МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(ВИНИТИ РАН)

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г. **№ 7** **Москва 2018**

ИНФОРМАЦИОННЫЙ АНАЛИЗ

УДК 004.774.056 – 047.44:[001:061(100)]

Ю.М. Брумштейн, Е.Ю. Васьковский

Сайты международных ассоциаций организаций, работающих в научно-технической сфере: анализ функциональности, вебометрических показателей, роли в научном информационном пространстве

Показано место международных научных ассоциаций (союзов, объединений, федераций) в общей структуре научных организаций, роль таких ассоциаций в формировании научного информационного пространства – международного и национальных.

Для рассматриваемых ассоциаций проанализированы общеметодологические вопросы построения их сайтов, информационного наполнения страниц сайтов, обеспечения необходимой функциональности и удобства работы пользователей с сайтами, соблюдения мер информационной безопасности.

Рассмотрены источники и состав информации, размещаемой на сайтах международных ассоциаций организаций, работающих в научно-технической сфере; приведены особенности структуры конкретных сайтов, их функциональных возможностей. Охарактеризованы применяемые на этих сайтах дизайнерские и программно-технические решения. Выполнен подробный сравнительный анализ вебометрических показателей, указаны основные направления их использования при принятии решений российскими исследователями и их группами, а также контент-менеджерами сайтов российских НИИ, вузов, научных обществ.

Ключевые слова: *научная информация, международное сотрудничество, ассоциации организаций, интернет-сайты, агрегация информации, доступность информации, вебометрические показатели, web-аналитика, поведение интернет-пользователей, потоки научной информации, управление потоками, информационная безопасность.*

ВВЕДЕНИЕ

В предыдущих работах авторов было выполнено категорирование сайтов, связанных с информационным и сервисным обеспечением научной деятельности; рассмотрена функциональность и вебометрические показатели ведущих зарубежных и российских сайтов-агрегаторов научной информации – политематической [1] и специализированной [2]. В настоящей статье исследована проблематика сайтов международных научных ассоциаций (союзов, объединений, федераций), относящихся к научно-технической сфере. В том числе изучен состав размещенной на сайтах информации и ее доступность, функциональность и вебометрические показатели сайтов. Далее термин «ассоциации» будем использовать как обобщающий. Это позволит различить объединения организаций и сами эти организации. Указанные ассоциации могут рассматриваться как одна из подкатегорий для категории «К13» по классификации введенной в [3]. Деятельность рассматриваемых ассоциаций важна для решения следующих задач: обеспечение необходимых базовых условий для функционирования международного научного информационного пространства (НИП) [4–6] в сфере научных исследований и разработок; поддержка развития национальных НИП [6–8], в том числе в относительно недавно возникших государствах [9]; расширение международного научно-исследовательского и научно-технологического сотрудничества [10]; информационная поддержка реализации отдельных международных проектов и программ [11]; накопление (агрегация) научно-технической информации – в том числе научных публикаций, объявлений о международных конгрессах и конференциях, результатов проведения этих мероприятий; информационная поддержка создания и развития интеллектуального потенциала для обеспечения устойчивого социально-экономического развития отдельных стран и их групп [3]; координация направлений научной деятельности на международном и национальном уровнях; управление процессами мониторинга [12] и обмена [13] НТИ в рамках НИП; подготовка молодых ученых (исследователей); информационная поддержка международной научной мобильности исследователей [14] и др.

Для сайтов ассоциаций актуальность изучения номенклатуры, назначения, функциональных возможностей, контента, вебометрических показателей сайтов определяется, прежде всего, необходимостью усиления интеграции России в международное научное информационное пространство [5], прежде всего за счет использования возможностей современных информационно-телекоммуникационных технологий. Однако в существующей литературе для этих ассоциаций тематика, связанная с построением и использованием сайтов и, особенно, с их вебометрическими показателями, исследована неполно. Поэтому цель настоящей статьи – комплексный анализ указанной проблематики с позиций интересов российских исследователей, научных организаций, вузов, органов государственного управления научной деятельностью.

Рассматриваемые сайты выполняют следующие основные функции: служат агрегаторами информации, предназначенной для публичного доступа; являются своеобразными «коммутационными узлами» сети распространения такой информации за счет наличия входящих, внутренних и исходящих гиперссылок; обеспечивают возможности для налаживания научных контактов между ассоциациями, организациями, отдельными учеными.

Основные практические направления использования сведений по зарубежным ассоциациям научных организаций и их сайтам, изучаемых в настоящей статье:

1) расширение информационной поддержки администраторов/контент-менеджеров российских сайтов при принятии решений о размещении гиперссылок на зарубежные сайты, отдельные материалы на них и т.д.;

2) ознакомление российских исследователей с составом полезной информации, имеющейся на сайтах этих ассоциаций, а также с расположением информации на их страницах, возможностями поиска информации на сайтах;

3) использование результатов анализа применяемых на рассматриваемых сайтах дизайнерских и программно-технических решений в отношении элементов меню, расположения контента, управления им и т.д. для унификации разрабатываемых российских интернет-ресурсов с зарубежными. Это, в свою очередь, может дать такие преимущества: применение уже апробированного зарубежного опыта для улучшения качества российских сайтов, сокращения сроков их разработки, обеспечения их информационного продвижения в Интернете; повышение удобства использования англоязычных страниц российских сайтов зарубежными пользователями и, как следствие, улучшение посещаемости таких страниц.

ХАРАКТЕРИСТИКА МАТЕРИАЛА ИССЛЕДОВАНИЙ И НАПРАВЛЕНИЙ ЕГО АНАЛИЗА

Объектами анализа были сайты международных ассоциаций организаций, работающих в научно-технической сфере – преимущественно по математике, физике, информационным технологиям. При отборе ассоциаций для анализа учитывалась их известность в НИП, потенциальная полезность информации на их сайтах для российских исследователей. Для уменьшения объема настоящей статьи в отдельные работы вынесена тематика по сайтам ассоциаций и национальных организаций по ряду «специализированных» направлений исследований/разработок: ин-

новационные, ядерные, космические; медицинские, биологические, экологические и т.п. Данные о сайтах брались, прежде всего, с самих сайтов. При необходимости использовались сведения из Википедии [15] и научных статей, сведения о программных разработках, информация о специализированных ресурсах-агрегаторах [16] и другие.

На рассматриваемых в статье сайтах в открытом доступе находятся как собственно научная информация, так и инженерно-техническая, познавательная, научно-популярная, а также информация гуманитарного характера. Размещение патентно-технической информации практически не встречается. Отметим, что в формировании потоков НТИ стран участвуют и иные виды информации: секретная, служебная и пр. Возможности (правила) размещения в открытом доступе единиц НТИ могут регулироваться на международном уровне (прежде всего в отношении соблюдения «авторских прав» на такие материалы), на внутригосударственном уровне, на ведомственном или внутрикорпоративном – для служебной информации. Иногда специально допускаются «дозированные» утечки информации (включая НТИ), размещение на сайтах ее фрагментов, новостных сообщений (иногда несколько тенденциозных), планов будущих разработок и исследований. Для сферы собственно научной деятельности умышленная дезинформация со стороны организаций в отношении полученных результатов, фактических достижений и пр., как правило, не характерна.

Рациональным приемом анонсирования результатов научных исследований без полного раскрытия их содержания может быть размещение на сайтах только названий выполненных работ и/или их рефератов (аннотаций). В этом случае по совокупности названий (рефератов) можно сделать определенные выводы: об общей направленности деятельности ассоциации/организации (или группы организаций); о направлениях персональной научной активности отдельных исследователей и/или их групп (по данным о них в публикациях); о возможностях примененных исследователями средств вычислительной техники и составе программного обеспечения, научного оборудования; иногда – о составе «не публикуемых» работ (путем сопоставления с мировыми потоками НТИ в соответствующих предметных областях) и т.д. Кроме того, может быть информативен анализ динамики изменения во времени объемов НТИ по отдельным направлениям, размещаемой на конкретных сайтах.

Направления и методика исследований, особенности представления результатов

Оценивалось следующее: юридическая принадлежность рассматриваемых сайтов, их назначение, состав размещенной информации, ее источники, функциональность сайтов, применяемые дизайнерские и программно-технические решения, вебометрические показатели сайтов, подходы к вопросам управления «информационной безопасностью» персональных данных.

Определение показателей сайтов осуществлялось по методике, подробно описанной в [8] с дополнениями, указанными в [9]. Приводимые далее оценки вебометрических показателей относятся к периоду с 01.01.2018 г. по 08.02.2018 г. Используемые обозначения для показателей подробно расшифрованы в [8, 9].

Аналогично [9] оценки количеств страниц сайтов с помощью информационно-поисковой системы (ИПС) Google во всех приводимых далее таблицах с показателями для сайтов даны в угловых скобках («< > »). При этом необходимо учитывать различия в оценках вебометрических показателей сайтов по методикам из [8] и на основе использования Интернета [9]. Отметим, что если вход на сайт требует ввода логина и пароля, то внешние программные средства фактически не позволяют проанализировать объемы и содержимое сайтов. Кроме того, на большинстве общедоступных сайтов есть закрытые паролями секции, включая «личные кабинеты» пользователей, разделы «рабочих групп» и пр. Поэтому внешние программные средства обычно дают заниженные результаты в отношении объемов размещенной на сайтах информации.

Существующие программные средства для вебометрического анализа (платные и бесплатные) позволяют определить общее количество и состав входящих и исходящих ссылок исследуемых сайтов. Для конкретного сайта анализ состава входящих ссылок позволяет определить: страницы «лидирующие» по количествам ссылок; распределение ссылок по национальным (соответствующим отдельным странам) и интернациональным доменам (org, com и пр.). Для совокупности исходящих с сайта ссылок также можно определить распределение их по доменам. Если ссылок относительно немного, то для оценки их распределений по доменам может использоваться следующее: импорт полученной таблицы ссылок в MsExcel, сортировка строк по столбцу с адресами ссылок, ручной подсчет количеств ссылок, относящихся к отдельным доменам. Автоматизировать такой подсчет с помощью автономных программных средств может быть достаточно сложно.

На сайтах рассматриваемых в статье ассоциаций преобладают исходящие ссылки на организаций-участников, сайты тематически смежных ассоциаций, профильных (по тематике) национальных научных обществ. Ссылки на сайты национальных Академий наук встречаются редко. Практически отсутствуют ссылки на личные сайты отдельных ученых, материалы в Википедии.

Также авторам не удалось найти бесплатных программных средств, которые бы в автоматическом режиме проводили анализ «дублированности» материалов исследуемых сайтов на других интернет-ресурсах (в «ручном режиме» такой анализ является очень трудоемким). Потенциально этот анализ может проводиться для получения следующих показателей:

1) доля единиц хранения информации (ЕХИ) на сайте, которые дублируются хотя бы на одном внешнем по отношению к нему интернет-ресурсе;

2) среднее количество «дублей» (копий) – в расчете на все размещенные на сайте ЕХИ (или только на дублирующиеся ЕХИ);

3) выявление ЕХИ на сайте, имеющих на внешних интернет-ресурсах наибольшее количество дублей (копий).

Обнаружение дублирующихся единиц хранения информации может быть важно в отношении имущественных авторских прав владельцев сайтов на размещенные материалы.

В статье числовые значения вебометрических показателей, соответствующие непосредственно сайтам изучаемых ассоциаций (их названия приведены в левых колонках каждой из таблиц статьи), даются без {}. Если «группы страниц» ассоциаций размещены на сайтах других организаций, то показатели, соответствующие «сайтам в целом», указываются внутри «{ }». Комбинация фигурных и угловых скобок (в виде {< >}) соответствует количеству страниц для «сайта в целом», полученному с помощью ИПС Google. В таблицах даны абсолютные количества ссылок. Относительные показатели (нормированные на количества страниц сайтов) не приводятся.

Причины создания наднациональных (международных) ассоциаций в сфере научных исследований и разработок

Основные причины создания наднациональных научных ассоциаций (объединений, союзов, федераций и пр.):

- желание повысить эффективность использования имеющихся ресурсов (интеллектуальных, финансовых, материально-технических, информационных) – особенно при реализации крупных международных проектов/программ;
- в ряде случаев – дефицит (или отсутствие) необходимых специалистов на национальных уровнях;
- объективная целесообразность углубления взаимодействия между странами, организациями, отдельными исследователями за счет координации планируемых/выполняемых работ, объединение усилий (ресурсов) при работе над группами проектов, отдельными проектами, частями этих проектов [17];
- стремление исключить неоправданное дублирование проводимых исследований;
- необходимость улучшения информационного менеджмента [13] для уже проводимых и перспективных исследований – в т.ч. за счет концентрирования (агрегации) НТИ, ее структуризации и управления информационными потоками.

Этот перечень не относится к исследованиям/разработкам, проводимым конкурирующими корпорациями, к исследованиям по «закрытой тематике» в оборонных целях и т.п.

Типичные варианты интеграции научно-исследовательской деятельности (НИД) на международном уровне:

1. Создание наднациональных организаций с собственными значительными бюджетами, основными фондами, штатами сотрудников, территориями. Такие организации реализуют проекты, номенклатура и графики выполнения которых согласованы со всеми участниками, в том числе и с целью исключения конфликтов за «доступ» к общим ресурсам для проведения исследований. В качестве примера приведем Европейский Центр Ядерных Исследований (ЦЕРН).

2. Создание международных ассоциаций (союзов, советов, комитетов, объединений и пр.), осуществляющих, в основном, координирующие функции и редакционно-издательскую деятельность, а также проводящих международные мероприятия.

3. Разделение (распределение) видов работ по НИД (а также их обеспечению НТИ) между странами на основе межгосударственных соглашений больших групп стран, представленных их национальными научными организациями и/или их объединениями.

4. Частичная координация НИД (а также процессов ее обеспечения НТИ, методическими материалами, стандартами, программными средствами, базами данных и пр.) в рамках межгосударственных соглашений небольших групп стран или двух стран. Сюда же отнесем обмен экспериментальными данными, данными наблюдений и т.п.

5. Координация НИД на основе соглашений (договоров) между отдельными организациями из нескольких стран.

6. Координация деятельности на основе двусторонних соглашений организаций из разных стран, в том числе и в отношении взаимного обмена НТИ, научными результатами.

7. Координация работ на основе соглашений между отдельными исследователями или их рабочими группами, в том числе не оформленными в виде юридически значимых договоров.

Все перечисленные варианты способствуют интенсификации транснациональных потоков НТИ, в том числе между организациями и отдельными исследователями; появлению дополнительных гиперссылок на интернет-сайтах и т.п.

Роль центров координации научно-исследовательской деятельности и/или управления качеством ее результатов на международном уровне выполняют также редакции авторитетных интернациональных научных журналов, оргкомитеты крупных международных научных конференций (с отбором участников), некоторые фонды-грантодатели.

Отметим, что во времена СССР (являвшегося лидером группы социалистических стран) также предпринимались попытки создания интегрированных структур по НИД – преимущественно в рамках стран Совета Экономической Взаимопомощи.

Для направлений исследований, носящих «закрытый характер», создание международных объединений, как правило, исключается. При этом информация о полученных результатах обычно не распространяется (в частности, не публикуются в открытой печати и не размещаются в Интернете сведения о «секретных» исследованиях, разработках, изобретениях и т.д.), за исключением «дозированных утечек». Такая ситуация приводит к параллельному выполнению в различных странах дорогостоящих исследований и разработок, только часть из которых в дальнейшем может быть использована в «гражданских целях».

Организация интерфейсов с пользователями на сайтах ассоциаций

В большинстве случаев для поддержки работы с различными устройствами дистанционного доступа к информации (включая мобильные) разработчиками сайтов применяются адаптивные интерфейсы [2], предусматриваются возможности скроллирования страниц.

Для страниц сайтов возможен просмотр HTML-кодов и Java-скриптов. Эти скрипты предназначены для выполнения действий на пользовательских устройствах и не содержат информации о программных средствах, применяемых на стороне сервера. Поэтому HTML-коды и Java-скрипты могут использоваться как некоторые готовые «образцы для подражания» разработчиками других сайтов. Доступность этой информации практически никак не влияет на уровень уязвимости к хакерским атакам самих сайтов, размещенных на серверах.

Для всех рассматриваемых ниже сайтов на их стартовых страницах (СС) не оговаривается необходимость использования определенных браузеров или их версий. Это может свидетельствовать о том, что сайты «совместимы» [18] с различными браузерами.

На сайтах ассоциаций глобального или континентального уровней интерфейсы с пользователем, чаще всего, только англоязычные. Исключение – случаи, когда для организаций в качестве «рабочих» утверждены несколько языков. Отметим следующее: качество текстов переведенных «вручную» интернет-страниц лучше, чем при использовании интернет-переводчиков [19]; последние вообще не могут обрабатывать тексты в составе графических объектов, размещенных на сайтах; внешний вид разноязычных страниц сайтов, а также состав размещенных на них графических объектов, иногда отличаются.

При входе на мультиязычные сайты международных ассоциаций чаще всего устанавливается английский язык интерфейса с пользователем. Однако возможно и автоматическое установление языка с помощью программных средств сайта путем определения местоположения пользователя (страны и даже региона) на основе его IP-адреса, государственного языка этой страны (если на сайте имеется СС с таким языком).

На СС сайтов практически всех ассоциаций есть средства прямого перехода на их страницы в «социальных сетях» – в том числе Facebook, Twitter, LinkedIn. На таких страницах нередко отображается и некоторая статистика, характеризующая их «популярность» в этих сетях.

Использование названий, эмблем, девизов ассоциаций

Англоязычные названия большинства международных ассоциаций имеют общепринятые англоязычные аббревиатуры, применяемые для следующих целей:

а) как «мнемонические» части имен интернет-адресов сайтов – это облегчает запоминание или «угадывание» таких адресов (характерно также использованием доменного имени «org»). Однако иногда соответствующие адреса уже «заняты» другими организациями;

б) как части адресов электронной почты ассоциаций;

в) на страницах сайтов самих ассоциаций;

г) на других зарубежных сайтах – в том числе для гиперссылок;

д) на русскоязычных сайтах, в том числе в сочетаниях с русскоязычными названиями этих ассоциаций и/или их русскоязычными аббревиатурами.

