

РАЗРАБОТКА ИНСТРУМЕНТОВ АНАЛИТИКИ ДАННЫХ МЕЖДУНАРОДНЫХ НАУКОМЕТРИЧЕСКИХ СЕРВИСОВ

М. А. Митрохин, Д. А. Данилин

DEVELOPMENT OF DATA ANALYTICS TOOLS FOR INTERNATIONAL SCIENTOMETRIC SERVICES

M. A. Mitrokhin, D. A. Danilin

Аннотация. Описаны базы знаний наукометрических сервисов Scopus, Web of Science, eLIBRARY.RU, Mendeley, SCImago Journal & Country Rank. Рассмотрены возможности пользовательских интерфейсов данных сервисов, частота обновления данных, особенности получения доступа к информации, принципы их работы. Описаны возможности существующих интерфейсов программного получения данных (API) рассматриваемых сервисов. Обоснована целесообразность разработки программных модулей получения данных о публикационной активности сотрудников наукоемких предприятий, преподавателей и обучающихся образовательных организаций с помощью описанных API. Представлены результаты разработки сервиса аналитического сопровождения публикационной активности сотрудников Пензенского государственного университета.

Ключевые слова: аналитика, наукометрия, обработка данных, API, публикации, наука, автоматизация.

Abstract. The knowledge base of scientometric services Scopus, Web of Science, eLIBRARY.RU, Mendeley, SCImago Journal & Country Rank are described. The possibilities of user interfaces of these services, the frequency of data updates, the features of accessing information, the principles of their work are considered. The capabilities of the existing software data acquisition interfaces (APIs) of the considered services are described. It is concluded that it is advisable to develop software modules for obtaining data on the publication activity of employees of high-tech enterprises, teachers and students of educational organizations using the described APIs. The experience of Penza State University in the development of analytical support services for the publication activity of employees is presented.

Keywords: analytics, scientometrics, data processing, API, publications, science, automation.

Введение

Одним из основных процессов управления научными и научно-образовательными организациями является управление знаниями (knowledge management) [1]. Частной задачей управления знаниями является аналитическое сопровождение публикационной активности сотрудников (обучающихся). Эта задача является одной из наиболее актуальных и значимых в организации научно-исследовательской деятельности. Эффективный сбор, анализ и представление наукометрических данных позволяют не только оперативно и точно оценивать текущее состояние научного сектора учебного заведения или НИИ, но и применять экономико-организационные решения с целью по-

вышения как уровня публикационной активности, так и качества публикуемых материалов. Но информационно-аналитическое сопровождение публикационной активности сотрудников и обучающихся – трудоемкий процесс, связанный с обработкой большого объема данных. Одним из способов оптимизации этого процесса (уменьшения времени актуализации данных, повышения точности формируемых показателей, исключения дублирования данных и т.д.) является использование программных способов получения первичных данных о научных трудах. Такой подход позволяет не только снизить нагрузку на человеческие ресурсы (исключение необходимости ручного ввода), но и исключить необходимость модерации вводимых данных (ввиду их гарантированной достоверности, обеспечиваемой издательствами и наукометрическими сервисами), а также автоматизировать саму загрузку обрабатываемых данных в информационную систему организации, не выделяя ее в отдельный производственный процесс.

Материал и методика

В настоящее время наибольшей популярностью в Российской Федерации пользуются наукометрические сервисы: Scopus, Web of Science, ResearcherID, eLIBRARY.RU, Mendeley, SCImago Journal & Country Rank.

