



оригинальная статья

<https://elibrary.ru/ilsjdu>

Применение методов обработки естественного языка для прогнозирования перспективных направлений использования формальных онтологий в биомедицине

Шарнин Михаил Михайлович

Федеральный исследовательский центр «Информатика и управление» РАН, Россия, Москва

eLibrary Author SPIN: 3842-3331

<https://orcid.org/0000-0003-0450-5156>

Scopus Author ID: 6504713775

Калинин Степан Сергеевич

Международный славянский институт, Россия, Москва

eLibrary Author SPIN: 2157-7178

<https://orcid.org/0000-0001-7371-9655>

Scopus Author ID: 57206675409

rage_of_gods@inbox.ru

Аннотация: В работе представлены возможности использования методов обработки естественного языка для прогностического анализа и обзора перспективных направлений использования формальных онтологий в различных областях науки и практики. Выбор формальных онтологий в качестве объекта прогностического исследования обусловлен тем, что они помогают более строго формализовать признаки объектов в различных предметных областях, что дает в дальнейшем возможность для более успешного использования программ машинного обучения для автоматического выявления закономерностей и взаимосвязей между этими признаками. Для анализа перспективных трендов в развитии использования онтологий был использован эксперимент на основе метода машинного обучения и прогнозирования с помощью алгоритма CatBoost в сочетании с методами извлечения информации из текстов и алгоритмом кластеризации лексических единиц. При этом использовались векторные представления соответствующих лексических единиц, отражающих тот или иной тренд частной области знания, между которыми вычислялась мера близости на основе семантического расстояния. В результате эксперимента были выделены четыре наиболее перспективных направления, в которых возможно применение формальных онтологий: связь *генотип – фенотип*, персонализация, алгоритмы кластеризации и совместное управление. Для каждого направления приводятся собственные ключевые слова, отражающие прогнозируемые тренды развития данной конкретной области, а также анализируются примеры наиболее характерных научных работ (статей), отражающих эти тренды.

Ключевые слова: обработка естественного языка, формальные онтологии, библиометрический анализ, прогностическое исследование, метод машинного обучения, векторные представления лексических единиц, кластеризация лексических единиц, алгоритм CatBoost, тренды научного развития

Цитирование: Шарнин М. М., Калинин С. С. Применение методов обработки естественного языка для прогнозирования перспективных направлений использования формальных онтологий в биомедицине. *СибСкрипт*. 2024. Т. 26. № 4. С. 567–575. <https://doi.org/10.21603/sibscript-2024-26-4-567-575>

Поступила в редакцию 19.11.2023. Принята после рецензирования 19.03.2024. Принята в печать 25.03.2024.

full article

Natural Language Processing Tools for Predictive Modeling of Advanced Trends in Formal Ontologies in Biomedical Sciences

Mikhail M. Charnine

Computer Science and Control Federal Research Center, Russian

Academy of Sciences, Russia, Moscow

eLibrary Author SPIN: 3842-3331

<https://orcid.org/0000-0003-0450-5156>

Scopus Author ID: 6504713775

Stepan S. Kalinin

International Slavic Institute, Russia, Moscow

eLibrary Author SPIN: 2157-7178

<https://orcid.org/0000-0001-7371-9655>

Scopus Author ID: 57206675409

rage_of_gods@inbox.ru

Abstract: Natural language processing methods can be used to predict advanced application trends in formal ontologies. Formal ontologies help to formalize the characteristics of objects in various domains. As a result, machine learning programs identify patterns and relationships between these characteristics. The article describes an experiment based on machine learning methods in combination with text search methods. It involves the CatBoost algorithm for predictive modeling and clustering of lexical items. The vector models of the corresponding items reflect a trend in a particular domain of knowledge; proximity between them was calculated based on the idea of semantic distance. The experiment revealed four advanced areas for formal ontologies, i.e., genotype – phenotype; personalization; clustering algorithms, and collaborative task management. Each area that represented the predictable trends of development in this particular domain was provided with keywords. The article also contains a review of most popular scientific articles on these trends.

Key words: natural language processing, formal ontologies, bibliometric analysis, predictive modeling research, methods of machine learning, vector models of lexical items, clustering of lexical items, CatBoost algorithm, trends in intellectual development

Citation: Charnine M. M., Kalinin S. S. Natural Language Processing Tools for Predictive Modeling of Advanced Trends in Formal Ontologies in Biomedical Sciences. *SibScript*, 2024, 26(4): 567–575. (In Russ.) <https://doi.org/10.21603/sibscript-2024-26-4-567-575>

Received 19 Nov 2023. Accepted after peer review 19 Mar 2024. Accepted for publication 25 Mar 2024.