Для названий большинства международных ассоциаций существуют устоявшиеся русскоязычные переводы, которые используются в отечественных научных статьях и монографиях, обзорах, на российских интернет-сайтах и т.д. Однако в англоязычных частях статей должны быть вставлены «оригинальные» названия ассоциаций или англоязычные аббревиатуры.

Многие рассматриваемые ассоциации имеют специальные эмблемы, рассчитанные на ассоциативное восприятие направлений их деятельности пользователями. Эти эмблемы на сайтах отражаются рядом с названиями ассоциаций (иногда в сочетании со значком ®); широко применяются при проведении конференций и иных мероприятий; используются при публикациях различных материалов, издаваемых под эгидой ассоциаций. Можно считать, что эмблемы в сочетании с аббревиатурами названий являются частями своеобразных «брендов», играющих важную роль в поддержке места ассоциаций в научно-информационном пространстве.

На некоторой части сайтов ассоциаций отображаются также их «девизы» в виде текстов, рассчитанных на эмоциональное восприятие пользователями.

На рассматриваемых сайтах встречаются такие варианты представления состава стран-участников (членов): текстовые списки; контурные карты соответствующих континентов с выделением границ стран-участников; в виде совокупности государственных флагов стран-участников.

Функциональность сайтов и особенности их информационного наполнения

Для рассматриваемых сайтов типично использование в верхнем (или боковом) меню пунктов «About Us» (или «Who we are»), «What we do», «Join Us»; «Scientific Activities»; «Countries»; «Publications» и т.д.

Достаточно часто на сайтах ассоциаций встречается регистрация пользователей для их входа на сайт в целом или только в личный кабинет по логину и паролю. Это предоставляет пользователям расширенные (или дифференцированные) возможности доступа к информации, позволяет ее комментировать, оценивать, а при необходимости накапливать в личных кабинетах.

Практически все рассматриваемые сайты имеют ИПС по размещенной на них информации. Рассмотрим типичные варианты реализации ИПС. Первый (наиболее распространенный) – перенаправление введенных пользователями простых поисковых запросов (в виде слов или коротких фраз) к поисковым системам Интернета. При этом область поиска ограничивается содержимым соответствующего сайта. Отметим, что только на части сайтов прямо указывается, какая именно используется поисковая система Интернета; практически все ИПС ограничивают «область поиска» на сайте только тем языком, на котором был введен запрос (это важно для сайтов, содержащих информацию на разных языках); при начале ввода запросов в ИПС на рассматриваемых сайтах обычно не отображаются «выпадающие списки» с характерными формулировками запросов. Для срав-

нения – в электронных репозиториях (например, на www.elibrary.ru) возможности управления отбором материалов значительно шире [1], в том числе на основе анализа только названий ЕХИ, их ключевых слов и т.п..

Второй – «усиленный» по сравнению с первым вариантом. В ИПС используются формы для задания условий отбора; на основе введенной в них информации автоматически генерируются «расширенные запросы» к поисковым системам Интернета с ограничением области поиска информации сайтом (этот подход на рассматриваемых сайтах не встречается). Третий – полностью автономные ИПС сайтов. Они могут включать «формы» для подробного задания критериев отбора информации, выпадающие списки для полей этих форм и другие возможности. Однако это весьма трудоемко в реализации. Четвертый – комбинация вариантов: первого (для большинства видов запросов) и третьего – для решения некоторых частных задач.

Все ИПС рассмотренных сайтов выдают только совокупности отдельных ссылок, а возможности «агрегации и суммаризации» [20] информации не предусматриваются (последнее может быть весьма полезным при большом количестве ссылок).

На некоторых сайтах применяются и «рубрикаторы» информации – по темам, по странам, по направлениям деятельности и пр. При этом единицы хранения информации, отфильтрованные с помощью рубрикаторов, часто привязываются к датам размещения этих единиц на сайтах.

Специальные средства адаптации для лиц с ограниченными возможностями по зрению на сайтах рассматриваемых ассоциаций пока не предусматриваются. Видимо считается, что при необходимости может использоваться ручное масштабирование экрана, экранные лупы и т.п. Потенциально возможны такие решения по адаптации сайтов: 1) использование гиперссылок на «версии для слабовидящих» с измененными размерами шрифтов; 2) средства автоматического реферирования единиц хранения информации или их частей с представлением на дисплеях крупными шрифтами кратких рефератов и, возможно, частей изображений из реферируемых материалов [21]. (Конечно, для получения рефератов фрагментов текстов их можно копировать в соответствующие программные средства и просматривать в увеличенном масштабе; однако для слабовидящих пользователей это неудобно. В настоящее время программные средства автоматического реферирования имеются или разрабатываются для различных языков [22], включая персидский [23]); 3) гиперссылки на версии сайтов для лиц с нарушениями цветовосприятия (чаще всего это дальтонизм) с увеличенным цветовым контрастом; 4) комбинация пунктов 1 и 3: черно-белый вариант сайта с увеличенными шрифтами.

Адаптация цветовых решений и размеров шрифтов на интернет-страницах может также осуществляться автоматически на основе сведений в профилях пользователей при их входе на сайт или в рабочий кабинет по логину и паролю.

Для получения информации с сайтов возможные следующие варианты:

- по подписке (обычно, без возможностей фильтрации по «темам») – на указанные пользователями адреса электронной почты;
- с помощью средства RSS (Rich Site Summary, иконка). Однако это средство есть лишь на некоторых из рассматриваемых в статье сайтов. При этом пользователи могут самостоятельно «включать» ссылки на RSS сайтов в панели браузеров на своих компьютерах;
- потенциально возможно периодическое автоматическое дублирование информации с рассматриваемых в статье сайтов на российские интернет-ресурсы (при использовании соответствующих программных средств). Например, с зарубежных сайтов может браться информация о научных конференциях и размещаться на сайтах российских вузов;
- использование технологий «интеллектуальных агентов» [24] позволяет в автоматическом режиме осуществлять мониторинг появления на контролируемых сайтах информации нужной тематики, агрегировать такую информацию с различных сайтов, размещать результаты агрегации на российских сайтах.

Статистика использования сайтов и информация о гиперссылках

Такая статистика может использоваться для принятия решений по управлению сайтом, включая оптимизации распределения контента между отдельными страницами, количеством и номенклатурой внутренних гиперссылок на сайте [25–27]. Для получения статистики системные администраторы сайтов используют ряд средств/методов: анализ протоколов доступа к сайту, в том числе в отношении поисковых запросов пользователей и для иных целей; подключение внешних анализаторов вебометрических показателей [28–30] (однако это может влиять на уровень информационной безопасности сайтов – по крайней мере, косвенно); размещение на самих сайтах специальных счетчиков [31], средств анализа трафика [32, 33]; определение количеств скачиваний отдельных файлов пользователями – это сейчас рассматривается и как показатель «альтметрики» для отдельных единиц хранения информации [27]. При этом на СС сайтов ассоциаций, как правило, не отображается даже статистика по количеству входов пользователей на сайт.

У лиц, не являющихся системными администраторами сайтов, возможности получения (определения) статистики по сайтам также имеются, но они более ограниченные. Возможности бесплатных программных средств (в том числе по анализу гиперссылок) подробно описаны в [1, 2]. Рассмотрим подробнее некоторые направления анализа статистической (вебометрической) информации по сайтам.

1. Для совокупности пользователей поисковые запросы к ИПС конкретного сайта сейчас анализируются системными администраторами в основном в отношении частоты применения (всеми пользователями) поисковых слов или фраз, по количествам произведенных переходов пользователей внутри сайта с

использованием гиперссылок. При более глубоком анализе может исследоваться статистика по промежуткам времени между двумя последовательными запросами и/или другими действиями одного и того же пользователя. В частности, если при очередном запросе начальная «порция» ссылок, выданных ИПС, не удовлетворяет пользователя, то он, весьма вероятно, не будет проверять последующие «порции» [34]. Вместо этого он может сделать следующее: сразу же ввести новый поисковый запрос; перейти на «ручной» поиск нужной информации (например, по оглавлениям номеров изданий, размещенных в архивах сайта); покинуть сайт – обычно после нескольких неудачных попыток поиска [34].

2. Количество входов посетителей на сайт, в том числе уникальных посетителей. На ряде интернет-сайтов (не рассматриваемых в данной статье) отображается следующее: количество входов посетителей за определенный период (сутки, неделя, месяц, год); редко – текущее количество посетителей (например, на www.elibrary.ru). Для лиц, не являющихся системными администраторами сайтов, нет возможности получать информацию о количестве посетителей за заданный период времени; о распределение входов посетителей по часам суток; о распределении посетителей по тем странам, из которых производились входы на сайты.

3. Продолжительность нахождения посетителей на сайтах с учетом особенностей такого учета, описанных в [9].

4. Распределение входящих на сайты пользователей по типам используемых ими устройств, размерам и разрешению их экранов, применяемым операционным системам, используемым браузерам.

5. Для администраторов сайтов рассматриваемых международных ассоциаций важна статистика по номенклатуре стран, пользователи из которых входят на сайты. Это можно сделать на основе анализа IP-адресов пользователей, если они не используют «анонимайзеры».

6. Также для администраторов сайтов имеет значение статистика хакерских атак на сайты, в том числе попыток использования «инъекций вредоносного кода» для вывода сайтов из строя; попыток получения доступа к персональным данным пользователей, зарегистрированных на сайте и/или к материалам, закрытым для общего доступа. В тоже время DDOS атаки на рассматриваемые сайты не характерны.

7. Статистика «подписок» на получение информации с сайтов: по электронной почте, с использованием RSS и т.д.

8. Сбор информации о действиях любых пользователей возможен на основе cookies Информация о том, что они применяются, на многих рассматриваемых сайтах есть, хотя иногда она «спрятана» и открывается только при выборе соответствующих пунктов меню. Иногда пользователь должен в явной форме «согласиться» с использованием cookies.

При входе пользователей на сайты (или в «личные кабинеты») под индивидуальными логинами и паролями возможен сбор и анализ «персональной» статистики, в том числе о выполняемых действиях, о времени пребывания на сайте и пр. Однако, на СС рассматриваемых сайтов в явной форме декларации о сборе и направлениях использовании такой статистики, присутствуют не всегда.

Подходы к оценкам авторитетности и результативности использования сайтов

Для сайтов научного характера (в том числе сайтов рассматриваемых ассоциаций) ранжирование/ рейтингование [29] на международном уровне не характерно (в отличие, например, от сайтов вузов). Ранжирование сайтов ассоциаций может, в частности, осуществляться на основе количества входов пользователей, суммарного времени, проведенного ими на сайтах и т.п.. Возможно и экспертное оценивание отдельных сайтов внутри их групп определенной направленности (тематики).

Использование средств альтметрики по размещенным единицам хранения информации уже применяется в некоторых репозиториях научной информации [8,10] в виде учета «лайков» на ЕХИ, их бальных оценок пользователями, включения пользователями ЕХИ в «личные подборки» и т.п. Однако для сайтов рассматриваемых ассоциаций такие решения пока не характерны.

Некоторые вопросы обеспечения информационной безопасности сайтов

На информационную безопасность сайтов влияет ряд факторов [35, 36]. При создании аккаунтов новыми пользователями осуществляется контроль минимально необходимого (и/или желательного) количества символов в пароле. После предусмотренного на сайте периода непрерывного использования пароля от пользователя иногда принудительно требуется его смена.

Все системы регистрации новых пользователей (или смены ими логинов) проверяют, чтобы вновь вводимые логины не совпадали с уже «задействованными» (при совпадении выдаются соответствующие сообщения). Потенциально это позволяет злоумышленникам выявлять уже используемые чужие логины путем их подбора, в том числе, когда некоторые части логинов они знают.

Если злоумышленникам известны логины пользователей (они, обычно, отображаются на мониторах при их наборе), то при наличии информации о количестве символов пароля, о некоторых его фрагментах (и/или принципах конструирования паролей отдельными пользователями) потенциально возможен и «подбор» чужих паролей. «Вручную» это может делаться, например, многократным повторением попыток ввода паролей.

Обычно при регистрации пользователей на сайтах требуется указание не только логина и пароля, но и номеров личных сотовых телефонов и/или адресов электронной почты, в том числе для передачи пользователям по этим каналам информации, позволяющей завершить процедуру регистрации. Часто требуется также указание фамилии и имени (однако достоверность этой информация фактически никак не контролируется), а иногда и места работы. Причины, по которым пользователей может не устраивать пре-

доставление (передача) на сайты персональной информации: ограничения, установленные работодателями, уставами некоторых ведомств; опасения по поводу возможных утечек персональных данных, в том числе при взломах сайтов хакерами [35], в результате действий спецслужб и прочее.

Определенную угрозу для информационной безопасности пользователей могут представлять и средства «восстановления паролей» на сайтах. Это связано с тем, что при использовании запросов на восстановление пароля на многих сайтах требуется ввод только адреса электронной почты для отправки ссылки на страницу со средством активации нового пароля. E-mail адреса пользователей широко распространяются в Интернете (например, в составе сведений об авторах научных статей в электронных репозиториях). Поэтому злоумышленники, зная чужие адреса электронной почты, иногда могут «запрашивать от чужого имени» восстановление (изменение) паролей. Если злоумышленникам неизвестна комбинация «логин и пароль» для входа в электронную почту пользователя, то подтвердить изменение пароля для входа на сайт вместо самих пользователей они не смогут. Однако пользователи получат запросы на подтверждение «изменений паролей», вынуждены будут разбираться с ними; иногда по невнимательности или из-за спешки могут «механически» нажать на гиперссылку для подтверждения активации нового пароля. В последнем случае если присланная гиперссылка для активации не сопровождается указанием нового «логина и пароля», то пользователь может вообще утратить возможность работы со своим аккаунтом.

Типичные решения по дополнительной защите пользователей: отправка на личные сотовые телефоны пользователей (указанные ими в своих профилях) SMS-сообщений с кодами подтверждений на изменения логинов и паролей» (однако, «международный роуминг» SMS-сообщений обычно является платной услугой в отличие от отправок запросов на активацию по электронной почте через Интернет); ввод пользователями для подтверждения их аутентичности ответов на «секретные вопросы», также включенные в профили пользователей.

На рассматриваемых в настоящей статье интернет-ресурсах обычно не указываются и программные средства, примененные для создания сайтов, работы с их базами данных, а также специальные средства поддержки интерфейса с пользователями. Такая информация может быть важной и полезной при применении технологий «интеллектуальных агентов» [24]. Однако сведения об используемых программных средствах могут снижать степени устойчивости сайтов к взломам при хакерских атаках. Связано это с тем, что с течением времени в инструментальных средствах разработки, которые были использованы при создании сайтов, позже могут выявляться различные «уязвимости». Информация о них сразу же попадает в Интернет, быстро распространяется, оперативно используется злоумышленниками. При этом обновления, «закрывающие» выявленные «уязвимости» (для операционных систем, для программных средств разработки/поддержки сайтов, для антивирусных средств), а также рекомендации о необходимых «защитных действиях» системных администраторов сайтов часто появляются с запаздыванием по сравнению с информацией об «уязвимостях» [35]. На российских интернет-ресурсах часто указываются программные средства, обеспечивающие их антивирусную и/или антиспамовую защиту. Однако на рассматриваемых в статье зарубежных сайтах такой информации как правило нет.

На СС лишь некоторых рассматриваемых сайтов есть сведения по соблюдению информационной безопасности персональных данных пользователей; выдаются предупреждения об ограничениях (соблюдении условий) при использовании этих данных; делаются запросы «подтверждений» на право работы с такой информацией, а также на право использования *cookies* для различных целей.

Сведения о внешних организациях, выполнивших «дизайн» сайта, а также осуществляющих его программно-техническое сопровождение, на рассматриваемых сайтах приводятся, но не во всех случаях (это важно и с позиций информационной безопасности [35]).

Как правило, на сайтах нет и сведений о местах размещения серверов, наличии у сайтов «зеркал» и т.п. Однако известно, что у некоторых международных ассоциаций есть «зеркала» на сайтах российский академических организаций, вузов.

Сведения об авторских правах («копирайт») почти всегда указывается на СС сайтов – обычно в сочетании со значком © (копирайт) и фразой «All rights reserved».

Сведения о количестве страниц сайтов и некоторых других характеристиках могут определяться в процессе их «мелкозернистого исследования» [37], однако на самих сайтах такие характеристики не указываются.

На сайтах ассоциаций научных организаций можно выделить следующие виды и источники размещенной информации:

1) общие сведения об ассоциации-владельце сайта, целях ее деятельности, юридическом статусе, административной подчиненности и т.п.;

2) данные о структуре (подразделениях) этой ассоциации, руководителях подразделений. Совокупность названий подразделений иногда может быть весьма информативной для оценки профиля деятельности ассоциации;

3) контактные данные ассоциации и, иногда, ее отдельных подразделений;

4) сведения о редакционно-издательской деятельности ассоциации, включая номенклатуру научных изданий, об условия публикаций в них и составах редакционных коллегий;

5) собственно научная информация в виде архивов научных статей, кратких сообщений, тезисов, рефератов; материалов прошедших и анонсов предстоящих мероприятий (включая конференции, семинары), иных научных материалов;

6) сведения о наградах (дипломах, медалях и пр.) присуждаемых/присужденных данной ассоциацией физическим лицам и ассоциациям/организациям, а также о наградах, полученных самой ассоциацией;

7) информация познавательного характера (в том числе научно-популярная), предназначенная для ши-

рокого круга посетителей; популяризации деятельности ассоциации и входящих в нее организаций [38];

8) сведения о возможностях получения грантов, стажировок, в том числе предоставляемых самой ассоциацией, организациями – членами ассоциации, иными организациями.