Scopus – крупнейшая международная база данных, содержащая аннотации и информацию о рецензируемой издательским домом Elsevier научной литературе, имеющая встроенные инструменты анализа, отслеживания и визуализации данных [2]. Данные Scopus используются для составления таких международных рейтингов университетов, как QS, THE, Shanghai University Rankings [3]. Метрики данного сервиса также активно используются министерствами и ведомствами, курирующими научную и инновационную деятельность (не только РФ, но и других государств), а также осуществляющими финансирование НИР фондами. Scopus отличается глубиной и широтой охвата содержимого, а также ежедневной актуализацией данных. В базе Scopus содержится более 23 700 изданий (из них 280 – специализированные, 4000 журналов в открытом доступе) более 5000 международных издателей в области гуманитарных, общественных, естественных наук, а также техники, медицины и искусства. Кроме того, данный сервис хранит более 167 000 книг, ежегодно добавляя более 20 000 книг и 560 книжных серий, более 8,3 млн докладов конференций на 100 000 международных конференциях. Отличительной особенностью Scopus является возможность хранения публикаций в стадии доредакционной подготовки «Articles-in-Press» из 8000 журналов. Общее же число записей в базе знаний составляет 71 млн (из них 64 млн записей с 1969 г., остальные – до 1970 г., наиболее ранняя из которых датируется 1788 г.). Общее число патентных записей – более 39 млн от пяти мировых патентных ведомств. Scopus имеет собственный интерфейс программного предоставления данных – любой желающий может получить ключ API и использовать его бесплатно (при соблюдении политики использования API и данных). Вся информация, получаемая через API, имеет подробный вид, описанный в технической документации сервиса. Так, зная ID автора, можно получить все его публикации, а также расширенную информацию о публикациях: дата, издание, авторы, цитирование и т.д. При необходимости же получения большей

информации (издание, аффилиция авторов и т.д.) возможно оформление подписки на платный сегмент API.

Web of Science представляет собой поисковый веб-сервис, объединяющий реферативные базы данных публикаций в научных журналах и патентов (в т.ч. базы, учитывающие взаимное цитирование публикаций) [4]. Web of Science охватывает материалы по общественным, техническим, естественным, гуманитарным наукам и искусству. Платформа обладает встроенным поисковиком с возможностью анализа и управления библиографической информацией. Данный наукометрический сервис также предоставляет бесплатный API и поддерживает расширенный поиск по всем материалам для получения метаданных на уровне элементов, включая цитируемость и аффилицию авторов. Кроме того, возможно обнаружение связанных работ.

В число инструментов обработки данных Web of Science входит ResearcherID – международная идентификационная система для глобального учета уникальных авторов научных статей [5]. После добавления публикаций в профиль ResearcherID происходит автоматическая привязка соответствующих публикаций в Web of Science (создается прямая ссылка из Web of Science на профиль ResearcherID). Использование данного сервиса позволяет сформировать полный список статей автора, которые есть в базе Web of Science, а также получать статистические данные, по которым можно формировать инфографику. Однако существенным минусом данного инструмента является отсутствие API.

Mendeley – веб-сервис и бесплатная программа для управления данными, позволяющая просматривать и хранить исследовательские работы в формате PDF и имеющая подключение к международной социальной сети научных сотрудников [6]. Объем хранимых сервисом документов составляет порядка 100 млн работ. Данные Mendeley позволяют оценить общее число публикаций в целевом научном направлении, динамику изменения этого числа в разные годы, выявить наиболее просматриваемые публикации и их аудиторию по странам. Сервис позволяет выбрать как бесплатный базовый пакет услуг, так и платный, предоставляющий увеличенные квоты на хранение материалов и создание групп. С помощью API Mendeley, предоставляемого на бесплатной основе, возможно получение информации об атрибутах, размещенных в базе документов: список ключевых слов, аннотацию, а также число просмотревших этот документ ученых, их места работы, страны проживания, должности и т.д.

eLIBRARY.RU является крупнейшей в России электронной библиотекой научных публикаций, обладающей обширными возможностями анализа и поиска научной информации [7]. Библиотека интегрирована с Российским индексом научного цитирования (РИНЦ) – бесплатным общедоступным инструментом измерения публикационной активности ученых и организаций, созданным по заказу Министерства науки и высшего образования РФ [8]. База данных eLIBRARY.RU содержит более 29 млн публикаций и научных статей, в т.ч. электронные версии более чем 5600 (4800 – в открытом доступе) российских научно-технических журналов. Платная подписка eLIBRARY.RU предоставляет пользователю доступ к полнотекстовой коллекции из более 1100 ведущих российских журналов. При этом идентификация пользователей

с оформленной подпиской в библиотеке осуществляется на основе внешних IP-адресов компьютеров, с которых ведется работа (список IP-адресов, относящихся к организации-лицензиату, модерируется только официально утвержденным представителем организации в его административном интерфейсе). Личный кабинет пользователя (в т.ч. без подписки) позволяет осуществлять настройку извещений по электронной почте, настройку панели навигатора, истории поисковых запросов, вносить изменения в персональную карточку, формировать подборки статей и т.д. Несмотря на широкие возможности визуального интерфейса сервиса, долгое время у библиотеки не было интерфейса программного получения данных. Однако в 2019 г. был осуществлен релиз eLIBRARY.RU API, доступ к которому является платным (минимальная цена – 25 000 руб.) [9].

