPE: система диагностики воспалительных процессов [2].

10. PUFF: система диагностики заболеваний легких [2].

11. МОДИС-2: система диагностики симптоматической гипертонии [2].

В основе широкого класса экспертных систем в той или иной мере лежит следующая структура, показанная на рисунке [5, 6]. Кроме прочего, в рамках этой схемы может быть предусмотрено динамическое или статическое обновление базы знаний, основанное на внесении знаний оператором, или получение их иным способом – из диалогов с пользователем, из внешних источников, в результате обобщений и машинного обучения.

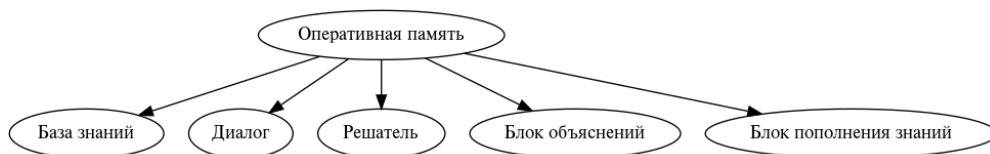


Рис. Типичная структура экспертной системы

Основными проблемами создания и применения ЭС на практике являются:

1) необходимость объяснения полученных при помощи ИИ результатов в терминах, понятных пользователю;

2) детали организации диалога с пользователем при помощи интуитивно понятного интерфейса или при помощи ограниченного естественного языка;

3) проблема актуализации знаний. Третья проблема, безусловно, является самой существенной, так как при неправильном, неточном, неполном заполнении база знаний не позволяет получить приемлемое качество решений. Поскольку логические манипуляции над существующими знаниями новых знаний не порождают, а лишь создают некие их интерпретации, необходимо пополнять базу знаний (БЗ) за счет привлечения экспертов, получения сигналов с датчиков, использования дополнительных онтологий и т.д.

В новых областях медицинских знаний, к которым, безусловно, относится информация в области диагностики COVID-19, проблема осложняется высокой динамикой изменения самих знаний, так как новая информация по поводу коронавирусной инфекции возникает очень часто, но при этом не сосредоточена в одном источнике, а появляется в широком спектре публикаций, размещенных в том числе и в Интернете. Их мониторинг, перенос на модель знаний, систематизация и агрегирование, безусловно, представляют собой проблему.

2. Извлечение правил из текста

Одним из способов актуализации базы знаний является диалог с экспертом. Однако эта работа требует привлечения некоторого круга специалистов и, как следствие, временных затрат. Возможное решение проблемы актуализации базы знаний может быть получено за счет применения методов и моделей извлечения данных из слабоструктурированных и неструктурированных источников. Этот подход рассматривается, в частности, в работе [7]. Автор предлагает метод формулирования правил для извлечения сложных связей из неструктурированного текста для простых и сложносочиненных предложений. Рассматриваются три наиболее распространенных типа сложных предложений.

Следует отметить, что в работе используется подход, основанный на применении графа грамматического разбора. В этом случае текст, подлежащий разбору, уже нельзя считать «неструктурированным», так как его структура должна удовлетворять некоторым требованиям, например [7]:

$$r_i : \{ \exists P_{e_i} \vee \{ l(e_i) = D', D' \subset D \} \}.$$

Здесь r_i кодирует правило, что если в графе зависимостей данного предложения существует определенная последовательность меток зависимостей, без учета порядка (определяемая множеством D), вместе с некоторой цепью зависимостей, то последовательность узлов (которые представляют слова соответствующего предложения) образует информационный компонент. Последовательность узлов (которые представляют слова соответствующего предложения) образует информационную составляющую. Подобные подходы могут успешно применяться, когда текст удовлетворяет заданным правилам, что очевидно следует из их четкой формулировки.