9) информация о направлениях развития внешних связей ассоциации с другими ассоциациями/организациями, научными группами, отдельными исследователями;

10) сведения о членстве данной ассоциации в других ассоциациях, об условиях индивидуального и коллективного членства в ассоциации, о фактическом членстве отдельных исследователей (иногда и о результатах их научной деятельности);

11) хроника «внутренней жизни» ассоциации: информация о достижениях штатных сотрудников самой ассоциации, а также входящих в нее организаций; анонсы предстоящих мероприятий; сведения о культурных мероприятиях;

12) информация о контактах ассоциации с другими ассоциациями, иными внешними организациями;

13) сведения о вакансиях в ассоциациях (а иногда и в организациях-участниках), о требованиях к кандидатам.

Источники информации для рассматриваемых в статье сайтов разделим на внутренние и внешние. Внутренние: научные материалы (включая статьи), предоставляемые руководством ассоциаций, отдельными сотрудниками, структурными подразделениями ассоциаций, организациями-участниками; иные материалы, «создаваемые» внутри ассоциации. Внешние источники: статьи и иные материалы внешних авторов, публикуемые в изданиях ассоциации и размещаемые на ее сайте; информация, представляемая «смежными» ассоциациями/организациями; сведения, поступающие от информационных агентств, от агрегаторов НТИ; тематически ориентированная информация с ведущих интернет-порталов. На рассматриваемых сайтах размещается в основном информация, созданная внутри ассоциаций или предоставленная организациями-участниками. Поэтому следует ожидать, что количество входящих ссылок на сайты в большинстве случаев будет больше, чем исходящих.

Обычно информация (прежде всего новостная), ранее размещенная на ведущих страницах сайтов, по мере ее устаревания заменяется на более актуальную, а «прежняя» – перемещается в архивы сайтов. При этом для архивируемой информации типична структуризация по датам первоначального размещения или номерам изданий, а не по тематике (единицы хранения информации обычно не снабжаются собственными тематическими дескрипторами). Поэтому внутренние ИПС сайтов обычно проводят поиск только по «встречаемости» заданных слов в ЕХИ.

Сайты некоторых ведущих международных ассоциаций глобального и континентального характера, не специализированных по тематике деятельности

Приведем данные сначала по глобальным, потом по континентальным ассоциациям; рассмотрим особенности их сайтов, важные для доступа потребителей к научной информации.

1. UNESCO (ЮНЕСКО) – специализированное учреждение Организации Объединённых Наций по вопросам образования, науки и культуры. Его девиз на СС сайта: «Нести мир в сознание мужчин и женщин». Возможно переключение между «рабочими» языками: английским (https://en.unesco.org/); французским (https://ru.unesco.org/), испанским (https://es.unesco.org/), русским (https://ru.unesco.org/), арабским (https://ar.unesco.org/), китайским (https://zh.unesco.org/). Однако у части страниц сайта есть только англоязычные версии (или англоязычные и франкоязычные). ИПС сайта использует Google Custom Search. На стартовых страницах размещено достаточно много новостной информации, также на них (и иных страницах) есть кнопки перехода на сайты различных социальных сетей.

На вкладке «Ресурсы» в пункте «Публикации» имеются подпункты: «Книжный магазин он-лайн» (для приобретения изданий ЮНЕСКО); «Базы данных» (почти по 150 тыс. документов ЮНЕСКО); «Библиотека» (для обеспечения пользователей «ссылками и информационными услугами»); «Архивы» (доступ к документам, публикациям, мультимедийным и электронным источникам информации). Пункт «Статистика» на вкладке «Ресурсы» (только англо- и франкоязычные версии) обеспечивает возможность доступа к тематически структурированным данным примерно по 200 странам.

В нижней части СС есть пункты: «Ограничение ответственности» (заявление об отсутствии гарантий того, что информация, документы и материалы на сайте являются полными и безошибочными); «Политика конфиденциальности» (заявление о том, что информация, полученная о пользователе при посещении им сайта, будет использоваться исключительно для служебного анализа его посещаемости), «Условия использования», «Политика ЮНЕСКО в области представления доступа к информации»

2. International Council for Science (ICSU). Членами этой ассоциации являются 122 национальные научные организации, 24 научные ассоциации, 30 научных союзов. На СС (https://www.icsu.org/) размещено следующее: ИПС для запросов, оперативная (новостная) информация, тематические рубрикаторы для доступа к определенным видам информации, вкладка «публикации» (только документы самого ICSU), на 05.01.2018 было размещено объявление о «слиянии» ICSU и International Social Science Council (ISSC).

3. У ISSC на 05.01.2018 г. был и свой сайт (http://www.worldsocialscience.org/). На его СС есть гиперссылки для перехода: на YouTube (с автоматическим открытием русскоязычного интерфейса), на страницы в социальных сетях, на SoundCloud – «Search for artists, bands, tracks, podcasts» (иконка ), на сервис Flicr – для работы с фото (иконка).

4. World Federation of Engineering Organizations (WFEO) – сайт (http://www.wfeo.org/). На его СС есть: ИПС, гиперссылка на LoginPage, новостная информация. Во вкладке Events верхней ленты-меню в выпадающем списке имеется пункт Call for Papers (архивная информация о мероприятиях структурирована по годам и тематикам мероприятий); возмож-

ность перехода на Flickr, на сайты социальных сетей; в нижней части – гиперссылка на страницу оформления подписки на получение по электронной почте «newsletters and flash infos», доступа к архивным flash infos. Во вкладке Knowledge верхнего меню на СС отметим пункты списка «Publications and E Newsletters», «IDEAS Engineering Education Journal», «Archives and Library».

5. International Association of Science Parks and Areas of Innovation (IASP). В верхней части СС сайта (www.iasp.ws) размещены: поле ИПС, гиперссылки на страницы социальных сетей и Login-страницу. В верхней ленте-меню во вкладке Our Industry есть пункт Knowledge Room, предназначенный для приобретения посетителями сайта «платных публикаций». В центральной части СС расположена громоздкая панель-меню со сменяющимися фото и соответствующими им кнопками доступа к функциональным разделам. В нижней части СС – горизонтальная лента организаций-участников (включая российские), эмблемы которых выполняют функции гиперссылок. Эта лента «прокручивается» автоматически, но есть и ручное управление с помощью навигационных стрелок. Также внизу страницы можно увидеть предупреждение об использовании cookies для целей улучшения работы сайта.

6. European Federation of National Engineering Associations (FEANI) – ориентирована, в основном, на поддержку инженерного образования в Европе. На стартовой странице сайта (https://www.feani.org) присутствует гиперссылка на Login-страницу; в верхней ленте-меню –вкладки European Engineering Education Database, Links, Publications (только «Annual Reports»); в центральной части – располагается основное меню с крупными картинками и кнопками; в нижней части – есть лента с эмблемами организаций-участников.

7. EuroScience (Европейская организация продвижения науки и технологий) – некоммерческая организация исследователей. Имеет девиз «Your voice on Research in Europe», является основателем EuroScience Open Forum (ESOF), издает сетевой журнал EuroScientist. На СС ее сайта (http://www.euroscience.org/) в верхней части есть: гиперссылка на страницу Log in, кнопка RSS, средство подписки для получения по электронной почте «Euroscience newsletter», гиперссылки на страницы социальных сетей (последние три позиции дублируются и внизу СС). В верхней ленте-меню отметим пункты «Join», «My Profile». В центральной части СС размещены новостная информация, запросы, пресс-релизы, твиты. Возможности для пользователей: различные виды членства в организации, управление «персональным профилем», возможность присоединения к «местным секциям и рабочим группам» по направлениям деятельности.

Сводки вебометрических показателей для сайтов рассматриваемых ассоциаций этой группы представлены в табл. 1 и 2.

Отметим: 1) детальный анализ структуры исходящих ссылок на сайте UNESCO (пункты 1a…1e, колонки Ω и Ψ) показал, что для всех языков значительно преобладают ссылки на страницы UNESCO на других языках. Таким образом, количество входящих и исходящих ссылок на сайты этой «группы в целом» в несколько раз меньше, чем «внутригрупповых»; 2) для сайтов ICSU и IASP имеется весьма большое количество входящих и внутренних ссылок при достаточно скромном количестве исходящих; 3) для всех сайтов со 2-го по 7-й, как и следовало ожидать, количество исходящих ссылок многократно меньше чем входящих; 4) для этой же группы сайтов отметим: весьма скромное количество (за месяц) «уникальных посетителей» и просмотров (примерно четыре просмотра на каждого «уникального посетителя»); 5) наибольшие значения AVD имеют два сайта: IASP и FEANI, а наименьшее значение этого показателя – сайт EuroScience.

Таблица 1

Вебометрические показатели для сайтов «универсальных» ассоциаций (первая часть)*

№	Название ресурса	AT, сек.	Count, URLs			Size, Mb	Scholar Google (SG)
			Text/html	image	application		
1	UNESCO						
	на английском	0,44	4841	1550	704	818,19	60
	на французском	0,49	2196	950	50	128,46	15
	на испанском	0,63	10854	917	182	825,85	2
	на русском	0,48	6158	1599	107	283,74	14
	на арабском	0,55	2897	885	50	203,55	9
	на китайском	0,5	3541	1010	83	176,14	5
2	ICSU	1,81	569	31	256	290,26	6
3	ISSC	0,3	1312	472	259	356,82	28
4	WFEO	0,52	849	691	249	618,3	28
5	IASP	2,49	30390	1269	220	4539,75	0
6	FEANI	1,06	165	109	29	61,34	0
7	EuroScience	0,68	577	0	1	19,31	1

**Примечания*: 1) показатели «Text/html» на сайтах UNESCO для разных языков существенно отличаются, причем наибольшее значение соответствует испанскому языку, а не английскому; 2) сайт IASP имеет набольшую длительность открытия СС; максимальные показатели для «Text/html» и объема сайта. В тоже время по показателю «application» лидирует англоязычная версия сайта UNESCO; 3) «видимость» ресурсом Scholar Google материалов на всех этих сайтах очень низкая.

Таблица 2

Вебометрические показатели для сайтов «универсальных» ассоциаций (вторая часть)

	Название ресурса	Абсолютные количества ссылок				КУнПос. за месяц	КПросм.	AVD, чч:мм:сс
		Входящие	Вн	Исх				
				Ω	Ψ			
1	UNESCO							
	на английском	7342173	283049	10464	131842	216110	823803	00:02:21
	на французском	1115362	31170	7928	53844	36804	140296	00:01:45
	на испанском	379416	391863	30658	277961	102656	391323	00:01:39
	на русском	361125	128383	15913	139983	5996	22860	00:02:33
	на арабском	10468	39011	7646	60519	9335	35586	00:02:33
	на китайском	6320	86249	13837	99740	1908	7274	00:02:33
2	ICSU	2188820	39704	1154	4180	9120	36480	00:02:44
3	ISSC	321895	22342	930	7085	6892	27570	00:01:42
4	WFEO	355893	36460	1459	6602	10121	40470	00:02:39
5	IASP	4138675	4403137	523	90752	6067	24270	00:03:15
6	FEANI	269509	5120	313	2201	4254	17010	00:03:17
7	EuroScience	303864	31669	597	5787	6173	24690	00:01:12

Международные ассоциации, специализированные по тематическим направлениям деятельности (глобальный уровень)

Отметим, что сайты таких ассоциаций рассматриваются лишь выборочно.

1. International Mathematical Union (IMU) – всемирная некоммерческая и неправительственная организация, созданная для сотрудничества учёных всех стран в области математики. Члены IMU – национальные математические организации из 83 стран. Подробный анализ сайта этой организации, проведенный в 2009 г., есть в [16]. На СС сайта (http://www.mathunion.org/) в правом верхнем углу имеется несколько гиперссылок: на Login-страницу, на ICMI (International Commission on Mathematical Instructions – деятельность комиссии относится к сфере математического образования), CDC (открывается страница, посвященная «поддержке математики в развивающихся странах»), CEIC (Committee on Electronic Information and Communication), CWM (по гиперссылке открывается страница по теме «Женщины в математике»), ICHM (открывается страница, посвященная материалам по «истории математики»). При переходе по указанным гиперссылкам открываются верхние ленты-меню этих страниц, причем каждая из вкладок на них имеет собственные выпадающие списки с пунктами и подпунктами.

Доступ к ключевым страницам сайта (например, к World Digital Mathematics Library – WDML) дублируется с разных вкладок.

На странице ICMI отметим в частности вкладки Publications, Digital Library.

Со страницы CEIC есть доступ к: а) CEIC News – в том числе к пункту Blog, включая подпункт Blog on Mathematical Journals (2011–2012); б) Library – доступ к материалам в WDML (на 02.01.2018 г. – 9036 книги для приобретения в электронной форме); в) Publications; г) External Resources – с пунктами: Ulf Rehmann's list of Retrodigitized Mathematics Journals and Monographs, mini-DML – Cellule MathDoc's search engine over 15 digital math repositories, EuDML – European Digital Mathematics Library, а также на два «национальных» ресурса Math-Net.Ru (Всероссийский математический портал) и Numdam (the French digital mathematics library).

Верхняя лента-меню СС сайта содержит ИПС, вкладки About Us, Historia Matematica, Grants & Sponsorships, Prizes, Reports (относящиеся к деятельности IMU) и другие.

В нижней части стартовой страницы сайта есть средство организации подписки на Newsletter (с рассылкой по электронной почте). Также указывается, что «server host by WIAS» (Weierstrass Institute for Applied Analysis and Stochastics, Berlin).

2. Society for Applied and Industrial Mathematics – SIAM (http://www.siam.org/). На широкоэкранных мониторах вся СС умещается целиком, т.е. вертикального скроллирования не требуется. В правом верхнем углу СС есть кнопки перехода на страницы в социальных сетях, перехода на Google+, SiteMap, кнопки создания и управления аккаунтом пользователя, средство поиска по сайту (с фильтром по типам объектов поиска). Основное меню на СС организовано в виде левой боковой ленты. Оно включает, в частности, такие пункты: Books, Conferences, Digital Library (SIAM Journals on-line), Journals (гиперссылки на многочисленные журналы этого общества), Reports, SIAM News. Внизу СС имеются пункты: Statement on Inclusiveness (в отношении политики сайта), Privacy policy, «Social Media Policy», а также информация «Website developed by Zero Defect Design LLC».

3. International Council for Industrial and Applied Mathematics – ICIAM (http://www.iciam.org/). Его СС не требует скроллирования; немногочисленные пункты верхней ленты-меню (включая News, Events, Prizes) дублируются внизу СС с расшифровками подпунктов. Имеется ИПС по сайту, две гипер-

ссылки на страницу «LogIn» с аналогичной функциональностью (вверху СС и внизу справа). Внизу СС можно найти пункты: Privacy, Terms of Service, Sitemap, Glossary – с подробной коллекцией гиперссылок на сайты математических обществ разных стран, на отчеты, на иные объекты (структурирование всех этих объектов производится по начальным буквам алфавита в их названиях).

4. International Society for Business and Industrial Statistics (ISBIS). На СС его сайта (http://www.isbis-isi.org/), имеющей небольшую ширину, основное меню – это вертикальная лента слева. Его наиболее полезные пункты: Conferences and Workshops; Publications (включает архивы номеров научных журналов), Links (ссылки на сайты ассоциаций/обществ по тому же или смежным направлениям деятельности), Blogs, Members Only (обеспечивает возможности входа на части сайта «закрытые от общего доступа» путем ввода логинов-паролей для трех различных категорий пользователей). Информация по журналам дублируется также в правой стороне центральной части сайта.

5. The Bernoulli Society for Mathematical Statistics and Probability (Bernoulli Society) насчитывает более 1000 представителей почти из 70 стран. Ее аскетически оформленная СС сайта (http://www.bernoulli-society.org/) имеет меню в виде вертикальной ленты слева, включающей пункты: Publications (в основном по журналам, которые общество издает или является одним из их спонсоров), Meetings, Prizes, Links (в том числе две гиперссылки на архивы публикаций – по статистике и по теории вероятностей).

6. Association for Computing Machinery (ACM) использует девиз «Advancing Computing as a Science & Profession». Считается, что эта организация со штабквартирой в Нью-Йорке объединяет около 83 тыс. специалистов отрасли. Она имеет около 500 вузовских отделений; занимается как прикладными исследованиями, так и некоторыми теоретическими. В верхней части СС сайта (http://www.acm.org/) имеется три горизонтальные меню-ленты. В верхней ленте отметим: Digital Library (в том числе по публикациям в журналах ACM), TechNews, Learning Center. В средней ленте: Join; MyACM (гибкое средство индивидуальной настройки для работы с сайтом), Search (это ИПС). В нижней ленте: Publications, Special Interest Groups, Conferences, Chapters (фактически – это средство информационного обеспечения деятельности Special Interest Groups), Public Policy. В российских вузах из числа мероприятий, проводимых ACM, наиболее известен командный студенческий чемпионат мира по программированию (https://icpc.baylor.edu/welcome.icpc). В нижней части СС: дублируются пункты «нижней ленты» (см. выше); есть кнопки перехода на сайты социальных сетей; кнопка «+» – дополнительные сервисы; в самом низу – пункты Privacy Policy, Social Media Policy, Accessibility.

7. International Union of Geological Sciencies (IUGS) – неправительственная организация со 121 «национальными членами» (http://www.iugs.org/). Ее девиз – Earth Science for the Global Community. На СС меню имеет вид вертикальной ленты слева. Оно содержит пункты: Documents, Publications, E-Bulletins, EC Members Only, Links (многочисленные гиперссылки на сайты научных комиссий, программ, рабочих групп, аффилированных организаций, международных союзов; на научные мероприятия, разделенные по тематике). В правой части СС есть немногочисленные кнопки для перехода на страницы социальных сетей, в том числе на твиттер. Отметим группу гиперссылок в нижней части СС на иноязычные «листки» (leaflets)– включая африкаанс, суахили, фарси, урду и некоторые другие.