Введение

Под формальными онтологиями понимается особый вид семантических представлений, которые можно определить как модель наблюдаемой действительности, либо как иерархическую форму представления знаний, отражающую структуру наблюдаемой действительности, либо как логическую теорию, которая дает возможность систематизировать выражаемые в языке значения либо категории действительности [Митрофанова, Азарова 2017: 67].

Существует несколько различных типов классификации онтологий [Митрофанова, Азарова 2017: 68–69]: по типу содержательного наполнения выделяются общие онтологии, онтологии задач и предметные онтологии. Общие онтологии представляют наиболее абстрактные понятия, которые являются общими для всех или нескольких предметных областей. Онтологии задач посвящены представлению знаний, связанных с решением какой-либо конкретной задачи

из науки или практики либо группы задач. Предметные онтологии связаны с перечислением объектов, которые используются в какой-либо области профессиональной деятельности.

По цели применения можно выделить онтологии представления, онтологии верхнего уровня, онтологии предметных областей, прикладные онтологии. Онтологии представления создаются для характеристики языка задания онтологий низших уровней. Онтологии верхнего уровня создаются для формального представления тех знаний, которые соотносятся с несколькими предметными областями. Онтологии предметных областей используются для описания основных понятий и терминов конкретной предметной области. Прикладные онтологии предназначены для формализации и структурирования знаний по конкретной задаче или группе задач, относящихся к определенной области науки или практики.

Формально структуру каждой онтологии можно записать следующим образом [Бова и др. 2014: 48]: $O = \langle X, R, F \rangle$, где X – конечное множество понятий-объектов данной предметной области, R – конечное множество тех отношений между понятиями-объектами, которые являются семантически значимыми, F – конечное множество функций интерпретации, которые заданы на объектах-понятиях данной предметной области и / или на семантических отношениях между ними.

В рамках данной работы основным объектом нашего анализа являются конкретные онтологии предметных областей (генетики, коммерции, машинного обучения). По типу наполнения их можно отнести преимущественно к онтологиям задач (например, онтологии, связанные с разработкой и применением алгоритмов кластеризации, или онтологии, относящиеся к управлению задачами). Таким образом, цель настоящего исследования – выявить перспективные направления, которые связаны с формальными онтологиями и которые имеют наиболее долговременные растущие тренды по количеству статей и цитирований, в области биомедицины. Задачи исследования, соответственно, включают в себя выявление ключевых слов, отображающих тренды научного развития в области биомедицины; выявление и анализ наиболее долговременных растущих трендов; иллюстрацию этих трендов при помощи наиболее характерно отражающих их научных работ (статей).

Данная работа основана на комбинации подходов, связанных с анализом формальных онтологий, и извлечением онтологической информации из текстов, соотносящихся с определенной предметной областью. Исследования, выполненные в данном русле, известны [Корсун, Пальчунов 2017; Munir, Anjum 2018], однако они основаны на атомарном подходе в рамках алгебраической теории моделей. Преимуществом настоящей работы является использование векторных представлений извлекаемых лексических единиц, а также методов и алгоритмов машинного обучения для обработки, прогнозирования трендов и кластеризации данных. Это позволяет не только рассчитывать встречаемость тех или иных терминов, но и делать прогностические выводы по их динамике.

Методы и материалы

В результате эксперимента был проанализирован 30031 заголовок статей по теме *Ontology* (выборку из числа которых мы назвали *Выборочная коллекция*) из коллекций PubMed и DBLP-v13. Коллекция DBLP-Citation-network V13 содержит 5354309 статей

по компьютерным наукам, а коллекция PubMed по состоянию на начало 2023 г. – более 34 млн статей по биомедицине. Из PubMed и DBLP-v13 была выделена 30031 статья, содержащая слова, отвечающие запросу *Ontolog**, где звездочка (*) означает любую последовательность букв. Большинство статей в Выборочной коллекции, соответственно, относится к теме онтологии в биомедицине.