В работе [8] дан обзор еще нескольких систем извлечения знаний их текстов на русском языке, основанных также на системах правил, задаваемых в отличие от предыдущей на специальных языках, например:

```
Rule:
  BornPlace (({Token.kind == word, Token.orth
== upperInitial }):
    person {Token.string == "родился"}
            {Token.string == "в"}
            ({Lookup.majorType == "City"}): city
    ) --> person.Name = {BirthCity =
city.Token.string}
```

Здесь очевидно, что требуется точное текстуальное совпадение слова «родился» с тем, что указано в правиле. Синонимичные и семантически близкие формы не рассматриваются.

В работе [9] рассмотрен подход, близкий к описанному в [10]. И. Мошков предлагает дополнить традиционный разбор текста, основанный на системах правил его семантическим анализом.

Однако в последнее время получили распространение системы, основанные на статистике языка, использующие механизмы машинного обучения, корпусную лингвистику. Такие подходы (в том числе описанный в [10]) предполагают рассматривать анализируемый текст в контексте остальных текстов корпуса, что позволяет учитывать более сложные связи.

Авторами разработан механизм извлечения связей из естественного текста, основанный на совокупности подходов:

- выделение колокаций (термов) из исходного текста при помощи статистических механизмов;
- фильтрация термов по тэгам;
- нормирование и сглаживание частотных характеристик термов;
- определение контекста для каждого термина на основе автоматического поиска его в Интернете;
- рекурсивное выделение колокаций их контекстов использования термина.

Выделение колокаций происходит при помощи библиотеки `gutermextract` [11]. Коллокации (термы) в этом случае представлены как кортежи объектов `Parse` [12], что позволяет отобрать только такие термы, которые не содержат определенных тегов `rumorphy` [13]:

```
{'ADJF', 'LATN', 'UNKN', 'NUMB', 'NUMR'}
```

Поиск контекста использования термина происходит в поисковой системе DuckDuckGo [14], выбранной в связи с относительной простотой использования в составе парсера. Для исключения «флуда» в поисковой системе запросы отправляются с задержкой, снижающей нагрузку на оборудование.

В результате из текста, представляющего собой инструкцию к медицинскому изделию для диагностики *in vitro* «Экспресс-антиген SARS-CoV-2», методом иммунохроматографического анализа «Экспресс-антиген SARS-CoV-2-ИХА» были извлечены следующие отношения (примеры):

- 1) «бактериальные респираторные инфекции» – «верхние дыхательные пути»;
- 2) «взятие биоматериала» – «биологический образец»;
- 3) «взятие биоматериала» – «пластиковая пипетка».

Важно отметить, что в тексте инструкции таких правил непосредственно нет, имеется текст: "«Экспресс-антиген SARS-CoV-2-ИХА» указывает только на наличие/отсутствие антигена SARS-CoV-2 в биоматериале, взятом со слизистой оболочки носоглотки" и «работники, которые собирают или транспортируют клинические образцы в лабораторию, должны быть обучены практике безопасного обращения с биоматериалом». Таким образом, связь «биоматериал» → «взятие» в один терм «взятие биоматериала» получена при помощи библиотеки *gitermextract*, а связь «взятие биоматериала» → «пластиковая пипетка» вычислена при помощи экстратекстуальных сведений, полученных из Интернета. При этом взятие биоматериала в данном экспресс-тесте производится пластиковой пипеткой:

Комплект №1

Тест-картридж для определения антигена SARS-CoV-2 – 20 шт.

Флакон-капельница с буферным раствором – 1 шт.

Микроцентрифужная пробирка – 20 шт.

Пластиковая пипетка – 20 шт.

Инструкция по применению – 1 шт.

Следует отметить, что извлечение правил из текста осложняется проблемами, которые свойственны семантическим сетям вообще и преодоление которых достигается только в их специализированных вариантах. К таким проблемам относится указание времени и модальности.

Например, имеется существенное отличие между отношением «нужно взять» и «можно взять». Равным образом есть разница между «взять в настоящий момент», «взять когда-нибудь в будущем», «взять через 3 минуты».

