

Автоматизация поиска научных работ сотрудников вуза с применением методов искусственного интеллекта

Н. И. Галлини¹, П. В. Четырбок²

Гуманитарно-педагогическая академия (филиал)
ФГАОУ ВО «КФУ им. В.И. Вернадского» в г. Ялте
¹gallini.nadi@yandex.ru, ²petr58@mail.ru

Н. С. Мотуз

Донской государственный технический университет
motuz0nik@yandex.ru

Аннотация. Автоматизация поиска научных работ сотрудников вуза с применением методов искусственного интеллекта необходима для упрощения сбора сведений о научной деятельности профессорско-преподавательского состава и обучающихся. Концептуальная модель интегрированной интеллектуальной системы научно-исследовательской деятельности вуза, представленная в данной статье, визуализирует рабочие процессы научной деятельности вуза, на базе которого проводится научное исследование. В процессе разработки интегрированной интеллектуальной системы научно-исследовательской деятельности вуза используется модуль Scikit-Learn и Keras (Keras).

Ключевые слова: автоматизация поиска научных работ, искусственный интеллект, научно-исследовательской деятельности вуза, Scikit-Learn, Keras

I. ВВЕДЕНИЕ

Актуальность автоматизации поиска научных работ сотрудников вуза с применением методов искусственного интеллекта очевидна, так как в процессе ведения научной деятельности современные ученые сотрудничают как на межвузовском уровне, так и на международном и используя современные технологии можно улучшить аналитическую обработку информации без использования ручного ввода информации.

Сложность задач и высокий темп развития научных экспериментов, всё более расширяющийся спектр использования автоматизации задач информационно-аналитических процессов вызывает потребность в оперативных алгоритмах поиска для парсинга данных, при этом в запросах необходимо учитывать такие параметры, как тип данных Big Data, стандартизацию и подлинность данных.

Цель исследования разработать интегрированную интеллектуальную систему научно-исследовательской деятельности вуза Аксиома.

Задачами исследования являются:

1. Разработать концептуальную модель интегрированной интеллектуальной системы научно-исследовательской деятельности вуза.
2. Описать работу разработанной и обученной нейронной сети, основанной на модуле Scikit-Learn.
3. Представить интегрированную интеллектуальную систему научно-исследовательской деятельности вуза Аксиома.

II. КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ ДЕЯТЕЛЬНОСТИ ВУЗА

Практика применения информационных технологий в вузах сформировала идеологию, необходимую для создания информационных технологий и процессов как информационной системы. В современном мире необходимо автоматизировать процесс сбора информации о научно-исследовательской деятельности вузов, для чего разрабатываются интеллектуальные системы, которые являются синтезом программного, лингвистического и логико-математического набора средств программирования для реализации главной задачи – поиска информации о научной работе сотрудников и обучающихся вузов в автоматическом режиме.

Для обеспечения целостности данных необходимо организовать единое пространство научно-исследовательской деятельности университетов и научно-исследовательских институтов. По статистическим данным в России насчитывается 607 государственных и 358 негосударственных университетов, в которых обучается 4,7 млн человек. После относительного спада в середине 90-х годов число обучающихся ежегодно растет. Так как каждый обучающийся должен вести активную научную деятельность под руководством профессорско-преподавательского состава кафедр, то в год выходит колоссальное количество публикаций, разрабатываются тысячи проектов и грантов.

Таким образом, наблюдается необходимость автоматизации сбора данных о научно-исследовательской деятельности обучающихся и профессорско-преподавательского состава при научно-исследовательской деятельности с использованием нейронной сети и технологий Big Data. Для визуализации информационных процессов была разработана концептуальная модель интегрированной интеллектуальной системе научно-исследовательской деятельности университета с учетом организационной структуры образовательной организации высшего образования, Гуманитарно-педагогическая академия (филиал) «Крымского федерального университета имени В.И. Вернадского» в г. Ялте.