Прогноз трендов был выполнен с помощью пакета CatBoost, а визуализация – с помощью алгоритма t-SNE. Долгосрочности трендов для характерных / ключевых слов мы рассчитывали, используя различные показатели для групп статей, содержащих эти ключевые слова. Наиболее важным индикатором тренда ключевого слова является количество цитирований ключевого слова (КЦКС). Для расчета этого показателя мы сначала находим все статьи с этим ключевым словом в определенном году, а затем подсчитываем все цитирующие ссылки на эти статьи. Для каждого слова мы рассчитали продолжительность его трендового роста, равную количеству лет непрерывного роста его средней цитируемости. Такая продолжительность роста тренда была целью для алгоритма машинного обучения CatBoost.

Регрессионная модель CatBoost была обучена для 20 признаков / параметров тренда ключевого слова в текущем и предыдущих годах, включая: текущий КЦКС, общее количество статей с ключевым словом, количество лет роста предыдущего тренда, общий рост цитируемости для предыдущего тренда, количество лет от начала тренда, количество цитирующих связей между статьями со словом от начала тренда, время от трендовой ситуации до текущего / базового года [Шарнин 2023].

В процессе анализа были рассчитаны графики роста количества статей за последние годы (рис. 1), характерные / релевантные ключевые слова в Выборочной коллекции и трендовые ключевые слова среди характерных.

Результаты

В последние годы наблюдается значительный рост количества ежегодно публикуемых научных статей с лексическими единицами *ontology* (онтология), *gene ontology* (генная онтология) и *phenotype* (фенотип) в коллекциях DBLP и PubMed (рис. 1). Так, с 2000 по 2022 г. количество статей со словом *ontology* (онтология) выросло в 18,1 раза, а за последние 2 года (с 2020 по 2022 г.) – в 1,39 раза. За эти 2 года количество статей со словом *gene ontology* (генная онтология)

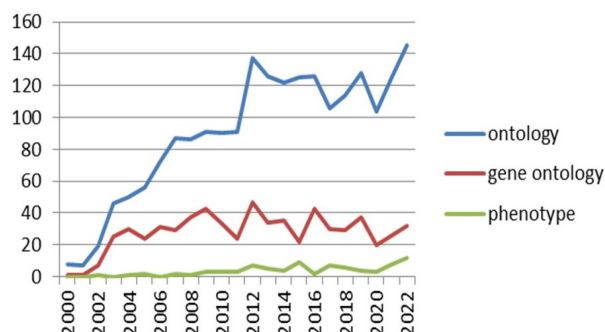


Рис. 1. Количество научных статей со словами *ontology* (онтология), *gene ontology* (генная онтология) и *phenotype* (фенотип) в коллекциях DBLP и PubMed за различные годы
Fig. 1. Research articles with the words *ontology*, *gene ontology*, and *phenotype* in DBLP and PubMed databases

выросло в 1,6 раза, а со словом *phenotype* (фенотип) – в 4 раза. Одновременное увеличение трендов этих трех терминов в последние 2 года подтверждает гипотезу, что общий интерес к онтологиям в последнее время подпитывается в связи с развитием генной онтологии и онтологии фенотипов. Онтология фенотипа человека – это стандартизированный, контролируемый словарь, содержащий фенотипическую информацию о генах и их продуктах. Онтология содержит более 12 тыс. терминов, описывающих фенотипические особенности. Она организована в виде трех независимых онтологий, которые включают различные категории: способ наследования, начало и клиническое течение, а также самую большую категорию фенотипических аномалий.

Характерными / релевантными мы считаем слова, которых относительно много в Выборочной коллекции по сравнению с коллекцией PubMed. Список некоторых наиболее характерных ключевых слов в Выборочной коллекции:

- *ontology* (онтология);
- *ontologies* (онтологии);
- *ontology-based* (на основе онтологических моделей);
- *ontological* (онтологический);
- *gene ontology* (генная онтология);
- *ontology-driven* (управляемый с помощью онтологий);
- *domain ontology* (онтология предметной области);
- *ontology matching* (сопоставление онтологий);
- *ontology-based approach* (подход на основе онтологий);
- *biomedical ontologies* (биомедицинские онтологии);