Такие представления рассмотрены в рамках темпоральной логики [15], однако использование их в семантических сетях затруднительно. Выходом в данном случае является или составление гибридной модели, или унификация отношений, в определении которых модальность и время учитываются заранее по типу темпоральной логики.

3. Анализ извлеченных правил специалистами

Полученные на предыдущем этапе правила были подвергнуты критике со стороны специалистов. В частности, отмечено следующее:

... читаю следующие: искусственные материалы в 4-й графе цифра 1, которая указывает на биологический материал. Но искусственные материалы не могут быть биологическим материалом, поскольку биологический материал предполагается брать у человека или у животного.

То есть имеется предположительно ошибочная связь «искусственные материалы» – «биологический материал».

... дальше – лимфоидная ткань, здесь у вас указана слизистая оболочка носоглотки, но я думаю, надо будет еще указать и носоглотку человека,

То есть эксперт предлагает расширить связь «лимфоидная ткань» – «слизистая оболочка носоглотки» связью «лимфоидная ткань» – «носоглотка человека».

... дальше, нижняя носовая раковина – нижние дыхательные пути. Так будет правильнее, потому что из всех этих биологических объектов берется мазок.

То есть эксперт предлагает создать связь «нижняя носовая раковина» – «нижние дыхательные пути».

... дальше читаю – тут ниже идет назальная полость, тут нужно будет указать полость носа, ... потому что назальная полость – это есть не что иное, как полость носа, то есть цифра 1 – нижние дыхательные пути, но правильнее было бы указать полость носа.

Здесь эксперт предлагает иное прочтение термина «назальная полость» заменить на «полость носа».

Предположительно ошибочная связь «искусственные материалы» – «биологический материал» может быть вызвана как ошибкой алгоритма, так и ошибкой интерпретации экспертом результатов его работы. Дело в том, что отношение между понятиями в графе не указано в силу текущих ограничений работы программы. Таким образом, при разборе текста может быть обработано утверждение «искусственный материал может быть получен переработкой биологического материала» (например, [16]).

Хотя, разумеется, в контексте рассматриваемой проблемной области эта связь, скорее всего, ошибочна. Связь «слизистая оболочка носоглотки» – «лимфоидная ткань» продолжается связью «назальная область» – «слизистая оболочка носоглотки», что позволяет выйти на отношение «назальная область» – «нереактивные образцы» и далее на «полость носа» и другие органы. Эксперту не удалось отследить этот путь, так как он работал с матрицей смежности, а не с графом. Как известно, матрица смежности показывает лишь наличие связи между вершинами V_i и V_j , но не наличие пути (что может быть получено из матрицы достижимости или визуального анализа графа).

Связь «нижняя носовая раковина» – «нижние дыхательные пути» не выявлена алгоритмом, однако существует связь «нижние дыхательные пути» – «полость носа» – «нижняя носовая раковина». Ситуация аналогична вышерассмотренной.

Таким образом, критику правил, выявленных алгоритмом, можно свести к следующим случаям:

- 1) невыявление или некорректное выявление связи типа «является синонимом»;
- 2) невыявление связи первого порядка при наличии связи 2-го и более высоких порядков;
- 3) выявление «лишних» связей.

Заключение

Таким образом, представленные в работе модель и методика позволяют извлекать информацию медицинского характера из слабоформализованного текста на естественном языке. Несмотря на то, при этом возникает некоторое количество артефактов, предложенный подход позволяет с высокой эффективностью автоматизировать процесс заполне-

ния базы знаний экспертной системы. Полученные правила при этом будут представлены в понятной для специалиста форме и в дальнейшем могут редактироваться с целью их улучшения.

Библиографический список

1. Головин П.А., Нечаев В.А., Нечаев Д.А. Экспертные системы для классификации болезней в медицинской диагностике // Научно-технический вестник информационных технологий, механики и оптики. – 2006. – Т. 6, № 6.

2. Головин П.А., Нечаев В.А., Нечаев Д.А. Экспертные системы для классификации болезней в медицинской диагностике // Научно-технический вестник информационных технологий, механики и оптики. – 2006. – № 29.