Концептуальная модель, которая представлена в данной статье, разрабатывалась для визуализации рабочих процессов научной деятельности вуза. В структуру входят базы данных, такие как IEEE Xplore

Digital Library, из которой поступает информация о статьях в Scopus и WoS; ELIBRARY, веб-сервер, который соединяет базы данных с тонким клиентом интегрированной интеллектуальной системы научно-исследовательской деятельности вуза. Также разрабатываемой интегрированной интеллектуальной системе научно-исследовательской деятельности вуза необходимо иметь доступ к базам данных Российского научного фонда (РНФ) и Платформы «Россия – страна возможностей» объединяющей 26 конкурсов, проектов и олимпиад (РСВ), так как подача заявок на конкурсы тоже важная часть научной деятельности ППС и обучающихся. В случае необходимости можно подключить и другие базы данных.

Основная функция разрабатываемой интегрированной интеллектуальной системы научно-исследовательской деятельности вуза это автоматическая визуализация результатов и формирование отчетности. При этом для составления необходимых отчетов пользователю необходимо только зарегистрироваться в системе, указав такие данные, как ФИО и электронный адрес, данный подход значительно ускорит и упростит сбор необходимых данных [1].

С помощью интегрированной интеллектуальной системы научно-исследовательской деятельности университета будет организован доступ к данным о научно-исследовательской работе обучающихся и сотрудников в структурные подразделения: отдел кадров (отвечает за актуальность информации при приеме на работу профессорско-преподавательского состава и зачисления обучающихся, что обуславливает необходимость отслеживания информации о научно-исследовательской деятельности сотрудников и обучающихся), институты (необходимо обеспечивать всей информацией о научной деятельности обучающихся и сотрудников института), колледж (необходимо обеспечивать всей информацией о научной деятельности обучающихся и сотрудников колледжа), центры (необходимо обеспечивать всей информацией о научной деятельности сотрудников центров), учебно-методический отдел (необходимо обеспечивать информацией о научной деятельности профессорско-преподавательского состава (ППС) для заполнения таблиц кадрового обеспечения основных образовательных программ), научный отдел (необходимо обеспечивать всей информацией о научной деятельности обучающихся и сотрудников университета), администрация (контролируют процесс научно-исследовательской деятельности университета), так как вуз, на базе которого происходит разработка описываемой системы является филиалом на концептуальной модели обозначен блок головного вуза с логотипом КФУ. Связи между этими структурными подразделениями показаны на концептуальной модели данной технологии (рис. 1).

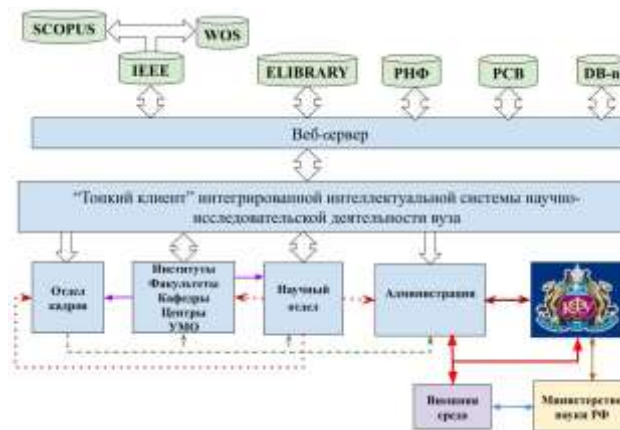


Рис. 1. Концептуальная модель интегрированной интеллектуальной системы научно-исследовательской деятельности вуза

Таким образом, решаются проблемы координации деятельности ППС и обучающихся, занимающихся научно-исследовательской деятельностью университета, в структурные подразделения, а также вопросы обеспечения их всей необходимой информацией. В то же время административный персонал получает своевременный доступ к актуальным данным о научно-исследовательской деятельности и располагает инструментами для быстрого формирования всей необходимой отчетности [7, 8].