- *ontology alignment* (согласование онтологий);
- *ontology mapping* (картирование онтологий);
- *ontology development* (разработка онтологий);
- *ontology learning* (обучение онтологий);
- *ontology design* (проектирование онтологий);
- *gene* (ген);
- *ontology engineering* (проектирование онтологий);
- *ontology-based data* (данные, полученные с использованием онтологий);
- *ontology construction* (построение онтологий);
- *ontology driven* (управление онтологией);
- *ontology terms* (термины онтологий);
- *ontology evolution* (эволюция онтологий);
- *ontology semantic* (семантика онтологий);
- *ontology-based semantic* (онтологическая семантика);
- *ontology-based information* (информация, полученная при помощи онтологий);
- *web ontology* (веб-онтология);
- *ontology model* (модель онтологий);
- *ontologies semantic* (семантика онтологий);
- *fuzzy ontology* (нечеткая онтология);
- *ontological analysis* (онтологический анализ);
- *ontology annotation* (аннотация онтологий);
- *formal ontology* (формальная онтология);
- *semantic web* (семантическая сеть);
- *phenotype ontology* (онтология фенотипа);
- *ontology-based knowledge* (знания на основе онтологий);
- *ontological framework* (онтологическая структура);
- *biomedical ontology* (биомедицинская онтология);
- *ontology integration* (интеграция онтологий) и т.д.

С помощью алгоритма CatBoost, который был описан в исследовании [Prokhorenkova et al. 2018] по методике, описанной в работах [Charnine et al. 2021a; 2021b], были рассчитаны прогнозируемые долгосрочности будущих трендов для характерных слов и построен следующий рейтинг перспективности ключевых слов в Выборочной коллекции:

- *bioinformatics* (биоинформатика);
- *interactive* (интерактивный);
- *clustering* (кластеризация);
- *personalized* (персонализированный);
- *automated* (автоматизированный);
- *phenotype* (фенотип);
- *integration* (интеграция);
- *clinical practice* (клиническая практика);
- *collaborative* (совместная работа);
- *managing* (управление);
- *smart* (умный);

- Данный рейтинг отражает результаты прогноза, причем ключевые слова с более долгосрочными прогнозами роста их трендов находятся ближе к началу рейтинга.

Далее была рассчитана совместная встречаемость характерных ключевых слов в заголовках статей в Выборочной коллекции, а также их мера близости, которую неформально можно определить как прямо пропорциональную зависимость частоты совместной встречаемости лексических единиц от значения семантического расстояния между ними. По этим данным построена семантическая карта с помощью алгоритма t-TSNE (рис. 2).

Чем чаще термины встречаются вместе в заголовках статей, тем ближе они расположены на рисунке. Красным цветом выделены наиболее перспективные термины, имеющие долгосрочные тренды, синим – средние тренды, черным – минимальные тренды. Группы рядом расположенных перспективных терминов образуют перспективные направления.

Перспективные направления выделяют в Выборочной коллекции группы статей, в заголовках которых присутствуют перспективные термины. Эти группы статей будут со временем увеличиваться в плане количества входящих статей и количества ссылок цитирования. Продолжительность этого роста соответствует рассчитанной / прогнозной долгосрочности тренда терминов из перспективных областей. Таким образом, перспективное направление применения онтологий – это группа терминов, отображающих перспективную область, с близкими по долгосрочности трендами, которые часто встречаются совместно в заголовках статей, тематически связанных с онтологиями.

На рисунке 2 красным цветом представлены перспективные термины. Можно видеть, что группы терминов образуют кластеры, например:

- *discovering, phenotype, bioinformatics, integration;*
- *personalized, interactive;*
- *clustering, automated;*
- *collaborative, managing, editing.*

Этим кластерам соответствуют перспективные направления, т. к. одинаковые прогнозы рядом встречающихся слов подтверждают друг друга. Итак, можно