SCImago Journal & Country Rank (SJCR) – общедоступный аналитический портал, который предоставляет наукометрические показатели по журналам и странам, а также отображает статистику цитирования журналов и рейтинги публикационной активности [10]. Основными показателями, размещаемыми на ресурсе, являются SCImago Country Rank (SCR) и SCImago Journal Rank (SJR). SCR учитывает многие факторы: индекс Хирша, взвешенные показатели цитирований по годам, общее количество опубликованных цитирований, статей. SJR (показатель престижности или влияния журнала) отображает среднее число цитат, полученных в текущем году на статьи, опубликованные в журналах за три предыдущих года [11]. Данный сервис позволяет сравнивать рейтинги как отдельных журналов, так и стран в целом. Данные о журналах могут быть сгруппированы по тематическим областям (27 основных тематических областей), тематическим категориям (313 конкретных тематических категорий) или по странам. Объем обрабатываемых SJCR данных составляет порядка 34 100 изданий от более чем 5000 международных издателей из 239 стран. Обновление данных осуществляется 2 раза в год. Данный сервис не предоставляет API, однако позволяет отображать основные метрики журнала в веб-интерфейсе с помощью встраиваемого кликабельного виджета.

В условиях существования API крупных международных наукометрических сервисов (Scopus, Web of Science, eLIBRARY.RU), повсеместной цифровизации современной научно-образовательной отрасли, а также отсутствия готовых решений, ориентированных на интеграцию с информационными системами вуза или научной организации, становится актуальной разработка ПО автоматизированного получения и анализа соответствующих данных.

***Разработка веб-сервиса аналитического сопровождения
публикационной активности сотрудников
Пензенского государственного университета***

Одним из этапов разработки соответствующего программного обеспечения (ПО) является выбор архитектуры его построения. Среди наиболее распространенных архитектур можно выделить нативное приложение, независимый веб-сервис, решение на базе конфигурируемой системы или фреймворка, а также веб-модуль расширения функционала уже используемых систем. Сравнительная характеристика указанных подходов приведена в табл. 1.

Таблица 1

Сравнительная характеристика архитектур построения ПО

Подход	Плюсы	Минусы
Нативное приложение	– возможность буферного хранения обрабатываемой информации на стороне клиента и отложенной синхронизации данных; – возможность использования аппаратных функций устройства; – возможность работы при отсутствии сетевого соединения	– необходимость разработки ПО для каждой целевой платформы и операционной системы; – необходимость доставки объемных обновлений на устройства пользователей; – сложность однообразного и одновременного изменения бизнес-логики сервиса, реализуемой на стороне пользователя
Отдельный веб-сервис	– легкость доставки информационной услуги до пользователя и обновления функционала сервиса; – кроссплатформенность	– необходимость дублирования на стороне сервера информации, возможно, хранимой в уже эксплуатируемых программно-вычислительных комплексах организации ; – необходимость разработки не только целевого функционала, но и типовых модулей и подсистем (авторизация, взаимодействие с БД и т.д.); – необходимость наличия постоянного сетевого соединения
Фреймворк	– наличие уже готовых служебных модулей ПО	– необходимость приобретения лицензии; – ограниченный функционал, заложенный возможностями базового ядра системы; – отсутствие гарантированной кроссплатформенности
Веб-модуль расширения функционала	– наличие уже готовых служебных модулей ПО; – легкость доставки информационной услуги до пользователя и обновления функционала сервиса; – кроссплатформенность	– необходимость адаптации разработки к уже существующему ПО; – необходимость наличия постоянного сетевого соединения

В условиях существования уже используемых на предприятии или в образовательной организации информационных систем наиболее рациональным подходом можно считать разработку веб-модуля расширения функционала. Его достоинства сочетают в себе гибкость адаптации существующих

модулей к новой задаче с кроссплатформенностью и простотой использования конечным пользователем, а зависимость построенного таким образом ПО от наличия постоянного сетевого соединения может быть частично компенсирована кэшированием данных на стороне клиента.