III. ПРИМЕНЕНИЕ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ РАЗРАБОТКЕ ИНТЕГРИРОВАННОЙ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ ДЕЯТЕЛЬНОСТИ ВУЗА

Повышение эффективности управления электронной информацией научной деятельности вуза прямо пропорционален автоматизации рабочих процессов в данной области, для чего проектируются и разрабатываются новые методики использования систем и технологий. В методе, представленном в данном исследовании, применяются синтез технологий автоматической обработки цифрового информационного моделирования и веб-технологии базовой платформы. Предполагается, что использование автоматического управления электронной информацией научной деятельности вуза будет способствовать развитию интеллектуальной среды сотрудников вузов [10].

Научная информация, загруженная в базы данных, размещенные в Интернете, является важнейшим источником для научной деятельности вузов, при этом составление отчетности в вузе, на базе которого проводится эксперимент, до сих пор формируется ППС вручную, что очень неудобно и занимает много времени. Разрабатываемую систему можно использовать не только для составления отчетности, но и для формирования рейтинга ППС и обучающихся вуза. В том случае, если систему будут использовать несколько вузов одновременно, можно будет формировать межвузовский рейтинг.

Интегрированная интеллектуальная система научно-исследовательской деятельности вуза разрабатывается в Scikit-learn, являющимся модулем Python, объединяющем широкий спектр современных

алгоритмов машинного обучения для задач среднего масштаба с учителем и без учителя. Этот пакет ориентирован на предоставление машинного обучения неспециалистам с использованием языка высокого уровня общего назначения. Акцент делается на простоте использования, производительности, документации, и согласованность API. Он имеет минимальное количество зависимостей и распространяется под упрощенной лицензией BSD, что поощряет его использование как в академических, так и в коммерческих целях. Исходный код, двоичные файлы и документацию можно загрузить с <https://scikit-learn.org> [2].

Использование пакетов программного обеспечения и библиотек Scikit-Learn, которые были разработаны для ускорения запуска моделирования с использованием нескольких известных классификаторов и алгоритма программного машинного обучения, в котором различные классификаторы получают различную производительность, зависят от типа данных, используемых для входных данных этапа классификации, при этом, работая с постоянными данными, надлежит осуществить их дискретизацию. Также при анализе данных необходимо учитывать, что взаимосвязь данных зависит не только от дискретизации, так как используемые данные чаще всего большие и неточные, а представление знаний помогает исследователям осмыслить данные в процессе дискретизации.

Так как на этапе дискретизации данного исследования используется нечеткая логика и влияние нечетких элементов на точность классификаторов Scikit-learn и результата моделирования с помощью нечеткого графа принадлежности, лингвистических переменных и значений дискретных интервалов то алгоритм, полученный в результате показал необходимость производительности для работы обученной нейронной сети. X-кратной перекрестной проверки используются классификаторы пакетов Scikit-learn [6, 9].

Методы поиска научной деятельности ППС и обучающихся обычно имеют проблему пропуска, и на поиск необходимой информации требуется много времени. В алгоритме, объединяющем последовательность поиска данных на основе простого правила матричного преобразования, определяются ключевые слова для поиска необходимой информации и коэффициентов распределения потока мощности. Используемый метод повышает точность расчета за счет охвата переменных и отсутствия предварительной настройки зоны поиска.

Можно избежать сложных логических операций алгоритма обхода, принятых в большинстве традиционных методов, и эффективно повысить скорость вычислений. Кроме того, алгоритм также имеет характеристику памяти, которую можно применять для онлайн-поиска ключевых участков передачи с помощью IEEE Metadata Search API (OpenAPI).

Для поиска метаданных в IEEE Xplore® в Аксиоме используется IEEE Metadata Search API (OpenAPI) с помощью которого разработчики могут запрашивать и извлекать записи метаданных, включая аннотации для более чем 5 миллионов документов, размещенных в IEEE Xplore®, включая журналы, материалы

конференций, книги, курсы и стандарты. API метаданных поддерживает как простой, так и логический поиск. Также можно сформировать запрос и получение полного текста статьи, которые обозначены в открытом доступе [3, 4, 5].