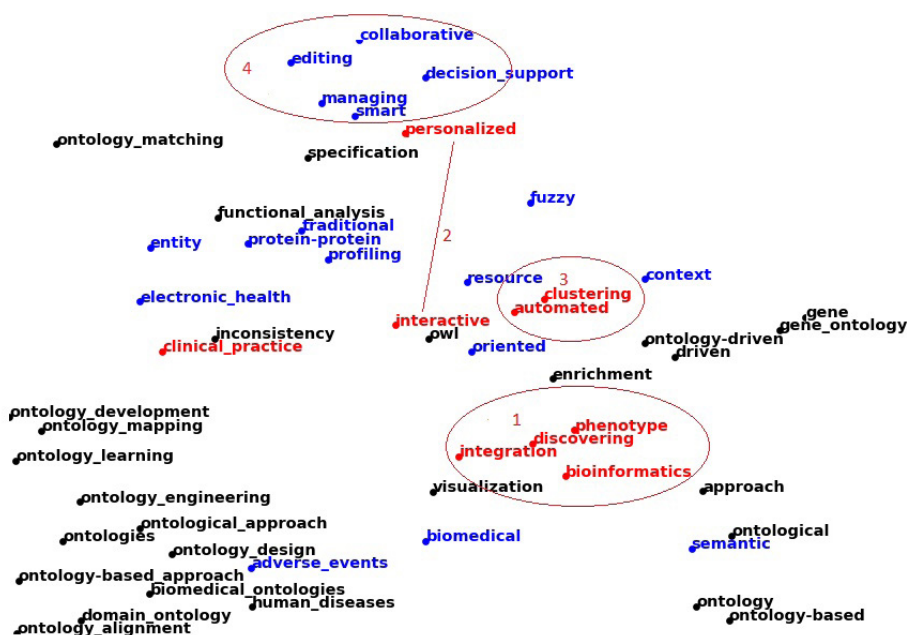


Рис. 2. Семантическая карта терминов, относящихся к предметной области онтологии
Fig. 2. Semantic map of ontology terms

выделить четыре перспективных направления использования онтологий. Этим перспективным направлениям можно дать следующие названия: (1) связь *генотип – фенотип*, (2) персонализация, (3) алгоритмы кластеризации, (4) совместное управление.

Связь *генотип – фенотип*

К первому направлению можно отнести статьи, в которых описано применение онтологий в биомедицине для установления соответствий между признаками в генах (*генотип*) и видимыми внешними проявлениями (*фенотип*). Онтологии помогают формально и строго классифицировать признаки, что позволяет более точно выявлять взаимосвязности / взаимосвязи признаков при анализе больших объемов медицинских данных в электронной форме. Современные методы биоинформатики позволяют автоматически выявлять связи *генотип – фенотип* путем интеграции знаний из различных источников, включая генетическую информацию и семантический анализ текстов медицинских карт. Это направление имеет следующие ключевые слова, большинство из которых выделены красным цветом на рисунке 2: выявление (*discovering*), фенотип (*phenotype*), биоинформатика (*bioinformatics*), интеграция (*integration*), семантический (*semantic*), электронные медицинские карты (*electronic health*). Рассмотрим некоторые статьи из этого направления.

В работе [Holloway 2002] обсуждались вопросы соединения онтологий биоинформатики, описывающих генотипы, и онтологий медицинской информатики, описывающих фенотипы. Статья [Hoehndorf et al. 2007] посвящена вопросам представления знаний по умолчанию в биомедицинских онтологиях и интеграции онтологий анатомии и фенотипа. В работе [Coulet et al. 2008] обсуждается подготовка данных на основе онтологий для выявления связей *генотип – фенотип*. В работе [Hoehndorf et al. 2012] рассматриваются вопросы семантической интеграции физиологических фенотипов с применением к онтологии клеточных фенотипов. В статье [Zhao et al. 2022] описывается разработка онтологии фенотипа расстройств аутистического спектра путем обработки естественного языка в электронных медицинских картах.

Персонализация

Второе перспективное направление содержит статьи, в которых описано применение онтологий для коммерции (покупка билетов), персональной рекламы, образования и персональной медицины.

Персонализация (по аналогии с компьютеризацией) коснулась таких разных сфер, как реклама, образование и медицина. Персонализация учитывает особенности и желания людей и позволяет сделать бизнес, обучение, лечение и рекламу более интеллектуальными и эффективными. Здесь также онтологии помогают формально и более строго классифицировать признаки клиентов, учащихся и предлагаемых услуг, что позволяет более точно выявлять их взаимосвязи при анализе больших объемов данных в электронной форме. Это направление имеет следующие ключевые слова, большинство из которых выделены красным цветом на рисунке 2: персонализированный (*personalized*), интерактивный (*interactive*), интеллектуальный (*smart*). Рассмотрим некоторые статьи из этого перспективного направления.

Работа [Ivanova et al. 2010] посвящена интеграции онтологии с разработкой персонализированных средств электронного обучения для дизлексиков. В ряде статей описаны: система персонализированной целевой рекламы на основе онтологий для интерактивного телевидения [Kim, Kang 2013]; предложение по автоматизации персонализированного домена бронирования авиабилетов путем интеграции безопасной нечеткой онтологии с мультиагентной платформой [Bukhari, Kim 2012]; применение биомедицинских онтологий в персонализированном здравоохранении для системы интеллектуальных запросов [Zhang et al. 2016]; платформа интеграции данных на основе онтологий, поддерживающая разработку прогностических моделей для персонализированной медицины [Weiler et al. 2018].