Именно такой подход был выбран для разработки веб-сервиса аналитического сопровождения публикационной активности сотрудников Пензенского государственного университета – «Scopus PSU»¹.

Проект «Scopus PSU» представляет собой комплекс программных средств, позволяющих интегрировать данные международного наукометрического сервиса Scopus в информационное пространство ПГУ. Данный сервис позволяет автоматизировать загрузку данных о публикациях, проиндексированных Scopus, в электронную библиотеку ПГУ, проводить аналитику показателей публикационной активности сотрудников и подразделений (количество публикаций, их цитируемость, индекс Хирша авторов научных работ и т.д.), а также осуществлять выгрузку соответствующих показателей в графическом (диаграммы) и табличном виде.

Работа веб-сервиса «Scopus PSU» основана на взаимодействии электронной библиотеки ПГУ из состава электронной информационной образовательной среды (ЭИОС) с программным интерфейсом Elsevier API, предоставляющим данные наукометрического сервиса Scopus [12]. Схема взаимодействия элементов сервиса представлена на рис. 1.

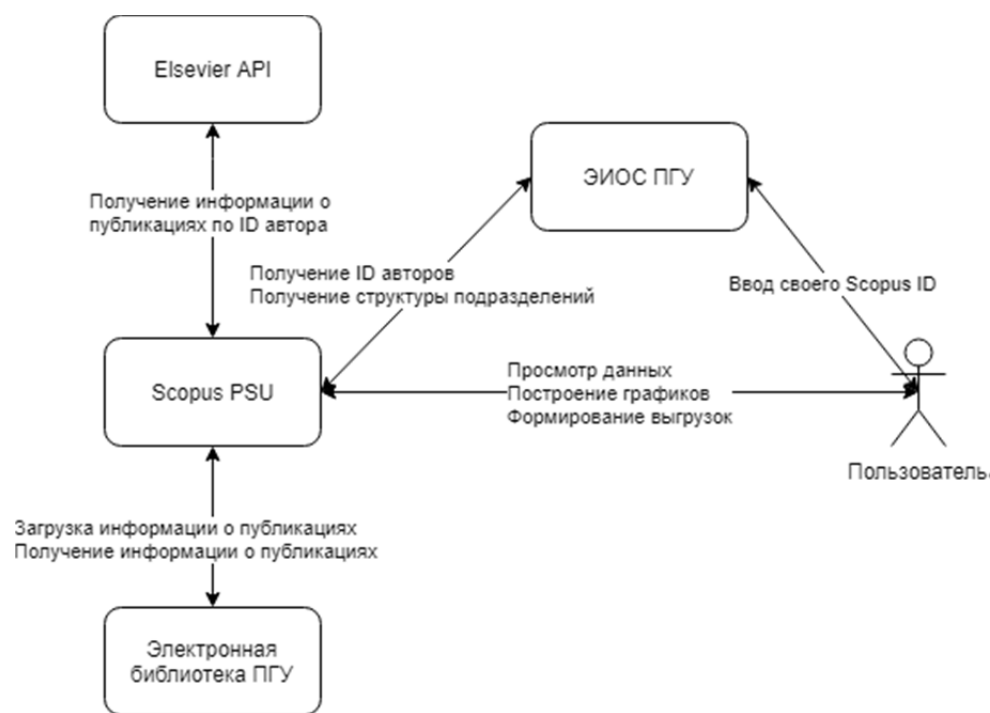


Рис. 1. Схема взаимодействия элементов веб-сервиса «Scopus PSU»

¹ Данный проект стал победителем конкурса «Ректорские гранты – 2019» (Договор на проведение НИР № 16,18,19/19 НИР) и сейчас находится на этапе внутреннего тестирования.

Результаты

Использование данного сервиса позволяет отказаться от процесса ручного ввода данных о научных работах и дальнейшей их верификации, и требует для своей работы лишь ввода сотрудником его уникального идентификатора Scopus в личном кабинете ЭИОС ПГУ. При этом привязка данных к иерархии подразделений университета, заложенной в ЭИОС, не требует создания дополнительных инструментов администрирования: при редактировании данных о месте работы или иной информации о сотруднике данные обновляются автоматически.