IV. ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА УЧЕТА И СИСТЕМАТИЗАЦИИ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ ДЕЯТЕЛЬНОСТИ ВУЗА

Интеллектуальная система учета и систематизации научно-исследовательских деятельности вуза для помощи принятия управленческих решений и эффективного использования информационных технологий, получила название «Аксиома».

Основной целью Аксиомы является автоматический сбор, хранение, анализ и систематизация информации о научно-исследовательской деятельности вуза с целью подготовки и принятия решений.

Задачами Аксиомы являются:

1. Проектирование и моделирование механизмов, развивающих научно-исследовательскую и педагогическую деятельность в вузе, основываясь на объективную оценку деятельности обучающихся и преподавателей.
2. Организация единого информационно-аналитического пространства научно-исследовательской деятельности в Интернете, для повышения рейтинговых показателей вуза
3. Создание современной интеллектуальной, интегрированной системы для управления научно-исследовательской деятельностью вуза.

Рассмотрим алгоритм работы Аксиомы. Для регистрации в Аксиоме пользователю необходимо ввести фамилию, имя и отчество на русском и английском языках и электронный адрес, сделать выбор баз данных в которых необходимо искать научную активность. После регистрации Аксиома самостоятельно выбирает все выходные данные о публикациях из баз данных и заполняет отчет. Также в Аксиоме учитывается возможность ручного ввода данных о публикациях, которые еще не отразились в базах данных, но у автора есть справка о том, что публикация будет издана. Так как каждая публикация имеет свои баллы. Аксиома позволяет автоматизировать рейтинг ППС и обучающихся.

Рейтинг ППС нужен для продления эффективного контракта и для получения выплат, которые зависят от научной активности ППС за последние полгода. По этой причине каждому преподавателю довольно часто приходится вручную заполнять разнообразные отчеты для подтверждения своей научной деятельности.

Обучающимся рейтинг необходим для получения повышенной стипендии и после каждой сессии необходимо заполнить очень большой пакет документов для его подтверждения, и на подготовку документов дается очень маленький период времени. Поэтому автоматизация подтверждения научной активности обучающихся является актуальной задачей.

Для большей заинтересованности обучающихся научной деятельностью в Аксиоме используется метод

игрофикации научных исследований. Игрофикация – это применение принципов игры и элементов игрового дизайна в неигровых контекстах. Например, игровые темы и принципы, такие как накопление виртуальных «баллов», а также выполнение поставленных задач или действий для перехода на следующий уровень, могут использоваться в учебном процессе для стимулирования студентов [8].

Метод игрофикации научных исследований позволяет обучающимся проще формировать «команды» для работы над публикациями. Компоненты игрофикации охватывают весь учебный процесс, изучая игровой мир, студент выступает в роли персонажа, развивается, получает новые знания, достижения, игровую валюту и опыт, перемещается в визуальной среде и взаимодействует в социальной системе. Мир является образовательной системой, которая через задачи, загадки и головоломки формирует образ мышления, которое направленно на совершенствование необходимых навыков.

Игрофикация способствует монетизации проекта. Монетизация встречается в изначально бесплатных играх, потому что они многопользовательские и требуют постоянного обновления и поддержки. На 2022 год есть много способов монетизации игр. Чаще всего это «скины» (игровые раскраски на различные вещи или же, к примеру, персонаж другой внешности). Самым популярным и прибыльным способом является так называемый «Боевой пропуск», они работают по сезонной системе. В современных играх есть как бесплатная версия, так и платная версия такого пропуска.

В бесплатной версии пропуска игрок может получать награды, но не каждый уровень пропуска, и не такие редкие, как в платной. К тому же в платной версии можно будет получить персонажа раньше всех и вдобавок эксклюзивный скин на него, который позже невозможно будет получить. Поэтому у игроков возникает желание купить именно платную версию. За счет монетизации Аксиомы авторы исследования будут поддерживать улучшение системы.