Алгоритмы кластеризации

Третье перспективное направление содержит статьи, описывающие алгоритмы кластеризации и их применение для различных задач, включая медицину, коммерцию, кластеризацию документов, автоматическое построение онтологий и установление соответствий между онтологиями. Алгоритмы кластеризации являются естественным методом решения задач, которые описаны выше в первых двух группах. Это направление имеет следующие ключевые слова, большинство из которых выделены красным цветом на рисунке 2: кластеризация (*clustering*), автоматизированный (*automated*). Рассмотрим некоторые статьи из этого перспективного направления.

В работе [Moshirpour et al. 2012] описывается автоматизированное построение онтологий на основе требований к программному обеспечению с использованием

методов кластеризации. В статье [Maleszka 2015] представлен метод кластеризации профилей с использованием выравнивания онтологий в персонализированных системах поиска документов. Работа [Sandhiya, Sundarambal 2019] посвящена кластеризации биомедицинских документов с использованием модели семантического сглаживания, обогащенной TF-IGM на основе онтологий, для телемедицинских приложений. В статье [Rojano et al. 2021] описывается кластеризация когорт генетических заболеваний на основе данных онтологии фенотипа человека.

Совместное управление

К четвертому направлению относятся статьи, в которых описано применение онтологий для управления задачами в совместных группах (коллективах), для интеграции знаний при совместной разработке продуктов, для управления сложными системами. В этих задачах онтологии помогают более точно и более формально описать признаки ситуаций, возникающих при управлении различными группами и объектами. В этом направлении представлены следующие ключевые слова, большинство из которых выделены синим цветом на рисунке 2: совместная работа (*collaborative*), управление (*managing*), редактирование (*editing*).

Рассмотрим некоторые статьи из выделенного перспективного направления, тематика которых представляется наиболее характерной для него. В исследованиях описаны: управление знаниями, необходимыми для поддержки электронного личного помощника, который является удобным для конечного пользователя инструментом редактирования графической онтологии [Einig et al. 2006]; подход на основе онтологий для интеграции знаний при совместной разработке продуктов [Yu et al. 2017]; управление устойчивостью умного города к вирусу Эбола с использованием онтологий [Liotine et al. 2016]; мобильное приложение на основе онтологий для управления задачами в совместных группах [Schmidt et al. 2016].

Заключение

Были выявлены четыре наиболее перспективных направления применения онтологий в области биомедицины: выявление связей между генотипом и фенотипом, персонализация различных услуг,

алгоритмы кластеризации, совместное управление. Исходя из полученных результатов представляется релевантным их использование в наукометрии в целом, поскольку с их помощью на основе анализа выборки из ключевых слов, соотносящихся с определенной научной областью, можно спрогнозировать изменение трендов в данной области либо появление новых трендов. Представляется, что предложенный в данном исследовании метод анализа научных трендов может быть применен к любой научной области (возможно, с определенными его адаптациями), а также для прогнозирования динамики их развития. Следовательно, предложенный в данном исследовании метод может использоваться и в науковедении для измерения динамики развития определенных научных областей, для изучения популярности и востребованности разделов определенной научной дисциплины, для определения факта, какие области исследований сейчас находятся на пике развития, а в каких постепенно снижается исследовательская активность. Тем самым результаты, полученные в данном исследовании, могут представлять и практический интерес для специалистов в частных областях науки и практики.

Конфликт интересов: Авторы заявили об отсутствии потенциальных конфликтов интересов в отношении исследования, авторства и / или публикации данной статьи.

Conflict of interests: The authors declared no potential conflict of interests regarding the research, authorship, and / or publication of this article.

Критерии авторства: М. М. Шарнин – общая концепция исследования, разработка алгоритмов, используемых в анализе, сбор и обработка фактического материала, проведение экспериментов, обсуждение результатов, написание текста статьи (60 %). С. С. Калинин – обсуждение результатов, обзор основного понятийного аппарата области формальных онтологий, написание текста статьи (40 %).

Contribution: M. M. Charnine developed the research concept and the algorithms used in the analysis, collection, and data processing, conducted the experiments, summarized the results, and drafted the manuscript (60%). S. S. Kalinin analyzed the results, reviewed scientific publications in formal ontology, and drafted the manuscript (40%).