3. Artificial Intelligence Rheumatology Consultant System Ontology // NCBO BioPortal. – URL: <https://bioportal.bioontology.org/ontologies/AI-RHEUM> (дата обращения: 28.02.2022).

4. Mycin – Wikipedia. – URL: <https://en.wikipedia.org/wiki/Mycin> (дата обращения: 28.02.2022).

5. Васильченко В.А., Бурковский В.Л. Структура экспертной системы оперативной диагностики и лечения заболеваний легких // Вестник ВГТУ. – 2016. – № 1.

6. Молчанов А.Ю. Выбор структуры построения экспертной системы мониторинга режимов электроэнергетической системы // Известия ЮФУ. Технические науки. – 2010. – № 1.

7. Манучарян Л.А. Метод формирования правил для извлечения сложных связей из произвольного текста // Перспективы развития информационных технологий. – 2011. – № 5.

8. Большакова Е.И., Ефремова Н.Э., Шариков Г.Ф. Инструментальные средства для разработки систем извлечения информации из русскоязычных текстов // Новые информационные технологии в автоматизированных системах. – 2015. – № 18.

9. Мошков И.С. Автоматическое извлечение знаний о таксономиях из текста на естественном языке // Вестник Самар. гос. техн. ун-та. Сер. Технические науки. – 2011. – № 1.

10. Курушин Д.С., Леонов Е.Р., Соболева О.В. О возможном подходе к автоматическому построению денотатного графа гипертекста // Информационная структура текста. – М.: Амирит, 2018. – С. 113–118.

11. GitHub – igor-shevchenko/rutermextract: Term extraction for Russian language. – URL: <https://github.com/igor-shevchenko/rutermextract> (дата обращения: 28.02.2022).

12. Морфологический анализатор pymorphy2. – URL: <https://pymorphy2.readthedocs.io/en/stable/user/guide.html> (дата обращения: 28.02.2022).

13. Обозначения для грамем (русский язык) – Морфологический анализатор pymorphy2. – URL: <https://pymorphy2.readthedocs.io/en/stable/user/gramemes.html> (дата обращения: 28.02.2022).

14. DuckDuckGo – Privacy, simplified. – URL: <https://duckduckgo.com/> (дата обращения: 28.02.2022).

15. Ricardo Caferra. Logic for Computer Science and Artificial Intelligence. – John Wiley & Sons, 2013. – 537 p.

16. Биологический материал. – URL: https://wikijaa.ru/wiki/Biological_material (дата обращения: 27.04.2022).

17. Долгова Е.В. Распознавание как этап создания модели технической системы // Вестник Перм. гос. техн. ун-та. Электротехника, информационные технологии, системы управления. – 2009. – № 3. – С. 102–105.

18. Трехмерное моделирование челюстно-лицевых имплантов / Е.В. Ерискина, Е.В. Долгова, Р.А. Файзрахманов, В.П. Василюк // Математические методы в технике и технологиях. – 2020. – Т. 7. – С. 46–50.

19. Ретроспективное исследование дефектов и деформаций челюстей / В.П. Василюк, Е.В. Долгова, Р.А. Файзрахманов, Г.И. Штраубе // Уральский медицинский журнал. – 2019. – № 12. – С. 26–29.

20. Омельчиченко Я. Эволюционные нейросети на языке Python. – М.: ДМК, 2020.

21. Постолит А. Основы искусственного интеллекта в примерах на Python. – СПб: БХВ-Петербург, 2022.

22. Roberto Pasolini. Learning methods and algorithms for semantic text classification across multiple domains. – 2015. – 198 p.

23. Федюшкин Н.А., Федосин С.А. Краткий обзор методов и моделей интеллектуального анализа текста // Проблемы и достижения в науке и технике: сб. науч. тр. по итогам междунар. науч.-практ. конф. – Омск, 2017. – № 4. – 102 с.

