

информационно-советующей системы, позволит обеспечить достаточно высокую точность и скорость прогнозирования нагрузок.

ЛИТЕРАТУРА

1. Шаюхов, Т.Т. Математическое моделирование влияния внешних факторов на параметры электропотребления / Т.Т. Шаюхов // Наукоеведение. – 2017. – № 9.
2. Lozinskaia, A., Electricity consumption forecasting for integrated power system with seasonal patterns / A. Lozinskaia, A. Redkina, E. Shenkman // Applied Econometrics – 2020. – № 5. – С. 25.
3. Виноградов, Д.В. Автоматизация объектов электроэнергетики: состояние и пути развития / Д.В. Виноградов // Информационные технологии, системный анализ и управление (Итсау-2022): Сборник трудов XX Всероссийской научной конференции молодых ученых, аспирантов и студентов, Ростов-на-Дону, Таганрог, 01–03 декабря 2022 года. – Ростов-на-Дону, Таганрог: ДиректСайнс, 2022. – С. 329–332.
4. Бугаец, В.А. Краткосрочное прогнозирование электропотребления энергорайонов с учетом влияния метеофакторов: дис. канд. техн. наук. Новочеркасск. – 2015. – 241 с.
5. Карпенко, С.М. Эконометрическое моделирование энергопотребления с учетом влияния производственных факторов / С.М. Карпенко, Н.В. Карпенко // Энергобезопасность и энергосбережение. – 2020. – № 1. – С.14–17.
6. Статья «Методы прогнозирования электропотребления» / Официальный сайт «Евразийский союз учёных». Режим доступа: <http://euroasia-science.ru/> – Дата доступа: 10.11.2023.
7. Круглов В.В., Борисов В.В. Искусственные нейронные сети: 2-е издание. – 2020. – 380 с.

УДК 004.04, 004.41, 001.5

А.Д. Атрошкина, магистрант (БГЭУ, г. Минск);

Е. А. Севрук, специалист (ОАО «Гипросвязь», г. Минск)

АВТОМАТИЗИРОВАННЫЙ СБОР ДАННЫХ О НАУЧНЫХ ПУБЛИКАЦИЯХ

Введение. На современном этапе развития общества основным результатом исследовательской деятельности считается новое научное знание, содержание которого излагается преимущественно в опубликованных научных трудах. Именно поэтому задачи корректного сбора, обработки и подсчета информации о публикациях выходят на пер-

вый план в большинстве научно-исследовательских учреждений. Для решения указанных задач может использоваться автоматизированный сбор данных о научных публикациях, внедрение которого позволит собрать показатели результативности ученых, осуществить их мониторинг и комплексный учет.

В данной статье представлены механизмы автоматизированного сбора данных о научных публикациях на основании количества и частоты научных публикаций в социальной сети ResearchGate и научной электронной библиотеки КиберЛенинка.

Создание механизмов автоматизированного сбора данных о научных публикациях. При написании механизмов сбора информации ключевыми факторами выступали: цифровая форма представления информации и концепция «Открытая наука».

Первый механизм создан для платформы ResearchGate, структура и особенности которой заключаются в открытом доступе, высокой активности ученых на ресурсе, а также разнообразии цифрового контента, которое выражено в возможностях публикации авторами различного типа научно-технической информации (НТИ): предварительная статья (препринт), статья, тезис, презентация, наборы данных [1]. На данный момент на платформе ResearchGate зарегистрированы более 25 млн ученых из 192 стран мира, а совокупный объем материалов составляет 160 млн страниц НТИ [2].

Для разработки данного механизма были выбраны следующие технологии: язык программирования Python, система управления базами данных (СУБД) SQLite3, а также готовые библиотеки (Selenium, BeautifulSoup4 и др.). Также для визуализации структуры и динамики тематических цепочек платформы ResearchGate была использована библиотека Pyvis, включая программный пакет Networkx [3].