8. International Union of Geodesy and Geophisics (IUGG) – неправительственная организация. На ее сайте (http://www.iugg.org/) СС имеет малую ширину, содержит две ленты меню. Через аббревиатуры на верхней ленте осуществляются переходы на сайты входящих в союз ассоциаций. Из пунктов левой боковой ленты для темы настоящей статьи наиболее важны: Associations (расшифровки для аббревиатур верхней ленты и гиперссылки на сайты соответствующих организаций), Scientific Meetings, Research Programs, Grant Programs; Publications (публикации IUGG), Related Organizations – подробный список «тематически смежных» организаций с гиперссылками на их сайты. Внизу СС помимо декларации «копирайта» указан разработчик веб-страниц (This webpage was designed by Ametec & Blackcrystal). Есть переключатель на французский язык (в правом нижнем углу СС), но он малозаметен.

9. The International Cartographic Association (ICA) включает ряд «тематических комиссий». На СС её сайта (http://icaci.org/) в верхней части есть кнопки перехода на сайты социальных сетей; кнопки RSS, Flickr, Picasa – в виде  (архив альбомов). В основном меню на СС (верхняя лента) в отношении научной информации наиболее полезны пункты: Publications, ICC conferences, Research Agenda, The Association (через подпункт «Membership and sister societies» открывается список «родственных» организаций с их названиями, эмблемами, гиперссылками на официальные сайты). Новостные сообщения сгруппированы по ежемесячным выпускам «eCARTO News», а внутри них – по темам. Для просмотра сообщений необходим вертикальный скроллинг на много экранов. Есть возможность оформления подписки на Newsletter по электронной почте. Вход пользователей по логину и паролю на этом сайте организован через пункт в самой нижней части СС.

10. International Union of Pure and Applied Physics (IUPAP) – его членами являются 59 стран. На СС сайта (http://iupap.org/) есть ИПС. Основное меню оформлено как двухстрочная лента с рядом пунктов, включая Conferences, Working Groups (тематически специализированные комиссии/группы, в т.ч. Women in Physics). Имеется также Publications – кратко аннотированная коллекция гиперссылок на номера изданий союза. В нижней ленте-меню СС имеются пункты, связанные с информационной безопасностью (Privacy and Cookies; Terms and Conditions) и особыми условиями использования сайта (Disclaimer) .

11. International Union of Pure and Applied Chemistry (IUPAC) – имеет более 50 стран участниц. В правом верхнем углу СС сайта (https://iupac.org/) есть

пункты Join, Login, ссылка на ИПС сайта. В меню «верхняя лента» (оно сохраняется на экране при скроллинге) пунктов немного, включая Events, Projects, News. В средней по высоте части СС можно найти кнопки перехода на типы страниц: по определенной тематике, со списками «смежных» организаций, по «индивидуальным» членами. В нижней части СС дана ссылка на организацию, разработавшую и сопровождающую сайт (Web Design & SEO by TheeDesign Studio).

12. International Astronomic Union (IAU) – его членами являются более 70 государств и примерно 11 тыс. физических лиц. Ширина СС сайта (https://www.iau.org/) невелика; меню на СС организовано в виде двух горизонтальных лент. В верхней ленте отметим: MemberDirectory (в т.ч. подробная половозрастная характеристика индивидуальных участников); login. В нижней ленте (главное меню) важны пункты: Science (в выпадающем списке есть подпункты Scientific Meetings, Scientific Collaboration Programs, Grants and Prizes), Publications (включая подпункты Information Bulletins, IAU Related Publications, E newsletters), IAU for the Public. Иерархия ссылок через пункты главного меню насчитывает от 3 до 5 уровней. На СС сайта имеются средства «Like» и «Share» (для работы в Facebook), вход по логину и паролю для принятых в IAU членов, средство оформления подписки на информационные рассылки по электронной почте, два средства поиска – по новостным сообщениям (слева) и в целом по сайту (справа). В самой нижней части СС отметим два «не типичных» пункта: Credit (аннотированный перечень лиц, отвечающих за отдельные направления работы по сайту), Technology (характеристика информационных технологий, применяемых для разработки и ведения сайта).

13. International Union on History and Philosophy of Science and Technology (IUHPST) – всемирная неправительственная организация. На СС его сайта (http://iuhps.net/) есть ИПС; меню организовано в виде левой боковой ленты с несколькими пунктами (включая Projects). Специально отметим графическое представление на СС совокупности связей IUHPST со «смежными» организациями и членами, включая ICSU (см. выше) и International Social Sciences Council (ISSC). Однако к объектам этой инфографики «не привязаны» гиперссылки на сайты соответствующих ассоциаций (организаций).

Сводки вебометрических показателей для сайтов, рассматриваемых в статье научных ассоциаций этой группы представлены в табл. 3 и 4.

Отметим некоторые особенности для этой группы сайтов: длительное «среднее время открытия» СС для сайтов IUPAC и ICA; значительно бо̀льший объем сайта IAU по сравнению с другими сайтами (однако для SIAM объем сайта, возможно, еще больше); для сайта IUPAC – большое количество объектов «Text/html» при сравнительно скромном размере сайта; во многих случаях Scholar Google практически не видит документов на сайтах.

По количеству входящих ссылок значительно преобладает сайт ACM, их также много на сайтах IAU и IUGS. Внутренних ссылок особенно много на сайтах IUPAC и IAU (для сайта SIAM их количество оценить не удалось). По количеству уникальных посетителей и просмотров значительно лидирует сайт ACM. Показатель AVD самый большой у Bernoulli Society – при небольшом количестве посетителей.

Таблица 3

Вебометрические показатели для сайтов тематически специализированных ассоциаций, отнесенных к «глобальному» уровню (первая часть)

№	Название ресурса	AT, сек.	Count, URLs			Size, Mb	Scholar Google (SG)
			Text/html	image	application		
1	IMU	0,82	8167	3966	1294	8162,23	83
2	SIAM	0,81	>1000000 { 22500}	-	-	-	229
3	ICIAM	1,8	1231	538	156	305,55	0
4	ISBIS	0,22	41	38	27	70,23	0
5	Bernoulli Society	0,66	302	30	31	65,33	1
6	ACM	0,52	2409	600	1720	782,17	20
7	IUGS	0,71	940	793	425	1056,57	3
8	IUGG	0,4	307	233	536	582,72	21
9	ICA	8,36	3808	1616	3583	4524,66	1390
10	IUPAP	3,16	935	364	438	207,74	9
11	IUPAC	14,46	125227	15051	11399	8968,98	716
12	IAU	0,92	9898	2550	372	109831,62	12
13	IUHPST	0,45	12	5	0	0,42	0

Таблица 4

Вебометрические показатели для сайтов тематически специализированных ассоциаций, отнесенных к «глобальному» уровню (вторая часть)

	Название ресурса	Абсолютные количества ссылок				КУнПос. за месяц	КПросм.	AVD, чч:мм:сс
		Входящие	Вн	Исх				
				Ω	Ψ			
1	IMU	868706	467661	4458	33648	18119	72480	00:02:34
2	SIAM	1076380	-	-	-	103181	412710	00:01:49
3	ICIAM	96065	40983	300	6673	2729	10920	00:03:13
4	ISBIS	3248	573	86	181	0-100	0-100	00:00:02
5	Bernoulli Society	1929814	4977	667	2940	2206	8820	00:04:06
6	ACM	5921050	301166	7857	108923	476956	1907820	00:02:24
7	IUGS	250774	20641	656	5073	3925	15690	00:01:36
8	IUGG	210423	6911	359	2371	7746	30990	00:01:19
9	ICA	179095	148954	5584	70487	11810	47250	00:01:32
10	IUPAP	47854	12881	737	3206	8279	33120	00:01:30
11	IUPAC	1242633	2653636	161738	439409	48147	192600	00:01:30
12	IAU	2875530	2642484	4443	244546	25408	101640	00:00:50
13	IUHPST	9	88	29	34	0-100	0-100	00:00:02

Международные ассоциации по тематическим направлениям деятельности (континентальный уровень)

В этом разделе ассоциации также рассматриваются выборочно. Отнесение их к континентальному уровню носит формальный характер (по названию).

1. American Mathematical Society (AMS) – основана в 1888 г. Насчитывает около 28 тыс. «индивидуальных» членов и примерно 560 организаций. Издает 10 исследовательских журналов и 4 переводных, а также книги. Подробная характеристика сайта AMS в версии 2009 г. дана в [16].

Девиз в верхней части СС сайта (www.ams.org) «Advancing research. Creating connections» подчеркивает важность коммуникативной функции AMS и ее сайта. В правом верхнем углу СС отметим такие иконки: My Account; My Shopping Cart (для приобретения изданий AMS). Представленная на СС сайта ИПС использует Google search; есть выпадающий список типовых запросов, соответствующих началу ввода фраз. В верхней ленте-меню (оно видно при переходах на все страницы сайта) отметим вкладки: а) Bookstore – реклама изданий, возможности бесплатного доступа к ним для членов AMS, включение в список рассылки и т.п.; б) MathScinet® – открывается окно запроса на русском языке к многофункциональной поисковой системе по публикациям, авторам, журналам, цитированиям; одно из полей отбора соответствует унифицированной (с ZbMath) схеме Mathematics Subject Classifications; в базе данных этой ИПС находятся около 3.5 млн. ЕХИ, включая HTML-документы; в) Journals – сгруппирована информация по журналам, есть дополнительное меню слева; г) Member Directory – открывается форма отбора ИПС для поиска в Combined Membership List по членам не только AMS, но и ряда «смежных» с ней организаций; д) Giving to AMS (форма по сбору спонсорской помощи). В центральной части СС расположены крупные «картинки с надписями», обеспечивающие переходы на соответствующие тематические разделы сайта. Под ними находятся восемь пунктов меню с подпунктами (они частично дублируют «картинки с надписями» и вкладки в верхней ленте-меню). В описываемой версии сайта (на 01.02.2018 г.) «не просматривается» подпункт «In Memory of» (с краткими некрологами о ведущих математиках ушедших из жизни в прошлом году), который был в конце 2017 г.. В самой нижней части СС располагаются гиперссылки на Notices of the AMS, AMS Blogs, Bulletin of the AMS, страницы в социальных сетях, Privacy Statement.

2. American Physical Society. На СС сайта (www.aps.org) в правом верхнем углу важны такие гиперссылки: Journals (обзоры журналов), Physics Central (Learn How Your World Works), Physics – в том числе содержит ИПС по статьям. В верхнем правом углу СС есть также поле ИПС по сайту, гиперссылки на страницы Login, Become a member. В верхней ленте-меню отметим вкладки: Publications; Meetings and Events; Programs и другие. Они в основном дублируются на правом вертикальном меню-столбце в виде групп гиперссылок с заголовками таких групп. В нижней части СС располагаются гиперссылки на страницы в социальных сетях, а также еще одно меню (содержит четыре пункта с подпунктами).

3. European Mathematical Society (EMS). Подробное описание структуры этого сайта и размещенной на нем информации для версии 2009 г. было сделано в [16]. На СС сайта (http://euro-math-soc.eu/) в верхней

части можно найти: поле ИПС, гиперссылку на страницу Login. Верхняя лента-меню включает вкладки Services, Scientific Activities, Publishing House (У EMS есть свое издательство с отдельным сайтом http://www.ems-ph.org/) . В нижней части СС есть еще одно меню (содержит восемь пунктов с подпунктами), которое в основном дублирует верхнюю ленту-меню. Особо отметим пункт Digital Resources, содержащий обширные архивы публикаций математической направленности. В самом низу СС есть еще одна лента с гиперссылками: Privacy, Conditions, UsefulLinks – это аннотированный список гиперссылок на «тематически смежные» организации глобального и европейского уровней. При этом на http://euro-math-soc.eu/european-applied-math-societies есть дополнительные ссылки на сайты некоторых национальных математических обществ. Однако среди них нет российских.

Внизу СС есть информация «Hosted by: University of Helsinki».

4**.** European Community on Computational Methods in Applied Sciences (ECCOMAS) – сайт http://www.eccomas.org/. На СС ограниченной ширины поле ИПС расположено вверху справа; есть два меню. В верхней ленте-меню (оно видно при переходах на любые страницы) наиболее важны пункты: Scientific Events, Conference Proceedings, Journals. В левом вертикальном меню вкладки содержат, преимущественно, информацию о деятельности ECCOMAS. В центре СС представлены аббревиатуры входящих в ECCOMAS организаций на фоне карты Европы. Для отображения «научных новостей» (порциями по четыре новости) используется лента в правой части СС.

5. The European Consortium for Mathematics in Industry (ECMI). На СС сайта (https://ecmiindmath.org/) есть поле ИПС, верхняя лента-меню с вкладками Research (включая пункты, относящиеся к рабочим группам), Publications (включая журналы, серии книг, годовые отчеты), Conference; «About this blog» – судя по его содержанию владельцы сайта рассматривают по крайней мере часть его как «блог по промышленной математике, обучению и пр.».

Таблица 5

Вебометрические показатели для сайтов тематически специализированных ассоциаций «континентального» уровня (первая часть)*

№	Название ресурса	AT, сек.	Count, URLs			Size, Mb	Scholar Google (SG)
			Text/html	image	application		
1	AMS	0.41	90664	5283	6164	4310.86	65300
2	APS	0.42	31611	13960	4974	7470.18	22
3	EMS	12.07	12440	492	458	1874.05	64
4	ECCOMAS	0.56	181	80	77	3242.42	0
5	ECMI	0.63	3034	0	0	96	2
6	EPS	1.13	6408	91	0	298.05	7
7	FEBS	0.92	1234	521	72	569.51	1

* По количеству объектов типа «Text/html» и «application» преобладает сайт AMS, однако наибольший объем имеет APS. Только на сайте AMS Scholar Google видит достаточно много документов. Сайт EMS имеет длительное время открытия СС.

Таблица 6

Вебометрические показатели для сайтов тематически специализированных ассоциаций «континентального» уровня (вторая часть)

	Название ресурса	Абсолютные количества ссылок				КУнПос. за месяц	КПросм.	AVD, чч:мм:сс
		Входящие	Вн	Исх				
				Ω	Ψ			
1	AMS	15501262	763466	120883	282359	140654	562620	00:02:14
2	APS	2918441	1959433	15710	520542	449506	1798020	00:03:17
3	EMS	1411909	569553	2905	114529	190319	761390	00:01:55
4	ECCOMAS	5635	5829	243	324	7905	31620	00:01:31
5	ECMI	6883	208936	5267	32707	4647	18600	00:01:56
6	EPS	688115	332659	4894	25774	6419	25680	00:02:28
7	FEBS	1540307	14290	661	1809	6044	24180	00:02:31

6. Eropean Physical Society (EPS) — некоммерческая организация, объединяющая физические общества 40 европейских стран и примерно 2500 индивидуальных членов. Девиз: «More than ideas». На сайте (http://www.eps.org/) СС имеет ограниченную ширину; в ее верхнем правом углу есть гиперссылки на страницы Sign In (для регистрации), Join EPS, поле ИПС с названием Community Search. В верхнем ленте-меню отметим вкладки: Support (информация по конференциям), Publications; (в т.ч. для доступа к архивам публикаций), Policy, Resources (включая пункт «Useful Links»). В правой колонке СС есть сегменты: Sign In, Publications (по свежим публикациям), «Women in physics». Ссылки на страницы в социальных сетях немногочисленны, находятся в нижней части сайта.

7. Federation of European Biochemical societies (FEBS). На сайте (https://www.febs.org/) в верхней ленте-меню отметим вкладки News, Our Activity, Our Publications, Members. Для «вертикальной меню-ленты» используется инструмент прокрутки пунктов – они соответствуют вкладке «Our Activity» верхней ленты-меню. В нижней части СС располагаются: поле ИПС для поиска по членам ассоциации в отдельных странах; дополнительное меню, в основном дублирующее вкладки верхней ленты-меню. Также отметим пункты: Terms Of Use, Privacy Policy, Cookies. По последней гиперссылке открывается окно, где декларируется использование средства Google Analytics для сбора статистики о действиях пользователей на сайте, о продолжительности использования собранной информации. В частности для того, чтобы различать пользователей, применяется информация за два года, а в отношении последнего посещения пользователями сайтов – за один год. Указан дизайнер сайта: Web design by Studio 24.

Сводки вебометрических показателей для сайтов указанных в данном разделе ассоциаций этой группы представлены в табл. 5 и 6.

Из табл. 6 видно, что по количеству входящих ссылок (15,5 млн) значительно лидирует сайт AMS – их примерно в 3 раза больше, чем на сайт APS. В тоже время внутренних ссылок на сайте APS значительно больше, чем на AMS. По количествам уникальных посетителей и просмотров значительно лидирует сайт APS, на втором месте находится EMS и только на третьем – AMS. Наибольший показатель AVD у APS (3,17 мин.), далее идут FEBS и EPS, только на 4-м месте находится AMS.

АССОЦИАЦИИ (СОЮЗЫ) ОПРЕДЕЛЕННЫХ ГРУПП СТРАН

1. Исламская организация по вопросам образования, науки и культуры (ISESCO). Помимо англоязычного варианта сайта (https://www.isesco.org.ma/) есть также арабоязычный (https://www.isesco.org.ma/ar/) и франкоязычный (https://www.isesco.org.ma/fr/). В верхней части СС из вкладки «Menu» открывается структурированный по тематическим разделам подробный перечень ссылок, в том числе на раздел Publications. В нижней части СС есть малозаметные гиперссылки на Periodicals, Publications, Digital Library (Majaliss) – в последнем случае демонстрируется лишь фото здания организации.