24. Федюшкин Н.А., Федосин С.А. Основные технологии интеллектуального анализа текста // Развитие технических наук в современном мире: сб. науч. тр. по итогам междунар. науч.-практ. конф. – Воронеж. – 2016. – № 3. – 128 с.

25. Соловьев А.Н. Язык, мышление и современные системы понимания речи // Вестник СПбГУ. Сер. Биология (3). – СПб.: Изд-во СПбГУ, 2008. – Вып. 1.

References

1. Golovin P.A., Nechaev V.A., Nechaev D.A. Ekspertnye sistemy dlia klassifikatsii boleznei v meditsinskoj diagnostike [Expert systems for classification of diseases in medical diagnostics]. *Nauchno-tekhnicheskii vestnik informatsionnykh tekhnologii, mekhaniki i optiki*, 2006, vol. 6, no. 6.

2. Golovin P.A., Nechaev V.A., Nechaev D.A. Ekspertnye sistemy dlia klassifikatsii boleznei v meditsinskoj diagnostike [Expert systems for classification of diseases in medical diagnostics]. *Nauchno-tekhnicheskii vestnik informatsionnykh tekhnologii, mekhaniki i optiki*, 2006, no. 29.

3. Artificial Intelligence Rheumatology Consultant System Ontology. *NCBO BioPortal*, available at: <https://bioportal.bioontology.org/ontologies/AI-RHEUM> (accessed 28 February 2022).

4. Mycin Wikipedia, available at: <https://en.wikipedia.org/wiki/Mycin> (accessed 28 February 2022).

5. Vasil'chenko V.A., Burkovskii V.L. Struktura ekspertnoi sistemy operativnoi diagnostiki i lecheniia zabolevanii legkikh [The structure of the expert system of operative diagnosis and treatment of lung diseases]. *Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta*, 2016, no. 1.

6. Molchanov A.Iu. Vybory struktury postroeniia ekspertnoi sistemy monitoringa rezhimov elektroenergeticheskoi sistemy [The choice of the structure of the expert system for monitoring the modes of the electric power system]. *Izvestiia Iuzhnogo federal'nogo universiteta. Tekhnicheskie nauki*, 2010, no. 1.

7. Manucharian L.A. Metod formirovaniia pravil dlia izvlecheniia slozhnykh svyazei iz proizvol'nogo teksta [Method of forming rules for extracting complex links from an arbitrary text]. *Perspektivy razvitiia informatsionnykh tekhnologii*, 2011, no. 5.

8. Bol'shakova E.I., Efremova N.E., Sharikov G.F. Instrumental'nye sredstva dlia razrabotki sistem izvlecheniia informatsii iz russkoiazыchnykh

tekstov [Instrumental means for the development of information extraction systems from Russian-language texts]. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh*, 2015, no. 18.

9. Moshkov I.S. Avtomaticheskoe izvlechenie znaniy o taksonomiiakh iz teksta na estestvennom iazyke [Automatic extraction of knowledge about taxonomies from the text in natural language]. *Vestnik Samarskogo gosudarstvennogo tekhnicheskogo universiteta. Tekhnicheskie nauki*, 2011, no. 1.

10. Kurushin D.S., Leonov E.R., Soboleva O.V. O vozmozhnom podkhode k avtomaticheskomu postroeniiu denotatnogo grafa giperteksta [On a possible approach to the automatic construction of a denotational graph of hypertext]. *Informatsionnaia struktura teksta*. Moscow: Amirit, 2018, pp. 113-118.

11. GitHub - igor-shevchenko/rutermextract: Term extraction for Russian language, available at: <https://github.com/igor-shevchenko/rutermextract> (accessed 28 February 2022).

12. Morfologicheskii analizator pymorphy2 [Morphological analyzer pymorphy2], available at: <https://pymorphy2.readthedocs.io/en/stable/user/guide.html> (accessed 28 February 2022).

13. Oboznacheniiia dlia grammem (russkii iazyk) - Morfologicheskii analizator pymorphy2 [Notation for grammem (Russian) - Morphological analyzer pymorphy2], available at: <https://pymorphy2.readthedocs.io/en/stable/user/grammemes.html> (accessed 28 February 2022).