V. ЗАКЛЮЧЕНИЕ

Интеллектуальная система учета и систематизации научно-исследовательских деятельности вуза Аксиома, предназначена для получения, структуризации информации, визуализации данных и автоматизации заполнения отчетов из иных информационных систем, связанных с научной деятельностью ППС и обучающихся вузов. Внедрение данной системы в каждом вузе будет способствовать обмену информацией о научной деятельности вузов, тематиках исследований ППС, что позволит облегчить взаимодействие между вузами.

Разработанная и обученная нейронная сеть основывается на модуле scikit. Благодаря этому модулю

машина будет правильно и быстро обучена. Далее будет строиться сверточная нейронная сеть, основанная на модуле Keras (Keras), для дальнейшего прогнозирования и экстраполяции данных.

Благодаря интегрируемой интеллектуальной системе научно-исследовательской деятельности администрациям университетов будет предоставлена информация для принятия решений, какую научно-исследовательскую деятельность вести в университете для получения наибольшей экономической эффективности.

СПИСОК ЛИТЕРАТУРЫ

- [1] A. Abu Jabal, E. Bertino and G. De Mel, "Provenance-Based Scientific Workflow Search," 2017 IEEE 13th International Conference on e-Science (e-Science), 2017, pp. 119-127, doi: 10.1109/eScience.2017.24.
- [2] A. Kolte, B. Mahitha and N. V. G. Raju, "Stratification of Parkinson Disease using python scikit-learn ML library," 2019 International Conference on Emerging Trends in Science and Engineering (ICESE), 2019, pp. 1-4, doi: 10.1109/ICESE46178.2019.9194627.
- [3] A.N. Kokoulin, A.A. Yuzhakov and D.A. Kiryanov, "Development of efficient fault-tolerant storage for multidimensional scientific data using the hough transform," 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EConRus), 2017, pp. 690-694, doi: 10.1109/EConRus.2017.7910650.
- [4] D. Zhang et al., "Fast Search Algorithm for Key Transmission Sections Based on Topology Converging Adjacency Matrix," in IEEE Access, vol. 8, pp. 108235-108246, 2020, doi: 10.1109/ACCESS.2020.3001077.
- [5] J. Machii, J. Murumba and E. Micheni, "Predictive Analytics and Artificial Intelligence in Blended Learning: A New Dawn for Institutions of Higher Learning," 2021 IST-Africa Conference (IST-Africa), 2021, pp. 1-14.
- [6] M.N.F. Hishamuddin, M.F. Hassan, D.C. Tran and A.A. Mokhtar, "Improving Classification Accuracy of Scikit-learn Classifiers with Discrete Fuzzy Interval Values," 2020 International Conference on Computational Intelligence (ICCI), 2020, pp. 163-166, doi: 10.1109/ICCI51257.2020.9247696.
- [7] N.I. Gallini, A. A. Denisenko, D. T. Kamornitskiy, P. V. Chetyrbok and N. S. Motuz, "Research on the Use of a Neural Network in the Intelligence System of University Research Activities," 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EConRus), 2021, pp. 886-890, doi: 10.1109/EConRus51938.2021.9396216.
- [8] N.I. Gallini, D.T. Kamornitskiy, M.A. Latysheva, P.V. Chetyrbok and L.S. Anisimova, "Gamification Components of a Software Solution Using a Neural Network in the Design of Unified Information and Analytical Space of a University," 2021 International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS), 2021, pp. 629-633, doi: 10.1109/ITQMIS53292.2021.9642740
- [9] Susanto, D. Stiawan, M.A.S. Arifin, M.Y. Idris and R. Budiarto, "IoT Botnet Malware Classification Using Weka Tool and Scikit-learn Machine Learning," 2020 7th International Conference on Electrical Engineering, Computer Sciences and Informatics (EECSI), 2020, pp. 15-20, doi: 10.23919/EECSI50503.2020.9251304.
- [10] Y. Chen, "Research on the Application Strategy of Automation Technology in Electronic Information Engineering Design System," 2020 12th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 2020, pp. 231-235, doi: 10.1109/ICMTMA50254.2020.00059.