2. Arab League Educational, Cultural and Scientific Organization (ALECSO). Сайт организации (http://www.projects-alecso.org/) арабоязычный. Однако в его заголовке на СС есть перевод названия организации на английский и французский языки. Применение Google-переводчика позволяет понять смысл текстов на СС кроме тех, которые находятся внутри графических объектов. После перевода СС на русский язык в верхней ленте-меню видны вкладки «Информация и связь», «Область науки и научных исследований». При этом выпадающие списки для этих вкладок отображаются на русском. В левом верхнем углу СС показаны флаги стран-участниц, а в правом – эмблема организации и ее аббревиатура по англоязычному названию (ALECSO). Это единственный сайт из числа рассмотренных в данной статье, при входе на который в явной форме выдается сообщение о «проверке браузера» – видимо на совместимость с сайтом.

3. Международный союз научных и инженерных общественных объединений (МСНиИОО) – общественная организация, членами которой являются (судя по информации на СС сайта) 10 национальных научно-инженерных объединений России, Украины, Казахстана и других стран Содружества Независимых Государств, а также 35 профессиональных обществ и ассоциаций. Сайт (http://www.rusea.info/intindex) – только русскоязычный. В верхней ленте-меню есть вкладки: о союзе (в пункте «Члены и партнеры» приведен перечень из 35 пунктов, но без гиперссылок на сайты организаций), «Мероприятия» (причем на 01.02.2018 г. был план только на 2017 г.), «Печатные издания» – на ней представлен единственный журнал (Наука и технологии в промышленности), причем архива его номеров нет. Сводка ВМП для сайтов этих организаций приведена в табл. 7 и 8.

Таблица 7

Вебометрические показатели для сайтов ассоциаций групп стран (первая часть)

№	Название ресурса	AT, сек.	Count, URLs			Size, Mb	Scholar Google (SG)
			Text/html	image	application		
1	ISESCO	2.97	9361	2411	1626	4898.6	88
Из них:							
	на английском		3562				
	на французском		2170				
	на арабском		3629				
2	ALECSO	-	0 <126>	0	0	0	0
3	МСНиИОО	1.97	6{<1220>}	5	0	0.06	{0}

Таблица 8

Вебометрические показатели для сайтов ассоциаций групп стран (вторая часть)

	Название ресурса	Абсолютное и относительное количества ссылок				КУнПос. за месяц	КПросм.	AVD, чч:мм:сс
		Входящие	Вн	Исх				
				Ω	Ψ			
1	ISESCO	645753	234875	220	28794	16269	65070	00:02:25
Из них:								
	на английском		89667					
	на французском		53276					
	на арабском		91932					
2	ALECSO	118036	0	0	0	2643	10560	00:02:08
3	МСНиИОО	6763	25	33	198	{4307}	{17220}	{00:03:29}

Как и следовало ожидать, показатели для этих сайтов значительно меньше, чем в предыдущих таблицах (нулевые значения для ALECSO – результат недоступности сайта для внешних программ в отношении анализа по рассматриваемым показателям).

Судя по табл. 8, количество входящих ссылок на сайт ISESCO достаточно большое. Однако количество уникальных посетителей и просмотров даже для него невелико. При этом показатель AVD для ISESCO находится на «среднем» уровне по сравнению с сайтами, рассмотренными в предыдущих разделах.

ВЫВОДЫ

Рассматриваемые в настоящей статье сайты групп организаций играют ключевую роль в поддержке деятельности этих организаций и агрегации научно-технической информации, особо отметим материалы публикаций и сведения о научных мероприятиях. Интерфейсы рассмотренных сайтов меняются с течением времени; не унифицированы (даже в пределах групп организаций с близкими областями деятельности). На многих сайтах ширина экранов мониторов используется не полностью, что может говорить об отсутствии их «адаптивности» к характеристикам применяемых пользователями устройств дистанционного доступа к информации. Практически на всех рассмотренных сайтах есть ИПС, ссылки на сайты социальных сетей. Информация об использовании cookies и направлениях их применения отражается лишь на части сайтов. Сведения о датах последних актуализаций сайтов (в том числе их отдельных страниц) приведены лишь в редких случаях. Для стартовых страниц рассмотренных сайтов характерно дублирование средств доступа к страницам с наиболее важной информацией. На многих сайтах есть средства регистрации пользователей, возможности их входа по логину и паролю. На сайтах международных научных ассоциаций значительно различаются объемы размещенной информации, в ряде случаев они сопоставимы по объему с сайтами, рассмотренными в [1, 2]. Для формирования контента сайтов научных ассоциаций используются как внутренние, так и внешние источники информации. Посещаемость рассмотренных сайтов зависит от ряда факторов: объема размещенной информации, её научной значимости, актуальности; удобства использования меню сайтов, поисковых систем и тематических рубрикаторов, предназначенных для обеспечения доступа к информации; международной известности сайтов; количества гиперссылок на сайты и т.д.

СПИСОК ЛИТЕРАТУРЫ

1. Брумштейн Ю.М., Васьковский Е.Ю. Анализ вебометрических показателей основных сайтов, агрегирующих политематическую научную информацию // Научно-техническая информация. Сер. 2. – 2017. – № 11 – С. 16–32; Brumshteyn Yu.M., Vas'kovskii E.Yu. Analysis of the Webometric Indicators of the Main Websites that Aggregate Multithematic Scientific Information // Automatic Documentation and Mathematical Linguistics. – 2017. – Vol. 51, № 6. – P. 250–265.
2. Брумштейн Ю.М., Васьковский Е.Ю. Исследование функциональности и вебометрических показателей специализированных сайтов, связанных с научной деятельностью // Научно-техническая информация. Сер. 2. – 2018. – № 1. – С. 16–30; Brumshteyn Yu.M., Vas'kovskii E.Yu. Studying the Functionality and Webometric Indicators of Specialized Science-Related Websites // Automatic Documentation and Mathematical Linguistics. – 2018. – Vol. 52, № 1. – P. 7–23.
3. Белл Э. Создание международного потенциала для обеспечения устойчивого роста: роль научных обществ // Форсайт. – 2010. – Т. 4, № 1. – С. 60–63.
4. Бетева Н.В. Управление информационным пространством в XXI в. // Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики. – 2014. – № 12-1 (50). – С. 31–33.

5. Борзов М.А. Проблемы интеграции России в современное информационное пространство // Сервис plus. – 2010. – № 3. – С. 64–70.
6. Кабанов Ю.А. Информационное пространство как новое (гео)политическое пространство: роль и место государств // Сравнительная политика. – 2014. – Т. 5, № 4 (16–17). – С. 54–59.
7. Зиновьева Е.С. Возможности России в глобальном информационном обществе // Вестник МГИМО Университета. – 2016. – № 3 (48). – С. 17–29.
8. Лагно А.Р. Обзор XIX международной конференции «Science Online: электронные информационные ресурсы для науки и образования» // Государственное управление. Электронный вестник. – 2015. – № 49. – С. 276–287.
9. Арский Ю.М., Гиляревский Р.С., Клещев Н.Т., Лаверов А.Н., Родионов И.И., Цветкова В.А. Информационное пространство новых независимых государств. – М.: ВИНИТИ, 2000. – 200 с.
10. Андреева Е.Л., Захарова В.В., Ратнер А.В. Научно-технологическое сотрудничество России в условиях становления международного экономического партнерства нового формата // Известия Уральского государственного экономического университета. – 2016. – № 6 (68). – С. 132–140.
11. Бударина Н.А. Международные программы научно-технического сотрудничества в рамках Европейского союза и Евразийского экономического сообщества // Вестник Полоцкого государственного университета. Серия D: Экономические и юридические науки. –2012. – № 6. – С. 112–122.
12. Калинин Ю.П., Хорошилов А.А., Хорошилов А.А. Принципы создания системы мониторинга и анализа мирового потока научно-технической информации // Системы и средства информатики. – 2016. – Т. 26, № 1. – С. 139–165
13. Брежнева В.В., Гиляревский Р.С. От информационного обслуживания к информационному менеджменту // Научно-техническая информация. Сер. 1. – 2015. – № 5. – С. 7–9.
14. Титова Т.П. Развитие международной мобильности научных кадров: опыт европейского союза // Социология науки и технологий. – 2015. – Т. 6, № 1. – С. 90–97.
15. Комаров С.Ю. Википедия как средство продвижения информационных ресурсов // Библиосфера. – 2012. – № 5. – С. 38–40.
16. Информационная система математических интернет-ресурсов MathTree / отв. ред. О.А. Клименко. – Новосибирск: Изд-во СО РАН, 2009. – 288 с.
17. Сюнтюренко О.В., Гиляревский Р.С. Задачи информационного обеспечения инновационного развития экономики и роль инжиниринга // Научно-техническая информация. Сер. 1. – 2017. – № 5. – С. 1–14; Syuntyurenko O.V., Gilyarevskii R.S. Tasks of Information Support of Innovative Economic Development and the Role of Engineering // Scientific and Technical Information Processing. – 2017. – Vol. 44, № 2. – P. 107-118.
18. Быстрицкий Н.Д. Алгоритм анализа интернет-страниц информационного ресурса // Фундаментальные исследования. – 2015. – № 6–3. – С. 443–446.
19. Ронжин Д.И., Мамонтов С.А., Тютюнник В.М. Особенности мультиязычного вебсайта научной организации // Психолого-педагогический журнал Гаудеамус. – 2013. – № 2 (22). – С. 174–176.
20. Спрысков А.А., Бидуля Ю.В. Агрегация и суммаризация текстов поисковой выдачи по запросу пользователя // Математическое и информационное моделирование сборник научных трудов. – Тюмень, 2017. – С. 435–439.
21. Орлова Ю.А. Методы адаптации текстовой информации для лиц с ограниченными возможностями по зрению // Прикаспийский журнал: управление и высокие технологии. – 2015. – № 4. – С. 210–21
22. Осминин П.Г. Современные подходы к автоматическому реферированию и аннотированию // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. – 2012. – № 25. – С. 134–135.
23. Резаиан Н., Новикова Г.М. Система автоматического реферирования текста на персидском языке // Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем. – Материалы Всероссийской конференции с международным участием. – М.: РУДН, 2016. – С. 163–165.
24. Кравец А.Д., Петрова И.Ю., Кравец А.Г. Агрегация информации о перспективных технологиях на основе автоматической генерации интеллектуальных агентов мультиагентных систем // Прикаспийский журнал: управление и высокие технологии. – 2015. – № 4 (32). – С. 141–148.
25. Aguillo I.F. Measuring the institutions' footprint in the web // Library Hi Tech. – 2009. – Vol. 27(4). – P. 540–556.
26. Fan W. Contribution of the institutional repositories of the Chinese Academy of Sciences to the webometric indicators of their home institutions // Scientometrics. – 2015. – Vol. 105. – P. 1889. DOI:10.1007/s11192-015-1758-4.
27. Thelwall Mike. Data Science Altmetrics //J DIS – Journal of Data and Information Science. – 2016. – Vol. 1, № 2. – P. 7–12. DOI: 10.20309/jdis.201610
28. Вараксин М.А. Инструменты для мониторинга посещаемости сайтов // Информационные системы и технологии в образовании, науке и бизнесе (ИСиТ-2014). Материалы Всероссийской молодежной научно-практической школы. – Кемерово: Кузбасский гос. тех. ун-т им. Т.Ф. Горбачева, 2014. – С. 161-162.
29. Рю Д. Оценивание и ранжирование веб-сайтов. Вебометрические рейтинги // Научный редактор и издатель. – 2017. – Т.2, №1. – С.14–17
30. Скородумов П.В., Холодев А.Ю. Анализ популярности веб-сайта научной организации с помощью различных систем сбора статистических данных // Вопросы территориального развития. – 2016. – № 1 (31). – С. 7.

31. Васьковский Е.Ю., Брумштейн Ю.М. Системный анализ функциональных возможностей счетчиков посещаемости сайтов // Прикаспийский журнал: управление и высокие технологии. – 2015. – №3. – С. 45–58
32. Гуськов А.Е., Быховцев Е.С., Косяков Д.В. Альтернативная вебометрика: исследование вебтрафика сайтов научных организаций // Научно-техническая информация. Сер. 1. – 2015. – № 12. – С. 12–28.
33. Дербишер В.Е., Силина А.Ю., Дербишер Е.В. Формализованные оценки информационных потоков для управления научной деятельностью // Информатизация образования и науки. – 2012. – № 16. – С. 113–132.
34. Брумштейн Ю.М., Васьковский Е.Ю., Куаншкалиев Т.Х. Поиск информации в Интернете: анализ влияющих факторов и моделей поведения пользователей // Известия Волгоградского государственного технического университета. Сер. Актуальные проблемы управления, вычислительной техники и информатики в технических системах. – 2017. – № 1 (196). – С. 50–55.
35. Брумштейн Ю.М., Бондарев А.А., Федотова А.В., Иванова М.В. Отражение научной деятельности региональных вузов на сайтах в Интернете: системный анализ вопросов информационной безопасности // Прикаспийский журнал: управление и высокие технологии. – 2014. – № 2 (26). – С. 85–100.
36. Дровникова И.Г., Алферов В.П., Змеев С.А., Хвостов А.В., Окрачков А.А., Макаров О.Ю. О показателях информационной безопасности элементов типового многоуровневого web-сайта // Технологии техносферной безопасности. – 2014. – № 6 (58). – С. 31.
37. Печников А.А. Структура веб-сайта: пример мелкозернистого исследования // Дистанционное и виртуальное обучение. – 2016. – № 8 (110). – С. 114–124.
38. Гулин К.А., Скородумов П.В. Интернет-портал как средство популяризации деятельности научной организации // Проблемы развития территории. – 2015. – № 5 (79). – С. 52–65.

Материал поступил в редакцию 12.02.2018.

Сведения об авторах

БРУМШТЕЙН Юрий Моисеевич – кандидат технических наук, доцент Астраханского государственного университета, доцент
e-mail: brum2003@mail.ru
ORCID http://orcid.org/0000-0002-0016-7295

ВАСЬКОВСКИЙ Евгений Юрьевич – аспирант кафедры информационных технологий Астраханского государственного университета, ведущий программист отдела Internet-технологий Астраханского государственного университета
e-mail: vaskovskiy_evgeniy@mail.ru
ORCID http://orcid.org/0000-0002-4937-3305

УДК [001.32(470):002–047.44]:004.65

Е.В. Бескаравайная, Т.Н. Харыбина

Сравнение библиометрических показателей некоторых лабораторий научного учреждения РАН

В настоящее время библиометрические исследования с использованием соответствующих баз данных являются не только востребованными в НИИ, вузах и библиотеках, но и широко применяются в сфере управления наукой. Рассматривается модель библиометрического анализа публикационного потока конкретных лабораторий. Особое внимание уделяется результатам библиометрического анализа, проведенного с учетом разработанных критериев. Подчеркивается важная роль, которую играют библиотеки при распространении библиометрических данных по запросам пользователей.

Ключевые слова: *библиометрический анализ, библиометрия, научная продуктивность, научные исследования, библиометрические базы данных, публикационная активность, индексы цитирования, индекс Хирша (h-индекс), импакт-фактор*

ВВЕДЕНИЕ

С внедрением информационных технологий и различных автоматизированных процессов увеличивается роль научной Библиотеки, как информационного центра, который наряду с уже традиционными формами обслуживания пользователей стремится предоставить новые виды и способы библиотечно-информационного сервиса.

В настоящее время в научных библиотеках одним из востребованных направлений их деятельности становятся библиометрические исследования. Проводимые мониторинги информационных потребностей пользователей Центральной библиотеки в Пущинском научном центре РАН (отдел Библиотеки по естественным наукам РАН) подтверждают заинтересованность читателей в получении информации такой направленности. В контексте информационного обслуживания поиск и предоставление подобных сведений рассматривается, как перспективная услуга в работе библиотеки.

В качестве предмета библиометрических исследований, как правило, выступает научная статья – это и монографии, и журнальные статьи, и обзоры, и целый ряд других печатных и электронных материалов. Суммируя в себе результаты деятельности ученых и выполняя информационную (тематическую) функцию для научного сообщества, статьи также осуществляют аналитическую и прогностическую функцию, сами становясь объектом научного исследования.

Количество ссылок, как наукометрический индикатор, может служить одним из показателей “продуктивности ученого” наряду с числом публикаций и/или патентов, наград, грантов, стипендий, участием в международных научных обществах, редколлегиях научных журналов.

Анализ библиометрических показателей дает возможность:

- оценить общую продуктивность научного коллектива и вклад отдельных специалистов согласно разработанным критериям;
- получить информацию о динамике и тенденциях развития той или иной научной области,
- определить наиболее выдающихся ученых и их значимые работы;
- выявить круг перспективных изданий для последующих публикаций в них статей научных сотрудников;
- определить электронные ресурсы, максимально отражающие тематику исследуемого научного направления.

Возросшее значение, библиометрических исследований, прежде всего, связано с внедрением рейтинговых систем оценки и анализа научной продуктивности ученых, организаций. По этой же причине, как показывают результаты изучения информационных потребностей пользователей, библиометрические исследования пользуются особым спросом не только у администраторов учреждений, но и вызывают несомненный интерес в самой научной среде в контексте понимания соответствия мировым трендам научного развития. Сегодня, ученые желают получать библиометрическую информацию по разным направлениям: Научному Центру, научно-исследовательскому институту (НИИ), по конкретному научному подразделению (лаборатории, отделу или сектору) в своем НИИ. В связи с этим, в данной статье мы предлагаем модель библиометрического анализа публикационного потока, ориентированную на ведущие лаборатории (на примере Института биофизики клетки РАН (ИБК РАН)).

В ходе реализации данного направления была разработана система индикаторов, позволяющая обеспечить сбор и анализ сведений о количественном

и качественном составе научных публикаций ученых как индикаторе вклада в развитие науки. В то же время мы хотим подчеркнуть, что использование библиометрических показателей в вопросе разработки оценки эффективности научной деятельности ученых является целесообразным, а в совокупности с другими критериями может служить мерой научной активности, как для небольших групп (лабораторий, диссертационных советов, редколлегий журналов, научных школ), так и институтов, в целом.