Схема хранения данных в базе данных представлена на рис. 1.

id	parent_id	name	link	count_publication	depth
Фил...	Фильтр	Фильтр	Фильтр	Фильтр	Филь...
1	0	Medicine	https://www.researchgate.net/topic/Medicine	585422	0
2	1	Allied Health Science	https://www.researchgate.net/topic/Allied-Health-Science	674	1
3	2	Allied Health Personnel	https://www.researchgate.net/topic/Allied-Health-Personnel	10000	2
4	0	Psychology	https://www.researchgate.net/topic/Psychology	804869	0
5	4	Clinical Psychology	https://www.researchgate.net/topic/Clinical-Psychology	207673	1
6	5	Mental Illness	https://www.researchgate.net/topic/Mental-Illness	513527	2

Рисунок 1 – Пример хранения результата в базе данных

Во время разработки механизма был выявлен ряд проблем. Например, в некоторых научных темах отсутствует количество публикаций, цепочки научных тем содержат в себе заикливание, кроме этого, зачастую по гиперссылке открываются страницы, которых не существует или были удалены.

Данная разработка позволила получить следующие результаты:

- количество тематик публикаций – 36 566;
- 1910 родительских разделов (тематик);
- среднее время обработки одной тематической ветви составляет 30 сек [1].

Помимо это была разработан механизм автоматизированного сбора данных о научных публикациях с платформы КиберЛенинка, которая представляет собой научную электронную библиотеку, построенную на парадигме открытой науки (Open Science), а также обеспечивает бесплатный оперативный доступ к научным публикациям в электронном виде. Количество посещаемости данной платформы достигает 72 млн человек [4].

Для разработки данной системы были выбраны следующие технологии: язык программирования JavaScript, программная библиотека Node.js, система управления базами данных (СУБД) SQLite3, а также готовые библиотеки (axios, cheerio) [5, 6].

В данной электронной библиотеке присутствует разделение научных публикаций на разделы: медицинские науки, естественные и точные науки, техника и технологии, гуманитарные науки, сельскохозяйственные науки, социальные науки. Помимо этого, каждый раздел включает в себя подразделы (всего 36).

База данных для хранения информации, полученной при автоматизированном сборе, хранит в себе следующие данные: название подраздела, название статьи, ссылка на статью, имена авторов, год публикации статьи, количество просмотров статьи, количество скачиваний, название журнала, в котором размещена данная статья, ключевые слова, время сбора данных, а также дополнительные данные о статье: BAK, Scopus, ESCI, RSCI, CAS. Схема хранения данных в базе данных представлена на рис. 2.

direction	name	link	authors	year	views
Filter	Filter	Filter	Filter	Filter	Filter
Фундаментальная ...	Проблемы внедрения ...	/article/n/problemy-vnedreniya-...	Вальдман Е.А.	2008	3833
Фундаментальная ...	Нейроиммунологические ...	/article/n/neyroimmunologicheski...	Гайнетдинова Д.Д.,Хакимова ...	2011	4479
Фундаментальная ...	Поиск и изучение ...	/article/n/poisk-i-izuchenie-...	Баженов Л.Г.	2009	2890
Фундаментальная ...	Повышение малыми дозам...	/article/n/povyshenie-malymi-...	Городецкая И.В.	2004	5225
Фундаментальная ...	Онкоиммунология, ...	/article/n/onkoimmunologiya-...		2009	9668

downloads	journal	keywords	bak	scopus	esci	rsci	cas	time
Filter	Filter	Filter	Filt...	Filter	Filt...	Filt...	Fil...	Filter
429	Ремедиум. Журнал о ...	ФАРМАКОГЕНЕТИКА,ИССЛЕД...	1	0	0	0	0	2023-12-17 16:13:43
615	Медицинская ...	КЛЮЧЕВЫЕ СЛОВА: ДЕТСКИЙ...	1	1	0	0	0	2023-12-17 16:13:43
339	Сибирский ...		1	1	0	0	0	2023-12-17 16:13:43
560	Вестник Витебского ...	тиреоидные ...	0	0	0	0	0	2023-12-17 16:13:43
1080	Медицинская ...		1	1	0	0	0	2023-12-17 16:13:44

Рисунок 2 – Пример хранения результата в базе данных

На данный момент собрано:

- 29 подразделов из 36;
- 1 933 130 научных статей.