14. DuckDuckGo - Privacy, simplified, available at: <https://duckduckgo.com/> (accessed 28 February 2022).

15. Ricardo Caferra. Logic for Computer Science and Artificial Intelligence. John Wiley & Sons, 2013, 537 p.

16. Biologicheskii material [Biological material], available at: https://wikijaa.ru/wiki/Biological_material (accessed 27 April 2022).

17. Dolgova E.V. Raspoznavanie kak etap sozdaniia modeli tekhnicheskoi sistemy [Recognition as a stage of creating a model of a technical system]. *Vestnik Permskogo gosudarstvennogo tekhnicheskogo universiteta. Elektrotekhnika, informatsionnye tekhnologii, sistemy upravleniia*, 2009, no. 3, pp. 102-105.

18. Eriskina E.V., Dolgova E.V., Faizrakhmanov R.A., Vasiliuk V.P. Trekhmernoe modelirovanie cheliustno-litseyvykh implantov [Three-

dimensional modeling of maxillofacial implants]. *Matematicheskie metody v tekhnike i tekhnologiiakh*, 2020, vol. 7, pp. 46-50.

19. Vasiliuk V.P., Dolgova E.V., Faizrahmanov R.A., Shtraube G.I. Retrospektivnoe issledovanie defektov i deformatsii cheliuste [Retrospective study of jaw defects and deformations]. *Ural'skii meditsinskii zhurnal*, 2019, no. 12, pp. 26-29.

20. Omel'chnchenko Ia. Evoliutsionnye neiroseti na iazyke Python [Evolutionary neural networks in Python]. Moscow: DMK, 2020.

21. Postolit A. Osnovy iskusstvennogo intellekta v primerakh na Python [Fundamentals of Artificial intelligence in Python examples]. Saint Petersburg: BKhV-Peterburg, 2022.

22. Roberto Pasolini. Learning methods and algorithms for semantic text classification across multiple domains, 2015, 198 p.

23. Fediushkin N.A., Fedosin S.A. Kratkii obzor metodov i modelei intellektual'nogo analiza teksta [A brief overview of methods and models of intellectual text analysis]. *Problemy i dostizheniia v nauke i tekhnike. Sbornik nauchnykh trudov po itogam mezhdunarodnoi nauchno-prakticheskoi konferentsii*. Omsk, 2017, no. 4, 102 p.

24. Fediushkin N.A., Fedosin S.A. Osnovnye tekhnologii intellektual'nogo analiza teksta [Basic technologies of intellectual text analysis]. *Razvitie tekhnicheskikh nauk v sovremennom mire. Sbornik nauchnykh trudov po itogam mezhdunarodnoi nauchno-prakticheskoi konferentsii*. Voronezh, 2016, no. 3, 128 p.

25. Solov'ev A.N. Iazyk, myshlenie i sovremennye sistemy ponimaniia rechi [Language, thinking and modern systems of understanding speech]. *Vestnik Sankt-Peterburgskogo universiteta. Biologiya* (3). Saint Petersburg: СПбГУ, 2008, iss. 1.

Сведения об авторах

Долгова Елена Владимировна (Пермь, Россия) – доктор экономических наук, профессор кафедры «Информационные технологии и автоматизированные системы» Пермского национального исследовательского политехнического университета (614990, Пермь, Комсомольский пр., 29, e-mail: itas@pstu.ru).

Комягина Оксана Владимировна (Пермь, Россия) – АО Медицинский центр «Философия красоты и здоровья» (614017, Пермь, ул. Ким, 64, e-mail: itas@pstu.ru).

Костарев Сергей Николаевич (Пермь, Россия) – доктор технических наук, профессор кафедры «Информационные технологии и автоматизированные системы» Пермского национального исследовательского политехнического университета (614990, Пермь, Комсомольский пр., 29, e-mail: itas@pstu.ru).