Литература / References

- Бова В. В., Лещанов Д. В., Кравченко Д. Ю., Новиков А. А. Компьютерная онтология: задача и методология построения. *Информатика, вычислительная техника и инженерное образование*. 2014. № 4. С. 44–55. [Bova V. V., Leshchanov D. B., Kravchenko D. Yu., Novikov A. A. Computational ontology: Objectives and methodology of construction. *Informatika, vychislitelnaia tekhnika i inzhenernoe obrazovanie*, 2014, (4): 44–55. (In Russ.)] <https://www.elibrary.ru/ubnfff>
- Корсун И. А., Пальчунов Д. Е. Методы извлечения и формального представления онтологических знаний из текстов естественного языка. *Знания – Онтологии – Теории (ЗОНТ-2017)*: Всерос. конф. с Междунар. уч. (Новосибирск, 2–6 октября 2017 г.) Новосибирск: Дигит Про, 2017. Т. 2. С. 21–25. [Korsun I. A., Palchunov D. E. Methods of data retrieving from texts written in natural language and formal representation of ontological knowledges. *Knowledges – Ontologies – Theories (ZONT-2017)*: Proc. All-Russian Sci. Conf. with Intern. Participation, Novosibirsk, 2–6 Oct 2017. Novosibirsk: Digit Pro, 2017, vol. 2, 21–25. (In Russ.)] <https://www.elibrary.ru/eidazb>
- Митрофанова О. А., Азарова И. В. Компьютерное представление значений. *Прикладная и компьютерная лингвистика*, ред. И. С. Николаев, О. В. Митренина, Т. М. Ландо. 2-е изд. М.: Ленанд, 2017. С. 59–93. [Mitrofanova O. A., Azarova I. V. Computational representation of meanings. *Applied and computational linguistics*, eds. Nikolaev I. S., Mitrenina O. V., Lando T. M. 2nd ed. Moscow: Lenand, 2017, 59–93. (In Russ.)] <https://www.elibrary.ru/zcwtrj>
- Шарнин М. М. Библиометрический анализ трендовых тем в англоязычной научной литературе по лингвистике. *Российский лингвистический бюллетень*. 2023. № 12. [Charnine M. M. A bibliometric analysis of trending topics in the English language linguistics research literature. *Russian Linguistic Bulletin*, 2023, (12). (In Russ.)] <https://doi.org/10.18454/RULB.2023.48.17>
- Bukhari A. C., Kim Y.-G. Integration of a secure type-2 fuzzy ontology with a multi-agent platform: A proposal to automate the personalized flight ticket booking domain. *Information Sciences*, 2012, 198: 24–47. <https://doi.org/10.1016/j.ins.2012.02.036>
- Charnine M., Klovov A., Kochiev L., Tishchenko A. Research trending topic prediction as cognitive enhancement. *2021 International Conference on Cyberworlds (CW)*. Caen, 2021a, 217–220. <https://doi.org/10.1109/CW52790.2021.00044>
- Charnine M., Tishchenko A., Kochiev L. Visualization of research trending topic prediction: Intelligent method for data analysis. *CEUR Workshop Proceedings: Proc. 31st Intern. Conf. on Computer Graphics and Vision*, Nizhny Novgorod, 27–30 Sep 2021. 2021b, vol. 3027, 1028–1037. <https://www.elibrary.ru/ptcljf>
- Coulet A., Smail-Tabbone M., Benlian P., Napoli A., Devignes M. D. Ontology-guided data preparation for discovering genotype-phenotype relationships. *BMC Bioinformatics*, 2008, 9(S4). <https://doi.org/10.1186/1471-2105-9-s4-s3>
- Einig M., Tagg R., Peters G. Managing the knowledge needed to support an electronic personal assistant – an end-user friendly graphical ontology editing tool. *Proceedings of the Eighth International Conference on Enterprise Information Systems – ISAS*, 2006, vol. 5, 304–309. <https://doi.org/10.5220/0002446003040309>
- Hoehndorf R., Harris M. A., Herre H., Rustici G., Gkoutos G. V. Semantic integration of physiology phenotypes with an application to the Cellular Phenotype Ontology. *Bioinformatics*, 2012, 28(13): 1783–1789. <https://doi.org/10.1093/bioinformatics/bts250>
- Hoehndorf R., Loebe F., Kelso J., Herre H. Representing default knowledge in biomedical ontologies: Application to the integration of anatomy and phenotype ontologies. *BMC Bioinformatics*, 2007, 8. <https://doi.org/10.1186/1471-2105-8-377>
- Holloway E. From genotype to phenotype: Linking bioinformatics and medical informatics ontologies. *Comparative and functional genomics*, 2002, 3(5): 447–450. <https://doi.org/10.1002/cfg.181>
- Ivanova T., Andreev R., Terzieva V. Integration of ontology with development of personalized e-learning facilities for dyslexics. *Artificial Intelligence: Methodology, Systems, and Applications. AIMSA 2010*: Proc. 14th Intern. Conf., Varna, Bulgaria, 8–10 Sep 2010. Berlin Heidelberg: Springer, 2010, 265–266. https://doi.org/10.1007/978-3-642-15431-7_29
- Kim J., Kang S. An ontology-based personalized target advertisement system on interactive TV. *Multimedia Tools and Applications*, 2013, 64(3): 517–534. <https://doi.org/10.1007/s11042-011-0965-0>
- Liotine M., Ramaprasad A., Syn T. Managing a smart city's resilience to ebola: An ontological framework. *49th Hawaii International Conference on System Sciences (HICSS 2016)*. Kauai, Hawaii, USA, 2016, 2935–2943. <http://dx.doi.org/10.1109/HICSS.2016.368>