Использование полученных данных. Данные о научных публикациях, полученных в ходе автоматизированного сбора, возможно использовать преследуя различные цели. Рассмотрим их более подробно:

1. Для *анализа публикационной активности ученых* (по количеству публикации). С помощью полученной информации возможно оценить эффективность научных исследований и выявить наиболее активных ученых в определенной научной области.

2. Для *анализа популярности научных статей* (по количеству просмотров, скачиваний и т.д.). Появляется возможность оценить в реальном времени, что интересует научную общественность, тем самым это поможет авторам научных статей адаптировать свои работы под «запрос» аудитории и определить перспективные направления для дальнейших исследований. Помимо этого, доступно выделение трендовых тематик по годам, что дает возможность проследить тенденцию изменения интересов исследователей.

3. Для *анализа эффективности научных журналов* (по количеству публикаций в журнале), путем сравнения количества просмотров и скачиваний статей, которые были опубликованы в различных журналах. Данные, полученные в ходе указанного анализа, отразят журналы, пользующиеся наибольшей популярностью среди научного сообщества.

4. Для *анализа тенденций в научных исследованиях* (по изменению тематик научных публикаций, используемых методов исследований и т.д.). С помощью данного анализа, возможно выявить новые направления научных исследований и определить их перспективы, что позволяет отразить тенденции в развитии конкретной области исследований.

Помимо анализа вышеуказанных показателей, данные можно применять для *создания профилей ученых и научных журналов*. В перспективе профиль может содержать в себе информацию о количестве публикаций, рейтинге журнала, уровне цитируемости и т. д. Кроме этого, для авторов, создание таких профилей облегчит поиск новых публикационных площадок.

Вместе с тем, с помощью полученных данных возможно создать *рекомендации по чтению научных статей*, основанных на таких факторах, как автор и тема статьи, цитируемость работы, рейтинг журнала и т.д. Такие рекомендации позволяют лучше ориентироваться в

бесконечном множестве научной литературы, подборе наиболее актуальных и авторитетных источников информации.

Заключение. Несмотря на возможности отслеживания тенденций практически по всем научным направлениям, созданные механизмы представляют наибольший интерес для авторов в сфере цифрового развития, поскольку подобные решения позволяют на ранних этапах отслеживать формирование новых технологий, а также обеспечить быстрое реагирование на их изменения.

ЛИТЕРАТУРА

1. Севрук Е.А., Качан Д.А., Юневич Н.Г. Выделение трендов цифрового развития с помощью автоматизированной аналитики частоты и тематики научных публикаций // Сборник докладов XXII Международной конференции «РИНТИ-2023». – 2023. – С. 148–153.

2. ResearchGate: Find and share research [Electronic resource]. – URL: <https://www.researchgate.net/>. – Дата доступа: 03.01.2024.

3. Interactive network visualizations. Pyvis Documentation [Electronic resource]. – URL: <https://pyvis.readthedocs.io/en/latest/> – Дата доступа: 03.01.2024.

4. Cyberleninka.ru [Electronic resource] // Similarweb: Website Traffic – Check and Analyze Any Website: [сайт]. – URL: <http://east-front.narod.ru/memo/latchford.htm> – Дата доступа: 03.01.2024.

5. Axios: [сайт]. – URL: <https://axios-http.com/docs/intro>.

6. Cheerio.js: [сайт]. – URL: <https://cheerio.js.org/>.

УДК 004. 032.26

Д.В. Шиман, доц., канд. техн. наук;
А.Н. Мушук, ассист.; Е.А. Гончар, ассист.
(БГТУ, г. Минск)

ЭФФЕКТИВНОСТЬ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ СОЦИАЛЬНЫХ ЯВЛЕНИЙ

На данный момент существует большое количество различных алгоритмов машинного обучения. Каждый из них был создан под разные задачи. В свою очередь, задачи делятся на подвиды. Одним из которых является прогнозирование.

В целом, практически любой алгоритм можно использовать для решения любого типа задач. Отличия состоят лишь в том, что точность работы алгоритма будет слегка ниже, чем если бы использовался предназначенный для конкретной задачи, однако иногда этим мож-