МЕТОДИКА ИССЛЕДОВАНИЯ И РЕЗУЛЬТАТЫ

Поиском методов оценки и анализом научной продуктивности ученых, организаций и стран занимаются ученые во всем мире [1–11], поэтому очевидным является интенсивный рост публикаций по данной проблеме, как в нашей стране, так за рубежом [12–13]. Если рассматривать библиометрику, как услугу библиотеки, то за последнее десятилетие это направление деятельности прочно вошло в практику библиотек. Этот факт подтверждают в своей статье специалисты университетской библиотеки Австралии, где, по их утверждению, происходит смещение вектора внимания научных библиотек от услуг для читателя к услугам для автора-ученого [14, с. 261]. Библиометрические исследования являются значимыми и для сотрудников библиотеки Венского университета. В своей статье они делятся опытом создания факультета библиометрии в Венском университете и рассказывают о его практической деятельности. Авторы подчеркивают, что библиометрические исследования являются для них инновационным видом обслуживания научного и административного персонала университета [15].

Библиотеки, располагая сегодня необходимой информационной базой и владея различными поисковыми технологиями, имеют широкие возможности для проведения наукометрических исследований. Сотрудники Центральной библиотеки в Пущинском научном центре РАН (ПНЦ РАН) – отдел Библиотеки по естественным наукам РАН (БЕН РАН) стараются предложить своим пользователям новые услуги и сервисы, разработать интересные методики для исследования конкретных научных направлений с использованием библиометрических методов [16–18]. Располагая большим спектром информационных ресурсов, таких как *«Web of Science CC»; «Scopus» – Elsevier*; «Российский индекс научного цитирования» (РИНЦ) – ООО «Научная электронная библиотека»; *«Chemical Abstracts» – CAS, MedLine* и др., мы имеем возможность проводить данные исследования, ориентируясь на разные категории пользователей – от ученого до администрации научного Центра. За 15 лет на основе системного подхода библиотекой была разработана методика, предназначенная для мониторинга и оценки научно–инновационного потенциала учреждений РАН, включающая в себя все основные библиометрические показатели публикационной и патентной активности ученых. Со временем разработанная система критериев изменялась и совершенствовалась под влиянием новых требований оценки научной работы учреждения.

В настоящей статье предлагается модель библиометрического анализа публикационного потока, ориентированную на отдельную лабораторию (на примере Института биофизики клетки РАН (ИБК РАН)). Данное направление сформировалось в результате проведении опроса сотрудников институтов ПНЦ РАН в рамках изучения информационных потребностей. Мы выяснили, что ученые желают получать информацию по основным библиометрическим индикаторам не только по научному Центру или институту для составления отчетов в различные вышестоящие организации, но и по лаборатории, отделу или сектору в своем НИИ.

Объектами исследования стали четыре лаборатории Института биофизики клетки РАН со своими собственными уникальными тематиками (табл. 1).

Таблица 1

Направления исследований анализируемых лабораторий

Подразделение	Направления исследований
Лаборатория 1.	1. Биофизические аспекты функционирования клетки. Механизмы рецепции и внутриклеточной сигнализации в различных типах клеток
	2. Медицинские аспекты физиологии клетки и клеточной биофизики. Разработка подходов к созданию новых медицинских препаратов, к диагностике и лечению различных заболеваний
Лаборатория 2	Изучение сигнальных процессов, протекающих в клетках при их возбуждении, реконструкция клеточных компонент в модельных системах, исследование механизмов молекулярного узнавания
Лаборатория 3	Изучение роли белков теплового шока в индукции противовирусного и противоопухолевого иммунитета, исследование экспрессии белков теплового шока в вирус-инфицированных клетках, выяснение тонкой эпитопной структуры gE и gB и локализация иммунологически важных областей
Лаборатория 4.	1. Моделирование структурной организации бактериальных промоторов и поиск их в бактериальных геномах
	2. Использование сигналов транскрипции для идентификации мест кодирования новых РНК-продуктов в геноме E.coli
	3. Изучение молекулярно-генетических механизмов возникновения стрессовых состояний у разных видов организмов (температурный шок, развитие гибридного дисгенеза)

Эти лаборатории выбраны не случайно. Их открытия имеют мировое значение и стоят на одном уровне с зарубежными исследованиями по биофизике. Так, например, благодаря исследованиям одной из лабораторий было впервые зарегистрировано распространение Ca^{2+} – волны в невозбудимых клетках, сопряженное с секрецией АТФ и обеспечивающее согласованное поведение клеточной популяции в ответ на стимуляцию определенных рецепторов. Этот тип межклеточной коммуникации носит универсальный характер и играет важную роль в регуляции функционирования клеток и тканей. Занимаясь клетками вкусовой почки млекопитающих I, II и III типов, сотрудники другой лаборатории обнаружили, что только клетки типа II, специализирующиеся в распознавании горьких, сладких и умами (вкус некоторых аминокислот) веществ, высвобождают АТР при их стимуляции, в то время как клетки I и III типов не способны секретировать нуклеотид. Полученные данные объясняются наличием уникального механизма для кодирования вкусовой информации. Сотрудниками еще одной лаборатории были заложены основы для целенаправленного создания субъединичных вакцин и средств специфической профилактики болезни Ауески (БА), что позволило разработать методы диагностики нового поколения, которые могут быть использованы в программах искоренения заболевания. Таким образом, данные подразделения являются ведущими в деятельности института и их показатели важны, как для научных сотрудников, так и для администрации учреждения.

Информация о составе лабораторий была взята на сайте ИБК РАН (http://www.icb.psn.ru/ index.php?option=com_content&view=featured&Itemid=101).

В качестве информационной базы для нашего исследования служил ресурс компании *CLARIVATE ANALYTICS «Web of Science Core Collection» (WOS CC)*: период доступа с 1975 г.

Критерии исследования в нашей модели делились на два направления: во-первых, анализ публикационной активности самих ученых, работающих в исследуемых лабораториях, включающий, анализ количества публикаций, их цитирование, h-индекс, сведения о финансовой поддержке и наличию грантов; во- вторых, анализ *цитирующих* публикаций, сведений об их издании, аффилиации авторов, тематике исследований.

Для нахождения всех публикаций автора производился поиск по фамилии с усечением и одним инициалом, вариантами разночтения при написании на иностранном языке. Затем каждая публикация разбиралась индивидуально, удалялись однофамильцы. Для создания таблиц по лабораториям в целом (табл. 2) к фамилии автора в обязательном порядке добавлялась его аффилиация. Суммарные показатели публикационной активности для каждой лаборатории собирались без повторов, а публикации сотрудников и их цитирование учитывалось только один раз.

Результатом работы ученых (за весь период по БД *WOS CC*) стали публикации в 19 книгах и научных сборниках, в 154 различных российских и зарубежных научных журналах. Анализирую табл. 3, мы видим, что по БД *WOS CC*, наибольшей популярностью пользуются журналы «Биологические мембраны», «Бюллетень экспериментальной биологии и медицины», «Биохимия», «Доклады академии наук». А снижение рейтинга профильного издания «Биофизика», скорее всего, связано с перерывом в его индексировании в данной базе. Из иностранных журналов чаще всего статьи публикуются в *«Journal of Biomolecular Structure & Dynamics», «FEBS Letters», «FEBS journal», «Biochimica et Biophysica Acta – Biomembranes», «PLOS ONE».* Исходя из анализа табл. 3, отсортированной по количеству публикаций за весь период и за последние 5 лет, можно сделать вывод, что у сотрудников лабораторий ИБК РАН на протяжении длительного срока признанием пользуются одни и те же издания, и эта популярностью не меняется с течением времени.

Несмотря на то, что исследования всех лабораторий входят в проблемно-тематический план научно-исследовательских работ по физико-химической биологии, научные интересы отдельных групп весьма разнообразны. Наряду с общими направлениями по *Biochemistry & Molecular Biology, Cell Biology, Biophysics, Multidisciplinary Sciences in Biology*, только у Лаборатории 3 есть работы по почвоведению; по физике атомного ядра – у Лаборатории 1; вирусологией занимаются исключительно сотрудники Лаборатории 2; а в Лаборатории 4 присутствуют статьи по спектроскопии и ядерно-магнитному резонансу. На рис. 1 представлены тематические категории, по которым было написано наибольшее количество работ.

Таблица 2

Сводные данные по подразделениям

Лаборатория	**Кол-во чел.**	**Кол-во публ. в *WOS CC***	**Цитировование по *WOS CC***	**Кол-во публ. в *WOS CC* за 2012-2016 гг**
1	15	154	1261	55
2	14	119	2142	58
3	14	134	658	46
4	20	68	378	19
Всего	63	475	4439	178

Таблица 3

Издания, в которых опубликовано наибольшее количество статей (по БД WOS CC)

Место в рейтинге	За весь период	Кол-во записей	За 2012-2016 гг.	Кол-во записей
1	Biologicheskie Membrany	57	Biologicheskie Membrany	35
2	Biofizika	52	Bulletin of Experimental Biology and Medicine	10
3	Bulletin Of Experimental Biology and Medicine	21	Doklady Physical Chemistry	9
4	Biochemistry-Moscow	21	Doklady Biochemistry and Biophysics	6
5	Molecular Biology	18	Russian Journal of Physical Chemistry B	6
6	FEBS Letters	12	Journal of Biomolecular Structure & Dynamics	5
7	Doklady Akademii Nauk SSSR	10	Biochemistry-Moscow	5
8	Doklady Physical Chemistry	9	FEBS Journal	5
9	Journal of Biomolecular Structure & Dynamics	9	Biochimica et Biophysica Acta-Biomembranes	4
10	Eurasian Soil Science	8	Biofizika	4
11	Tsitologiya	6	Eurasian Soil Science	4
12	Doklady Biochemistry and Biophysics	6	PLOS ONE	4
13	Russian Journal of Physical Chemistry B	6	Russian Chemical Bulletin	3
14	Bioelectromagnetics	6	Molecular Biology	3
15	Biochimica et Biophysica Acta-Biomembranes	5	Biochemical and Biophysical Research Communications	3
16	Journal of Biological Chemistry	5	International Journal drfs Radiation Biology	2
17	Instruments and Experimental Techniques	5	Biochemical Society Transactions	2
18	FEBS Journal	5	Biochemistry Moscow Supplement Series A-Membrane and Cell Biology	2
19	Nucleic Acids Research	5	Journal of Virological Methods	2
20	Biochimica et Biophysica Acta	5	Bioelectromagnetics	2
21	Russian Chemical Bulletin	5	Journal of Environmental Radioactivity	2
22	Biochemistry International	4	Neurophysiology	2
23	Neuroscience Letters	4	Journal of Cell Science	2
24	Journal of Virological Methods	4	Quaternary International	2
25	Biochemical and Biophysical Research Communications	4	Pflugers Archiv-European Journal of Physiology	2
26	PLOS ONE	4	Journal of Bioinformatics and Computational Biology	2
27	Pflugers Archiv-European Journal of Physiology	4	International Journal of Biochemistry & Cell Biology	2

Изучая динамику публикационной активности, мы можем сделать вывод, что наиболее стабильно публикует свои работы Лаборатории 1 и 3 (рис. 2), а наименее устойчивое положение у Лаборатории 4, что в равной степени относится и к цитированию (рис. 3). Однако, лаборатория 4 является продолжением и развитием отечественной научной школы по исследованию механизмов регуляции транскрипции, которая была преобразована в лабораторию в 2006 г. и, соответственно, на сегодняшний день представляет собой наиболее «молодое» из анализируемых подразделений. Высокий уровень исследований и большое количество молодых сотрудников дает нам право предполагать быстрый рост ее публикационной активности и цитируемости в будущем.

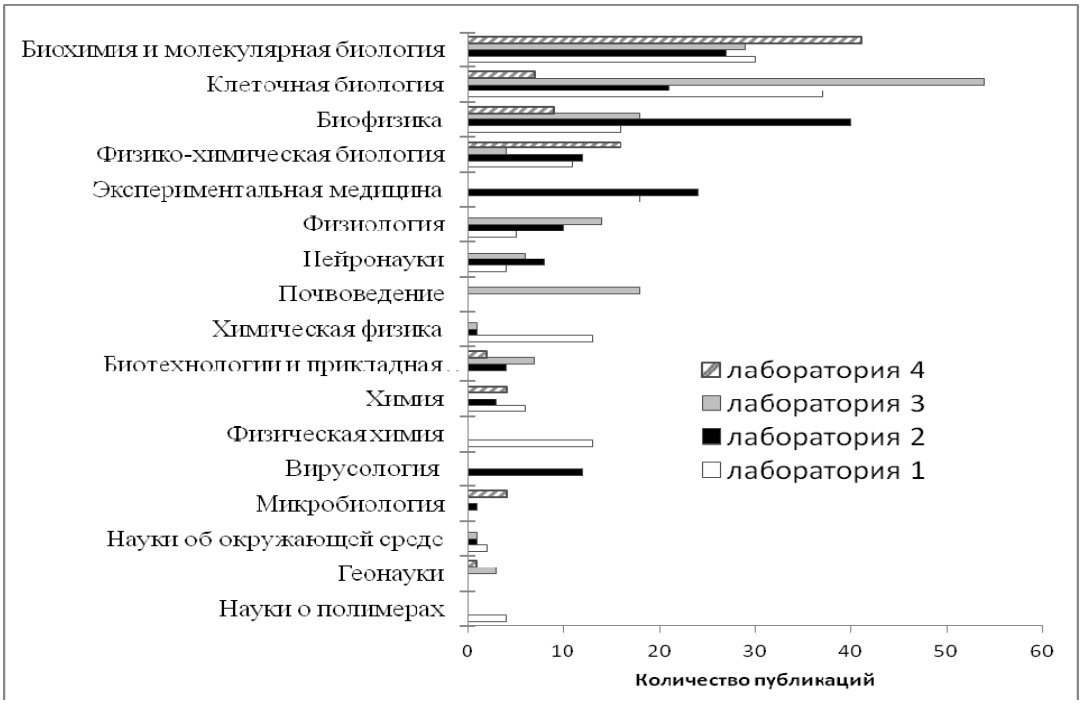

Рис. 1. Тематические направления, по которым было написано наибольшее количество работ по БД *Web of Science, SCOPUS* и РИНЦ сотрудниками исследуемых лабораторий

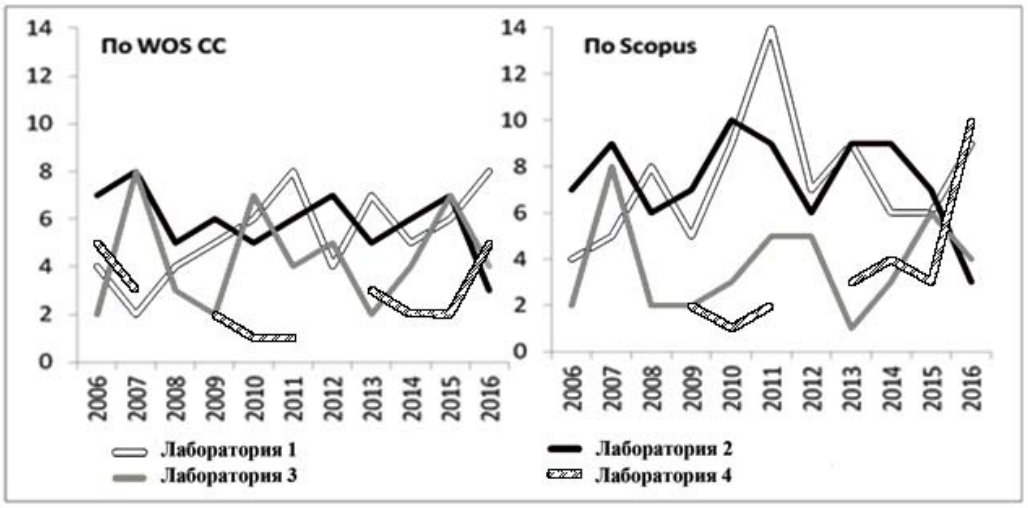

Рис. 2. Динамика публикационной активности за 2012-2016 гг. по БД *WOS CC* и *Scopus*

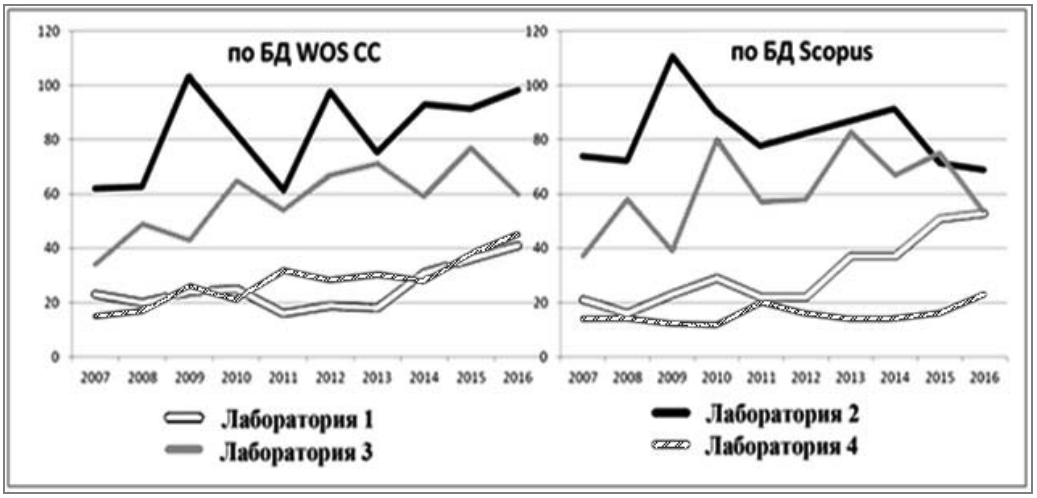

Рис. 3. Динамика цитирования всех публикаций лабораторий в период 2007–2016 гг. по БД *WOS CC* и *Scopus*