- Maleszka B. A method for profile clustering using ontology alignment in personalized document retrieval systems. *Computational Collective Intelligence. ICCCI 2015: Proc. 7th Intern. Conf.*, Madrid, Spain, 21–23 Sep 2015. Springer, 2015, 410–420. https://doi.org/10.1007/978-3-319-24069-5_39
- Moshirpour M., Mireslami S., Alhajj R., Far B. H. Automated ontology construction from scenario based software requirements using clustering techniques. *2012 IEEE 13th International Conference on Information Reuse & Integration (IRI)*. 2012, 541–547.
- Munir K., Anjum S. M. The use of ontologies for effective knowledge modelling and information retrieval. *Applied Computing and Informatics*, 2018, 14(2): 116–126. <http://dx.doi.org/10.1016/j.aci.2017.07.003>
- Prokhorenkova L., Gusev G., Vorobev A., Dorogush A. V., Gulin A. CatBoost: Unbiased boosting with categorical features. *Advances in neural information processing systems: Proc. Conf.*, Montreal, QC, 2–8 Dec 2018, 6638–6648. <https://www.elibrary.ru/nryevs>
- Rojano E., Córdoba-Caballero J., Jabato F. M., Gallego D., Serrano M., Pérez B., Parés-Aguilar A., Perkins J. R., Ranea J. A. G., Seoane-Zonjic P. Evaluating, filtering and clustering genetic disease cohorts based on human phenotype ontology data with cohort analyzer. *Journal of Personalized Medicine*, 2021, 11(8). <https://doi.org/10.3390/jpm11080730>
- Sandhiya R., Sundarambal M. Clustering of biomedical documents using ontology-based TF-IGM enriched semantic smoothing model for telemedicine applications. *Cluster Computing*, 2019, 22(S2): 3213–3230. <https://doi.org/10.1007/s10586-018-2023-4>
- Schmidt D., Pannison A., Freitas A., Bordini R. H., Meneguzzi F., Vieira R. An ontology-based mobile application for task managing in collaborative groups. *Proceedings of the 29th Florida Artificial Intelligence Research Society Conference (FLAIRS)*. Association for the Advancement of Artificial Intelligence (AAAI), 2016, 522–525.
- Weiler G., Schwarz U., Rauch J., Rohm K., Lehr T., Theobald S., Kiefer S., Götz K., Och K., Pfeifer N., Handl L., Smola S., Ihle M., Turki A. T., Beelen D. W., Rissland J., Bittenbring J., Graf N. XplOit: An ontology-based data integration platform supporting the development of predictive models for personalized medicine. *Studies in health technology and informatics*, 2018, 247: 21–25. <https://doi.org/10.3233/978-1-61499-852-5-21>
- Yu G., Yang Y., Qingsong X. An ontology-based approach for knowledge integration in product collaborative development. *Journal of Intelligent Systems*, 2017, 26(1): 35–46. <https://doi.org/10.1515/jisys-2015-0046>
- Zhang G., Bie R., Sun Y. Bring biomedical ontologies to personalized healthcare: A smart inquiry framework. *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. Washington, DC, USA, 2016, 84–88. <https://doi.org/10.1109/CHASE.2016.32>
- Zhao M., Havrilla J., Peng J., Drye M., Fecher M., Guthrie W., Tunc B., Schultz R., Wang K., Zhou Y. Development of a phenotype ontology for autism spectrum disorder by natural language processing on electronic health records. *Journal of Neurodevelopmental Disorders*, 2022, 14. <https://doi.org/10.1186/s11689-022-09442-0>