Курушин Даниил Сергеевич (Пермь, Россия) – кандидат технических наук, доцент кафедры «Информационные технологии и автоматизированные системы» Пермского национального исследовательского политехнического университета (614990, Пермь, Комсомольский пр., 29, e-mail: daniel@kurushin-perm.ru).

Соболева Ольга Владимировна (Пермь, Россия) – кандидат филологических наук, доцент кафедры «Иностранные языки, лингвистика и перевод» Пермского национального исследовательского политехнического университета (614990, Пермь, Комсомольский пр., 29, e-mail: itas@pstu.ru).

Татарникова Наталья Александровна (Пермь, Россия) – заведующая кафедрой «Инфекционные болезни» Пермского государственного аграрно-технологического университета (614000, Пермь, ул. 25 Октября, 10, e-mail: itas@pstu.ru).

Файзрахманов Рустам Абубакирович (Пермь, Россия) – доктор экономических наук, профессор, заведующий кафедрой «Информационные технологии и автоматизированные системы» Пермского национального исследовательского политехнического университета (614990, Пермь, Комсомольский пр., 29, e-mail: itas@pstu.ru).

About the authors

Elena V. Dolgova (Perm, Russian Federation) – Doctor of Economics, Professor of the ITAS PNRPU Department (614990, Perm, 29, Komsomolsky pr., e-mail: itas@pstu.ru).

Oksana V. Komyagina (Perm, Russian Federation) – JSC MC "Philosophy of Beauty and Health" (614990, Perm, 29, Komsomolsky pr., e-mail: itas@pstu.ru).

Sergey N. Kostarev (Perm, Russian Federation) – Doctor of Technical Sciences, Professor of the ITAS PNRPU Department (614990, Perm, 29, Komsomolsky pr., e-mail: itas@pstu.ru).

Daniil S. Kurushin (Perm, Russian Federation) – Candidate of Technical Sciences, Associate Professor of the ITAS PNRPU Department (614990, Perm, 29, Komsomolsky pr., e-mail: daniel@kurushin-perm.ru).

Olga V. Soboleva (Perm, Russian Federation) – Candidate of Philological Sciences, Associate Professor of the Department of IALP PNRPU (614990, Perm, Komsomolsky pr., 29, e-mail: itas@pstu.ru).

Natalia A. Tatarnikova (Perm, Russian Federation) – Head of the Department of Infectious Diseases of the Perm State Agrarian and Technological University (614990, Perm, 29, Komsomolsky pr., Perm, e-mail: itas@pstu.ru).

Rustam A. Fayzrakhmanov (Perm, Russian Federation) – Doctor of Economics, Professor, head of the ITAS PNRPU Department (614990, Perm, 29, Komsomolsky pr., e-mail: itas@pstu.ru).

Поступила: 15.10.2022. Одобрена: 31.10.2022. Принята к публикации: 01.12.2022.

Финансирование. Работа выполнена при финансовой поддержке проекта «Приоритет 2030», подпроект №4 «Человекоцентричные технологии и сервисы».

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов по отношению к статье.

Вклад авторов. Все авторы сделали эквивалентный вклад в подготовку публикации.

Просьба ссылаться на эту статью в русскоязычных источниках следующим образом:

О возможности автоматического извлечения системы правил из слабоформализованного текста / Е.В. Долгова, О.В. Комягина, С.Н. Костарев, Д.С. Курушин, О.В. Соболева, Н.А. Татарникова, Р.А. Файзрахманов // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. – 2022. – № 44. – С. 64–79. DOI: 10.15593/2224-9397/2022.4.04

Please cite this article in English as:

Dolgova E.V., Komyagina O.V., Kostarev S.N., Kurushin D.S., Soboleva O.V., Tatarnikova N.A., Fayzrakhmanov R.A. On automatic rule system extraction from a weekly formalized text. *Perm National Research Polytechnic University Bulletin. Electrotechnics, information technologies, control systems*, 2022, no. 44, pp. 64-79. DOI: 10.15593/2224-9397/2022.4.04