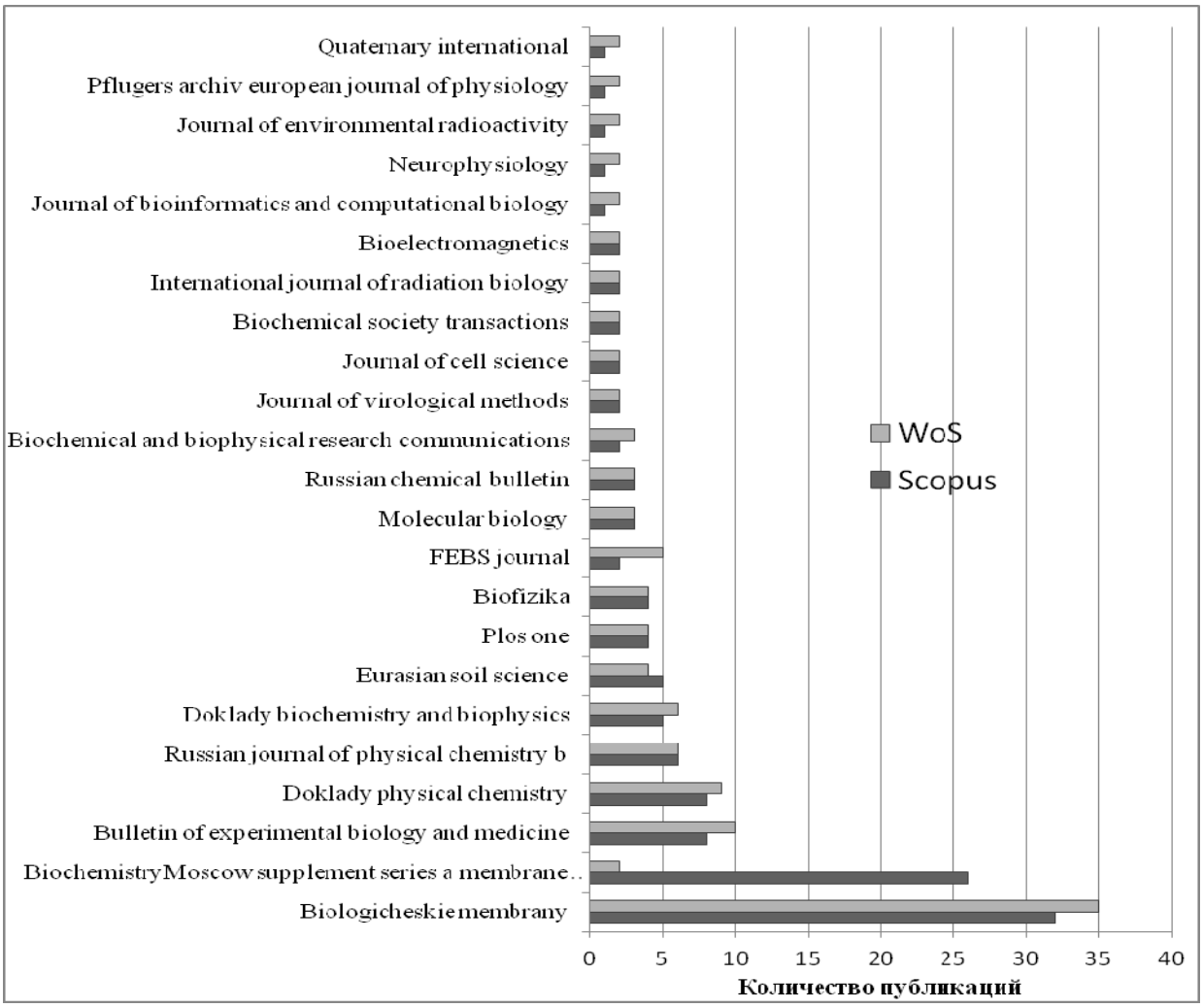

Рис. 4. Журналы, индексируемые в WOS CC и Scopus с наибольшим количеством публикаций сотрудников исследуемых лабораторий

Сравнивая динамику публикационной активности в основных международных базах, мы определили издания, которым отдается предпочтение в каждой из баз. Всего из 120 журналов, в которых были опубликованы труды сотрудников данных лабораторий, 102 индексируются в обеих базах (рис. 4).

Приведя названия источников к единообразию мы установили, что 42% изданий, публикующие статьи сотрудников исследуемых лабораторий, имеются только в *WOS CC*, а 35% только в *Scopus*. При этом довольно часто попадались издания, которые вошли в базу позже исследуемых лет, поэтому статей сотрудников лабораторий базе нет, хотя, на сегодняшний момент сами журналы в *Scopus* индексируются.

Далее мы проверили постатейно каждую публикацию и выяснили что даже когда источники индексируются в обеих базах, зачастую статьи присутствуют только в одной из них. Такое положение возможно, во-первых, если журнал не полностью переводится на иностранный язык, во-вторых, если издание является сборным и публикует избранные статьи из разных журналов, в-третьих, если публикация попадает в базу из списков цитирования (рис. 5) Данный анализ позволяет нам выявить источники, стабильно индексируемые в одной и более базах (включая российские журналы), ценные с точки зрения дальнейшей перспективы предоставления работ в печать.

Обратимся к анализу цитирования. С целью определения динамики роста цитирования по каждой из баз, мы выбрали статьи, имеющие цитирование в обеих базах уже на следующий год после издания (рис. 6). На начальном этапе (в год издания статьи) цитирование в *Scopus* равно или незначительно превышает цитирование в *WOS*. Кроме того, анализируя непосредственно каждую публикация, мы обнаружили, что в крупных изданиях, таких, как *BBA, JOURNAL OF CELL SCIENCE, ACTA BIOTHEORETICA, CYTOTHERAPY* цитирование в обеих базах появляется почти одновременно. В отношении российский переводных изданий, в которых напечатано большинство работ исследуемых лабораторий, сами журналы и их цитирование появляется с небольшим опозданием.

В процессе изучения закономерностей в различии цитирования мы пришли к выводу, что рост цитирования в ближайший от момента публикации год в *SCOPUS* идет за счет учета цитирования в российских журналах дополнительно к их переводной версии (табл. 4).

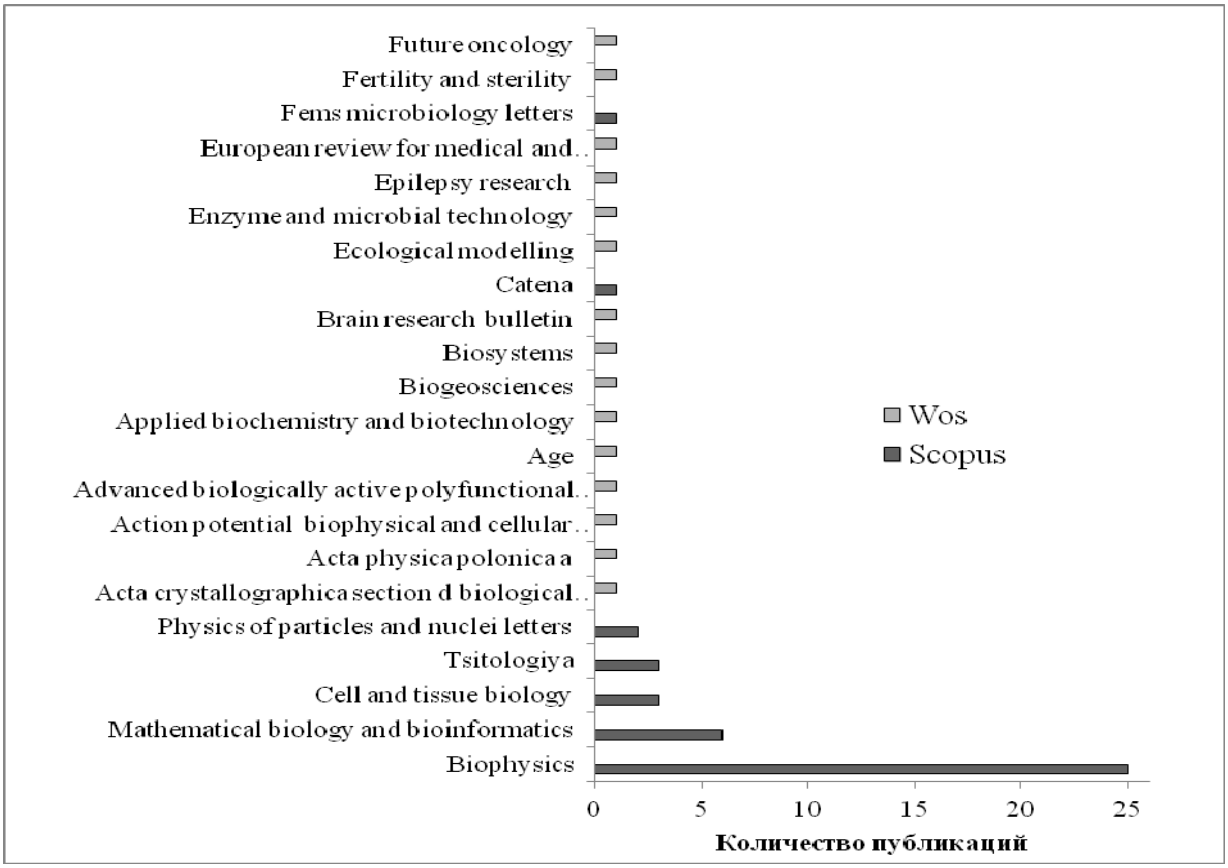

Рис. 5. Выборочная индексации изданий в БД WoS и Scopus на примере публикаций лабораторий ИТЭБ

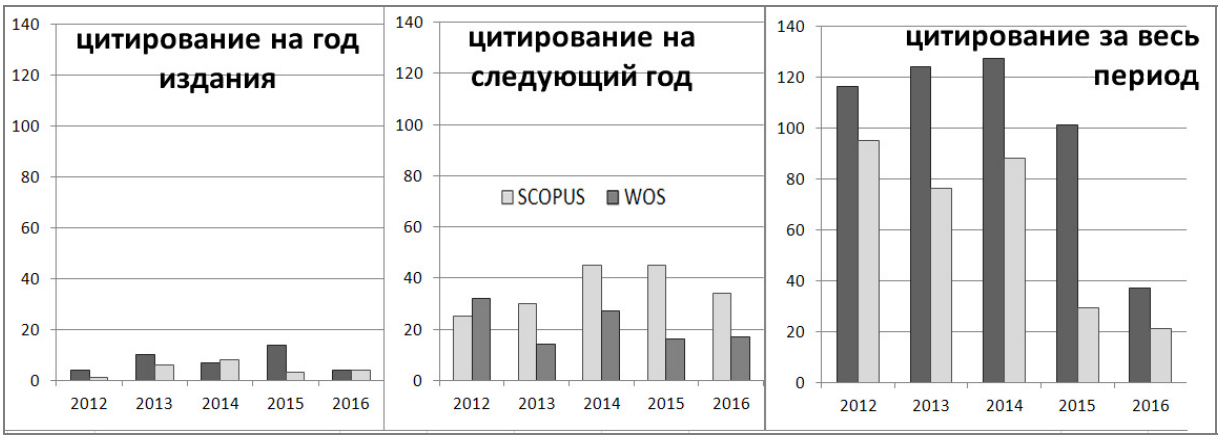

Рис. 6. Динамика цитирования публикаций в *WOS CC и SCOPUS*

Таблица 4

Пример цитирования публикаций в российских изданиях и переводных версиях в *БД SCOPUS*

ПЕРЕВОДНЫЕ И РОССИЙСКИЕ ИЗДАНИЯ В БД *SCOPUS*	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	ОБЩИЙ ИТОГ
Biochemistry (Moscow) Supplement Series A: Membrane and Cell Biology			2	1	1	3		2			9
Biologicheskie Membrany	2		3	1	3		2				11
Cell and Tissue Biology	1										1
Tsitologiya	2										2
Biophysics (Russian Federation)	1			1	4	2				1	9
Biofizika		1			2	1					4

Таким образом, включение в БД *SCOPUS* российских журналов увеличивает количество цитирования почти вдвое. Однако, нужно понимать, что разбивка цитирования одной статьи на несколько частей (отдельно для российской, отдельно для иностранной версий), приводит к уменьшению индекса Хирша авторов публикации.

Одним из важнейших аспектов работы научного коллектива является информация о высокоцитируемых публикациях. По сведениям баз данных *Web of Science* и *Scopus* для исследуемых лабораторий есть статьи, цитирование которых превышает отметку 100. Списки таких статей совместно с публикациями в изданиях с высоким импакт – факторомм ежегодно предоставляются администрации института и служат одним из аргументов при распределении премий.

Другим важным критерием оценки деятельности является международное сотрудничество. Среди иностранных организаций, выполняющих совместные работы в 2007 – 2016 гг. с лабораториями ИБК РАН можно назвать такие, как *Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, United States; Wenner-Gren Institute, Stockholm, Sweden; UCD Conway Institute, University College Dublin, Ireland; Ctr. for Molec. Biology and Medicine, Epworth Hospital, Melbourne, Australia*; *Department of Molecular Genetics, National Institute of Genetics, Mishima, Japan; Department of Structural Biology, The Weizmann Institute of Science, Rehovot, Israel* и др.

Чаще других партнерами в работе становятся ученые из США, Италии, Великобритании (следует обратить внимание, что в зарубежных базах публикации считаются отдельно для Англии и Уэльса); реже - Германии, Испании. А вот публикации со специалистами из бывших союзных республик редки: 2 публикации с Казахстаном и 1 с Узбекистаном за 10 лет (рис. 7).

Существенным показателем, демонстрирующим высокий научный уровень исследований, является факт индексации в международных библиометрических базах работ сотрудников ИБК РАН, написанных без иностранного участия (27%). Кроме этого, еще 20% статей написано лабораториями совместно с исследователями из других российских институтов: Института биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова РАН, Казанского физико-технического института имени Е.К. Завойского РАН, Московского государственного университета, Московского физико-технологического института, Тульского государственного университета.

Особое внимание при анализе публикационной активности лабораторий мы уделили научным направлениям, получившим дополнительную финансовую поддержку от различных научных фондов. Этот вопрос наиболее часто стоит перед сотрудниками при выполнении по грантам работ, тематически укладывающихся в несколько близких категорий. Наиболее активно финансировались публикации по тематическим направлениям, представленным в табл. 5.

Наиболее часто финансирование исследований осуществляли Российский фонд фундаментальных исследований, Фонд науки и образования Министерства образования и науки РФ, Российский научный фонд – некоммерческая организация, созданная для поддержки фундаментальных и поисковых научных исследований; Российский фонд технологического развития, президентская программа поддержки молодых ученых, Фонд «Сколково», Группа ОНЭКСИМ – частный российский инвестиционный фонд, Благотворительный фонд "Новая Мысль". Что касается иностранных фондов, поддержку оказывали Министерство экономики, промышленности и конкурентоспособности Испании, *AGAUR* – фонд, предоставляющий материальную помощь студентам и аспирантам в Испании, Национальный центр научных исследований Франции, Немецкое научно-исследовательское общество, Европейский исследовательский Совет. Названные при финансовом участии названных фондов работы являются совместными с иностранными, научными, образовательными или медицинскими организациями.

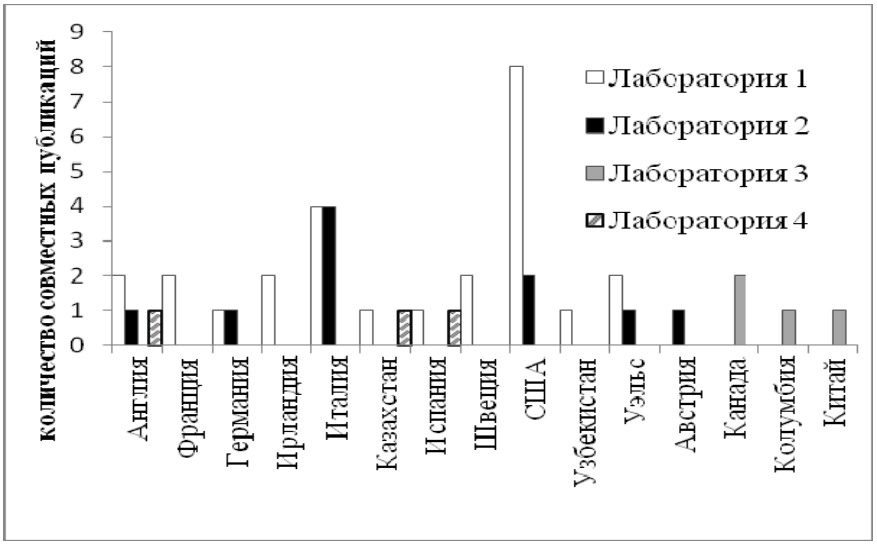

Рис. 7. Государства – партнеры, с которыми опубликованы совместные работы в 2007-2016 гг. по БД *WOS CC*

Таблица 5

Рейтинг научных направлений по частоте финансирования за 2007–2016 гг. (по БД *WOS CC*)

МЕСТО В РЕЙТИНГЕ	ОБЛАСТЬ НАУЧНОГО ЗНАНИЯ	ДОЛЯ ФИНАНСИРОВАНИЯ, %
1	Молекулярная биология	46
2	Биофизика	2
3	Цитология	2
4	Микробиология	8
5	Физиология	6
6	Биохимические методы исследования	4
7	Экология	4
8	Геология	4
9	Мультидисциплинарные науки	4
10	Нейрофизиология	4
11	Биология	2
12	Сердечные и сердечно-сосудистые системы	2
13	Аналитическая химия	2
14	Фармакология	2
15	Информатика	2
16	Кристаллография	2
17	Науки об окружающей среде	2
18	Физическая география	2
19	Иммунология	2
20	Математическая вычислительная биология	2
21	Ядерная технология	2
22	Медицинская радиология	2
23	Почвоведение	2

Таблица 6

Государства, ученые которых цитируют публикации сотрудников исследуемых подразделений