Пользовательский интерфейс страницы просмотра данных о подразделениях приведен на рис. 2, где область экрана № 1 – история перехода по иерархии подразделений вуза, область № 2 – панель быстрого поиска и перехода к подразделениям, область № 3 – данные о текущем подразделении, область № 4 – данные о подчиненных подразделениях, область № 5 – панель фильтрации данных по времени, типу публикации и т.д.

1. Бreadcrumb: Политехнический институт > Факультет приборостроения, информационных технологий и электроники

2. Поиск: Введите название подразделения [Найти]

3. Текущее подразделение: Факультет приборостроения, информационных технологий и электроники

Показатель	Значение
Всего авторов	6
Всего статей авторов	146
Всего статей подразделений	120
Количество цитирований	244

4. Таблица подразделений:

Подразделение	Всего авторов	Всего статей авторов	Всего статей подразделения
Кафедра "Конструирование и производство радиоприборов"	1	14	14
Кафедра "Автономные информационные и управляющие системы"	1	4	4
Кафедра "Физика"	2	88	62
Кафедра "Информационно-измерительная техника и метрология"	2	40	40

5. Фильтры:

- За период: [Выбор периода]
- По: 05.08.2019
- Тип публикации: ☒ Все, ☒ Article, ☒ Conference Paper, ☒ Review
- Показать все: [Обновить]
- Сбросить

Рис. 2. Пользовательский интерфейс страницы просмотра данных о подразделениях веб-сервиса «Scopus PSU»

Отображаемые данные о публикационной активности при этом включают в себя количество авторов в подразделении, количество публикаций за период, количество цитирований работ авторов. Также на странице подразделения отображается список подчиненных подразделений (название, количество авторов, количество публикаций) и список сотрудников непосредственного подчинения (ФИО, количество публикаций, их цитирований и индекс Хирша). При этом для каждого подразделения возможно скачивание электронной таблицы с соответствующими данными о нем в удобном для пользователя формате: .ods, .csv, .xls, .xlsx.

Пользовательский интерфейс страницы формирования графиков публикационной активности приведен на рис. 3.

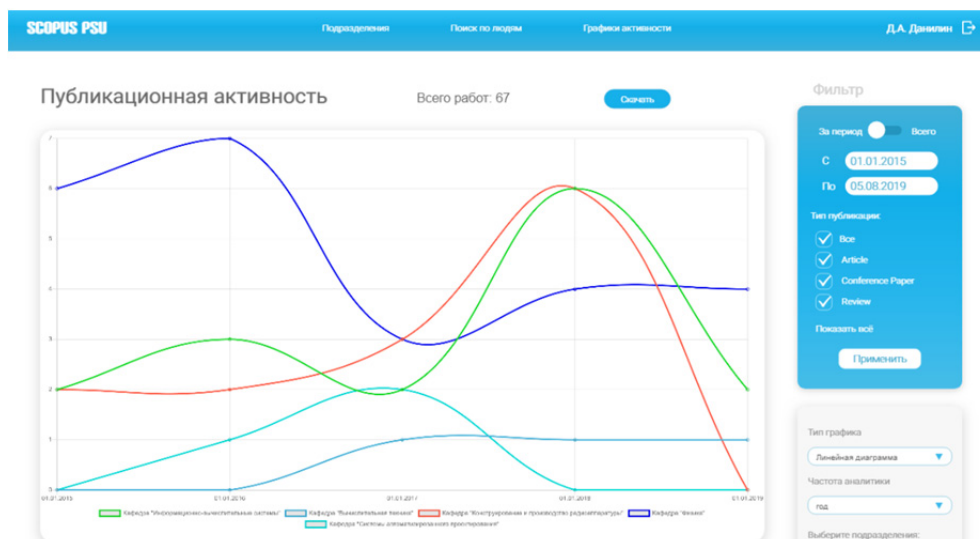


Рис. 3. Пользовательский интерфейс страницы формирования графиков публикационной активности веб-сервиса «Scopus PSU»

Обсуждение

Функционал представления данных сервисом включает в себя не только текстовое отображение информации о конкретном сотруднике или подразделении (а также формирование электронных таблиц с выгрузками данных о них и аффилированных публикациях), но и построение графиков и диаграмм на ее основе. При этом пользователь может выбрать необходимый тип диаграммы (круговая, линейная, столбчатая и др.), длительность единичного периода (месяц, 2 месяца, полгода, год и т.д.) для диаграмм, демонстрирующих динамику публикационной активности во времени (столбиковая, линейная и др.).

Загрузка данных с помощью API Scopus выполняется в фоновом режиме с заданной периодичностью, настраиваемой в зависимости от масштаба учебного заведения, интенсивности публикации научных работ его сотрудников, а также в соответствии с установленными еженедельными лимитами запросов к API (так, например, в Пензенском государственном университете синхронизация данных выполняется раз в неделю). Получение данных осуществляется путем выдачи API Scopus списка всех работ (а также данных по каждой публикации) автора по его уникальному идентификатору Scopus ID, который вводится в электронную информационную систему. При этом учитываются как ситуации, когда в системе еще нет информации о публикации, так и случаи, когда данные уже присутствуют (дополняются наиболее актуальными) или в системе присутствуют дубли записей (выполняется их объединение и дополнение). Данный подход позволяет постепенно перейти от подхода ручного заполнения электронной библиотеки университета к ее дополнению недостающей информацией без слома существующих бизнес-процессов и проблемы согласования разработанной системы с уже существующим банком данных.

Выводы

В результате реализации проекта разработан веб-сервис аналитического сопровождения публикационной деятельности сотрудников Пензенского государственного университета, позволивший автоматизировать ввод внешних наукометрических данных, а также обеспечить их удобное представление и выгрузку для дальнейшей обработки.

В настоящее время веб-сервис прошел этап внутреннего тестирования и в настоящий момент, после исправления выявленных недостатков, используется в деятельности научно-инновационного управления ПГУ.

Архитектура разработанного ПО позволяет проводить в дальнейшем расширение числа обрабатываемых наукометрических сервисов, имеющих средства программной выдачи данных (в первую очередь – eLIBRARY.RU), а также обеспечивает возможность его адаптации для работы в составе информационных систем большинства высших учебных заведений Российской Федерации.

Библиографический список

1. Blunt, R. Knowledge Management in the New Economy / R. Blunt. – iUniverse, 2001.
2. The largest database of peer-reviewed literature – Scopus. Elsevier. – URL: <https://www.elsevier.com/solutions/scopus> (дата обращения: 10.09.2019).
3. Москалева, О. В. Рейтинги университетов и научные журналы / О. В. Москалева // Научная периодика: проблемы и решения. – 2014. – № 4 (22). – С. 9–20.
4. Панкрашин, В. В. Индекс научного цитирования Web of Science / В. В. Панкрашин // Наука и технологии трубопроводного транспорта нефти и нефтепродуктов. – 2011. – № 3. – С. 102–104.
5. Акоев, М. А. Руководство по наукометрии: индикаторы развития науки и технологии. Thomson Reuters / М. А. Акоев, В. А. Маркусова, О. В. Москалева, В. В. Писляков. – Екатеринбург, 2014.
6. Reference Manager and Academic Social Network – Mendeley Database. Elsevier. – URL: <https://www.elsevier.com/solutions/mendeley> (дата обращения 01.09.2019).
7. Костюкова, М. В. О проекте «Российский индекс научного цитирования» / М. В. Костюкова // Педагогический журнал Башкортостана. – 2010. – № 4 (29). – С. 125–133.
8. Российский индекс научного цитирования – eLIBRARY.RU. – URL: https://elibrary.ru/project_risc (дата обращения: 15.09.2019).
9. Интерфейс программирования API – eLIBRARY.RU. – URL: https://elibrary.ru/projects/api/api_info (дата обращения: 15.09.2019).
10. SJR – About US. Scimago Journal & Country Rank. – URL: <https://www.scimagojr.com/aboutus> (дата обращения: 13.09.2019).
11. Балацкий, Е. В. Международные рейтинги университетов: практика составления и использования / Е. В. Балацкий, Н. А. Екимова // Экономика образования. – 2012. – № 2. – С. 67–80.
12. Кревский, И. Г. Электронная информационно-образовательная среда вуза как ключевой элемент его информатизации / И. Г. Кревский, А. В. Антонов // Новые информационные технологии и системы : сб. науч. ст. XIV Междунар. науч.-техн. конф., посвященной 70-летию кафедры «Вычислительная техника» и 30-летию кафедры «Системы автоматизированного проектирования». – 2017. – С. 365–367.

References

1. Blunt R. *Knowledge Management in the New Economy*. iUniverse, 2001.
2. *The largest database of peer-reviewed literature – Scopus*. Elsevier. Available at: <https://www.elsevier.com/solutions/scopus> (accessed Sept. 10, 2019).
3. Moskaleva O. V. *Nauchnaya periodika: problemy i resheniya* [Scientific periodicals: problems and solutions]. 2014, no. 4 (22), pp. 9–20. [In Russian]
4. Pankrashin V. V. *Nauka i tekhnologii truboprovodnogo transporta nefi i nefteproduktov* [Science and technology of pipeline transportation of oil and oil products]. 2011, no. 3, pp. 102–104. [In Russian]
5. Akoev M. A., Markusova V. A., Moskaleva O. V., Pislyakov V. V. *Rukovodstvo po naukometrii: indikatory razvitiya nauki i tekhnologii*. Thomson Reuters [Guide to scientometry: indicators of the development of science and technology. Thomson Reuters]. Ekaterinburg, 2014. [In Russian]
6. *Reference Manager and Academic Social Network – Mendeley Database*. Elsevier. Available at: <https://www.elsevier.com/solutions/mendeley> (accessed Sept. 01, 2019).
7. Kostyukova M. V. *Pedagogicheskiy zhurnal Bashkortostana* [Pedagogical journal of Bashkortostan]. 2010, no. 4 (29), pp. 125–133. [In Russian]
8. *Rossiyskiy indeks nauchnogo tsitirovaniya – eLIBRARY.RU* [The Russian index of scientific citation eLIBRARY.RU]. Available at: https://elibrary.ru/project_risc (accessed Sept. 15, 2019). [In Russian]
9. *Interfeys programmirovaniya API – eLIBRARY.RU* [Programming interface API – eLIBRARY.RU]. Available at: https://elibrary.ru/projects/api/api_info (accessed Sept. 15, 2019). [In Russian]
10. *SJR – About US. Scimago Journal & Country Rank*. Available at: <https://www.scimagojr.com/aboutus> (accessed Sept. 13, 2019).
11. Balatskiy E. V., Ekimova N. A. *Ekonomika obrazovaniya* [Economics of education]. 2012, no. 2, pp. 67–80. [In Russian]
12. Krevskiy I. G., Antonov A. V. *Novye informatsionnye tekhnologii i sistemy: sb. nauch. st. XIV Mezhdunar. nauch.-tekhn. konf., posvyashchennoy 70-letiyu kafedry «Vychislitel'naya tekhnika» i 30-letiyu kafedry «Sistemy avtomatizirovannogo proektirovaniya»* [New information technologies and systems : collection of scientific works. article XIV international. science.- tech. conf. dedicated to the 70th anniversary of the Department of "Computer engineering" and the 30th anniversary of the Department «Computer-aided design Systems»]. 2017, pp. 365–367. [In Russian]

Митрохин Максим Александрович
доктор технических наук, профессор,
заведующий кафедрой вычислительной
техники,
Пензенский государственный
университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: mmax83@mail.ru

Mitrokhin Maxim Alexandrovich
doctor of technical sciences, professor,
head of sub-department
of computer engineering,
Penza State University
(40 Krasnaya street, Penza, Russia)

Данилин Дмитрий Андреевич
студент,
Пензенский государственный
университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: timondimon98@gmail.com

Danilin Dmitriy Andreevich
student,
Penza State University
(40 Krasnaya street, Penza, Russia)

Образец цитирования:

Митрохин, М. А. Разработка инструментов аналитики данных международных наукометрических сервисов / М. А. Митрохин, Д. А. Данилин // Модели, системы, сети в экономике, технике, природе и обществе. – 2019. – № 4 (32). – С. 85–95.