Государство	**Количество цитирований, %**
США	31,165
Россия	21,721
Япония	7,870
Китай	7,870
Германия	7,345
Великобритания	6,296
Италия	5,352
Канада	4,302
Испания	3,253
Франция	3,148
Польша	2,938
Бельгия	2,623
Ирландия	2,623
Австралия	1,784
Индия	1,784
Бразилия	1,574
Израиль	1,574
Швейцария	1,574
Нидерланды	1,469
Южная Корея	1,469
Швеция	1,469
Чили	1,259
Дания	1,259
Румыния	1,049

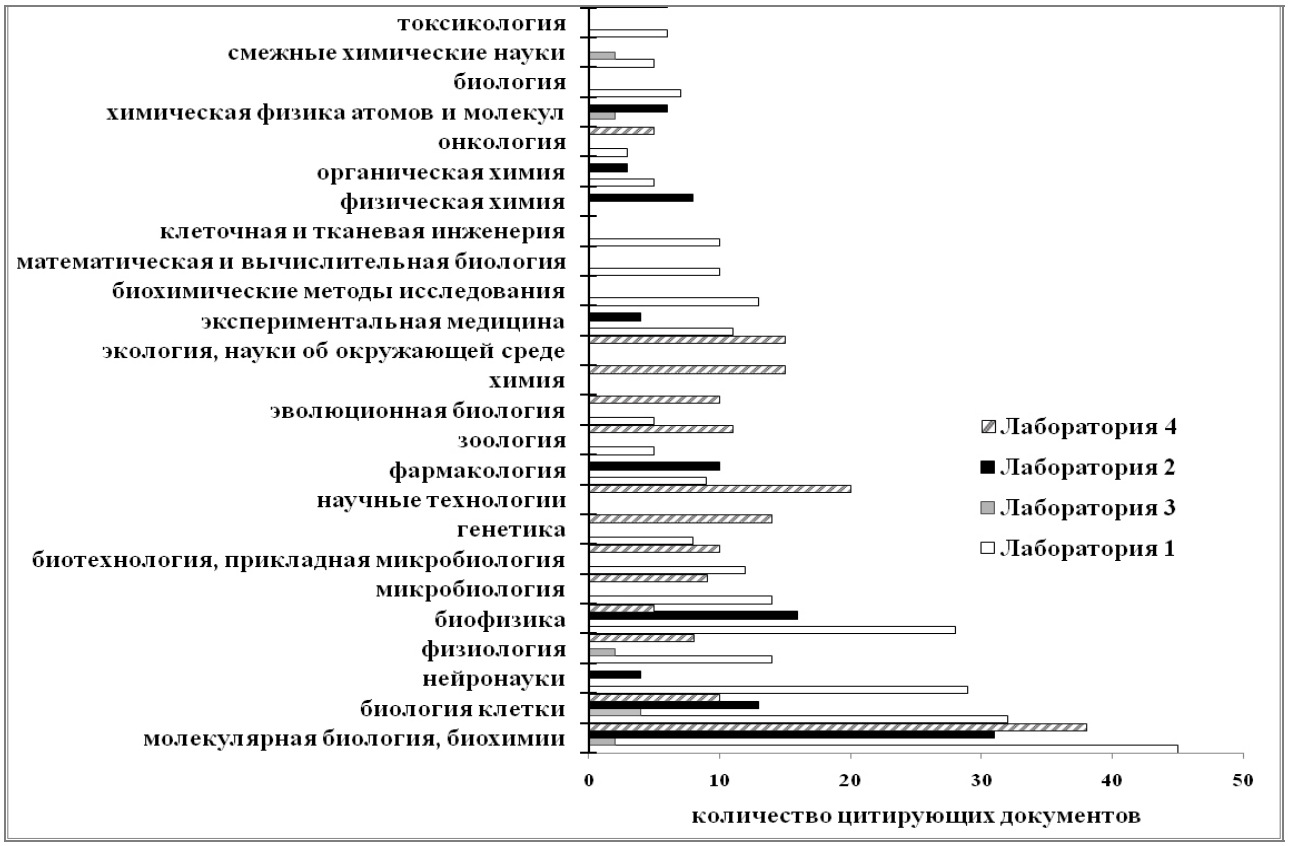

Рис. 8. Тематика публикаций, наиболее часто цитирующих работы исследуемых лабораторий

Одним из индикаторов востребованности научного направления в профессиональном сообществе являются ссылки на публикации автора. Недаром при ответах на вопросы анкеты многих ученых интересуют сведения о цитируемости своих работ: кто процитировал, из какой организации, какова тематика цитирующей публикации. Изучению взаимных связей, возникающих в процессе цитирования, посвящен еще один раздел нашего исследования. Анализ ссылок по *«Web of Science CC»* показал, что статьи рассматриваемых подразделений ИБК РАН за 2007–2016 гг. были процитированы в 953 публикациях учеными из 59 иностранных государств (табл. 6). На наш взгляд, такое разнообразие является результатом охвата большого количества тематических направлений, которыми занимаются ученые института.

Стоит отметить, что большое количество статей наши ученые опубликовали в иностранных или переводных журналах, индексируемых иностранными библиографическими базами, что повышает доступность их работ для мирового сообщества. Наибольшим разнообразием отличаются связи Лаборатория 1 и 2, ученые, цитирующие их работы, проживают в 57 и в 49 странах, соответственно**,** при безусловном лидерстве США- это 17,2% и 42,0% от общего числа цитирований. Цитирующие публикации других лабораторий отличаются меньшим территориальным разнообразием, и тут, кроме США, в первой пятерке публикации ученых из Италии, Китая, Германии, Японии и Франции.

Далее мы определили предметные области цитирующих публикаций и выяснили, что приоритетными направлениями для них являются: молекулярная биология, нейробиология, фармакология, физиология, микробиология (рис. 8).

Интересно, что такие научные категории, как фармакология, компьютерная биология, геронтология, токсикология, иммунология не входят в 20 распространенных тем исследуемых лабораторий, однако попадают в цитирующие публикации по этой теме. Кроме того, публикации сотрудников были процитированы в таких направлениях, как энергетика и топливо, прикладная физика, техническая химия, по своей тематике не связанных с профилем института.

Сравнение тематических направлений цитирующих публикаций, выявление закономерностей попадания научных статей в пристатейные списки, на наш взгляд, интересная и перспективная задача анализа цитирования. Возьмем, например, наиболее цитируемую публикацию: *Romanov RA., Rogachevskaja O.A., Bystrova M.F. et al. «Afferent neurotransmission mediated by hemichannels in mammalian taste cells.» // EMBO JOURNAL.– 2007. – Vol. 26.– Iss. 3. – P. 657–667. DOI: 10.1038/sj.emboj.7601526*, имеющую цитирование 218. Направление исследования данной работы – *Biochemistry & Molecular Biology; Cell Biology,* а вот цитируют ее в публикациях 39 различных научных категорий, включающих такие сферы, как диетология, бихевиоризм, пищевые технологии,

эндокринология, биоинженерия, медицинское приборостроение и т.д. Подобное исследование дает нам возможность определить тенденции развития коммуникационных процессов в науке, выявить междисциплинарные взаимоотношения ученых, установить связи, на основании которых в дальнейшем могут быть выделены кластеры работ, связанных по тематике.

Не менее интересно было узнать, читают и цитируют ли публикации наших авторов в странах ближнего зарубежья. К сожалению, из 561 цитирующей публикации всего по одной записи приходится на цитирование из Республики Беларусь, Латвии, Литвы, Украины.

ЗАКЛЮЧЕНИЕ

В ходе выполнения данного исследования была разработана методология для проведения анализа состояния развития научной и инновационной деятельности любого научного учреждения на примере лабораторий института ПНЦ РАН. Опираясь на суммарные показатели публикационной активности, цитируемости, международном сотрудничестве и т.п., мы можем сделать важные для оценки работы учреждения выводы, согласно которым наиболее стабильно публикует свои работы Лаборатории 1 и 3**,** а наименее устойчивое положение у Лаборатории 4, что в равной степени относится и к цитированию. Однако, лаборатория 4 является наиболее «молодым» из анализируемых подразделений, и такие факторы необходимо учитывать администрации при сравнении подразделений института с использованием библиометрического анализа. Стоит отметить, что большое количество статей наши ученые опубликовали в иностранных или переводных журналах, индексируемых иностранными библиографическими базами, что повышает видимость их работ зарубежными коллегами, а наибольшим разнообразием отличаются связи Лаборатория 1 и 2.

Особое внимание при анализе публикационной активности лабораторий мы уделили научным направлениям, получившим дополнительную финансовую поддержку от различных научных фондов. Российский Фонд Фундаментальных Исследований, Министерство Образования и Науки Российской Федерации, Российский Научный Фонд и наиболее часто осуществляли финансирование исследования по молекулярной биологии, микробиологии, физиологии, биохимии.

Изучению взаимных связей, возникающих в процессе цитирования, посвящен еще один раздел нашего исследования. Анализ ссылок по *«Web of Science CC»* показал, что статьи рассматриваемых подразделений за 10 лет, были процитированы в 953 публикациях учеными из 59 иностранных государств. По результатам анализа для исследуемых лабораторий были выявлены статьи, цитирование которых превышает отметку 100. Исходя из проведенного анализа, можно сделать вывод, что у сотрудников лабораторий ИБК РАН на протяжении длительного срока признанием пользуются такие издания, как Биологические мембраны, Биофизика, Бюллетень экспериментальной биологии и медицины, биохимия, молекулярная биология, Доклады академии наук.

Проведенный нами анализ публикационного потока лабораторий, с одной стороны, имеет значение для администрации учреждения, позволяя использовать наши данные о публикационной активности, цитируемости, финансировании, международных связях и пр. для предоставления отчетов в вышестоящие организации или учитывать их при распределении стимулирующих надбавок. С другой стороны, информация о присутствии статьи в определенном тематическом кластере, о наиболее финансируемых научных областях, о попадание публикации в разряд высоко- или быстроцитируемых, о присутствии издания в международных базах затребована непосредственно самими учеными, желающими ориентироваться в мировых научных течениях. Суммируя все вышесказанное, можно с уверенностью признать возможным использование наукометрического анализа публикаций с целью изучения закономерностей функционирования научного сообщества.

Учитывая, что научные библиотеки принимают на себя функции информационных центров, являясь посредниками между пользователями и информационными ресурсами, расширяют ассортимент услуг посредством обеспечения библиометрических, наукометрических, патентных поисков и их анализа, тем самым способствуют укреплению имиджа библиотеки, как важнейшего социального института.

СПИСОК ЛИТЕРАТУРЫ

1. Арчаков А.И., Карпова Е.А., Пономаренко Е.А. Международные критерии эффективности научно-исследовательской деятельности коллективов и отдельных ученых в области биологии и медицины // Вестник Российской академии медицинских наук. – 2013. – № 5. – С. 4–9.
2. Березкина Н.Ю., Хренова Г.С. Использование баз данных «Web of Science» для оценки результатов научной деятельности в республике Беларусь // Образовательные технологии и общество. – 2010. – Т. 13, № 3. – С. 311–316.
3. Вялков А.И., Глухова Е.А. Оценка качества научно-исследовательской деятельности медицинской организации с помощью наукометрических показателей // Здравоохранение Российской Федерации. – 2013. – № 3. – С. 3–5.
4. Гуськов А.Е. Реформа российской науки как импульс для развития наукометрических исследований (вступительная статья) // Труды ГПНТБ СО РАН. – 2015. – № 9. – С. 5–13.
5. Евдокимов В.И., Глухов В.А., Григорьев С.Г. Публикационная активность и наукометрические показатели статей в научных учреждениях по психиатрии и наркологии (2005-2014 гг.) // Вестник психотерапии. –2015. – № 56 (61). – С. 61–78.
6. Князева С.Ю., Слащева Н.А. Научно-техническое сотрудничество России и ЕС: библиометрический анализ // Форсайт. – 2008. – Т. 1, № 5. – С. 30-41.

7. Маркусова В.В., Иванов В.В., Варшавский А.Е. Библиометрические показатели российской науки и РАН (1997-2007) // Вестник РАН. – 2009. – Т. 79, № 6. – С.483–491.
8. Маршакова-Шайкевич И.В. Россия в мировой науке. Библиометрический анализ. – М.: Наука, 2008. – 227 с.
9. Мохначева Ю.В., Харыбина Т.Н. Научная продуктивность учреждений РАН и вузов: сравнительный библиометрический анализ // Вестник РАН. – 2011. – Т. 81, № 12. – С. 1065–1070.
10. Лаврик О.Л. Наукометрический анализ отечественного библиотековедения и библиографоведения // Библиосфера. – 2010. – № 2. – С. 51–59.
11. Гиляревский Р.С., Сюнтюренко О.В. Использование методов наукометрии и сопоставительного анализа данных для управления научными исследованиями по тематическим направлениям // Научно-техническая информация. Сер. 2. – 2016. – № 12. – С. 3–10.
12. Panat R. On the data and analysis of the research output of India and China: India has significantly fallen behind China // Scientometrics. – 2014. – № 2. – P. 471–481.
13. Gumpenberger C., Gorraiz J. Going beyond Citations: SERUM — a new Tool Provided by a Network of Libraries // LIBER Quarterly. – 2010. – Vol 20, №1. – P. 80–93. DOI: http://doi.org/10.18352/lq.7978.
14. Library Research Support in Queensland : A Survey / Joanna Richardson, Therese Nolan-Brown, Pat Loria, Stephanie Bradbury // Australian Academic & Research Libraries. – 2012. – Vol. 43, Iss. 4. – P. 258–277. DOI: 10.1080/00048623.2012.10722287.
15. Gumpenberger C., Wieland M., Gorraiz J. Bibliometric practices and activities at the University of Venna// Libr. Manag. – 2012. – Vol. 33, № 3. – P. 174–183.
16. Бескаравайная Е.В., Харыбина Т.Н. Динамика библиометрических показателей сотрудников научных школ Института белка РАН // Информационное обеспечение науки: новые технологии: сб. науч. тр. / ред. Н.Е. Каленов, В.А. Цветкова. – М.: БЕН РАН, 2015. – С. 63–73.
17. Мохначева Ю.В., Харыбина Т.Н. Сравнительная оценка научной продуктивности исследовательских учреждений РАН и сектора российской высшей школы по некоторым библиометрическим индикаторам (2000 – 2009 гг.) // Библиосфера. – 2011. – №3. – С. 57–64.
18. Бескаравайная Е.В., Харыбина Т.Н. Наукометрический анализ членов диссертационного совета одного из НИИ Пущинского научного центра РАН Науковедческие исследования // Сб. трудов ИНИОН РАН / отв. ред. А.Н. Ракитов. – М., 2016. – С. 74–90.

Материал поступил в редакцию 15.03.18.

Сведения об авторах

БЕСКАРАВАЙНАЯ Елена Вячеславовна – старший научный сотрудник Библиотеки по естественным наукам РАН, Москва
e-mail: elenabesk@gmail.com

ХАРЫБИНА Татьяна Николаевна – заслуженный работник культуры РФ, старший научный сотрудник Библиотеки по естественным наукам РАН, Москва
e-mail: natsl@vega.protres.ru

УДК 004.422.63.021:004.85

Д.Ю. Съедин

Разработка и реализация алгоритма связывания данных в государственной информационной системе гражданского назначения

Описана разработка алгоритма связывания данных. В основе алгоритма лежит вариант адаптивной меры подобия записей (АМПЗ), оптимизация которого для решения задачи связывания данных проводится при помощи алгоритма машинного обучения. Подход к оценке адекватности АМПЗ при связывании данных состоит во введении и автоматизированном подборе алгоритмом машинного обучения значений пороговых величин, что достигается введением функции приспособленности, представляющей модификацию $F_{measure}$.

Ключевые слова: *обработка информации, связывание данных, машинное обучение, распределенные вычисления*

ВВЕДЕНИЕ

Для реализации концептуальной модели Единой государственной информационной системы учета научно-исследовательских, опытно-конструкторских и технологических работ гражданского назначения (ЕГИСУ НИОКТР), позволяющей интегрировать информацию обо всех результатах научно-технической деятельности (РНТД) в стране, а также для обеспечения непротиворечивости данных, доступных для эффективного мониторинга, в работе [1] ставилась задача проектирования и реализации соответствующего информационного хранилища. Помимо реализации хранилища непосредственно, т. е. проектирования структур данных, обеспечивающих функционирование модели, необходимо было обеспечить эффективную координацию архивных данных, являющихся ретроспективной информацией о РНТД, накопленных в разрозненных информационных системах. Основная сложность такой задачи была обусловлена обилием текстовых данных, не связанных с компонентами нормативно-справочной информации (НСИ). Очевидно, что для решения этой проблемы необходимы инструменты, обеспечивающие нечеткое сопоставление данных с компонентами НСИ для дальнейшей координации. Настоящая работа посвящена реализации такого инструмента, а именно алгоритму связывания данных как одному из средств решения одноименной задачи.

ПОСТАНОВКА ЗАДАЧИ

Задача связывания данных может быть определена как задача сопоставления пар эквивалентных записей, различающихся синтаксически [2]. Существуют различные подходы к ее решению. Основы для решения этой задачи одними из первых заложили авторы в работе [3], предложив формальную математическую модель связывания данных. В дальнейшем в ряде публикаций исходная теория развивалась и совершенствовалась. Широкое распространение получили так называемые меры подобия (*similarity measures*) между двумя последовательностями символов, позволяющие путем осуществления определенных преобразований количественно определять сходство между последовательностями. Позднее на их основе были разработаны подходы, призванные повысить точность решения задачи сопоставления пар записей. Например, в работе [4] авторами подразумевается разбиение записей на отдельные слова. Это делается с целью определения среднего от мер подобия между наиболее похожими словами записей как меры подобия последних. В работе [5] при расчете мер подобия записей также указывается на необходимость учета статистических параметров слов в записях. Для повышения эффективности решения задачи авторы предлагают производить первоначальную предобработку источников данных: удалять стоп-слова, производить смысловое разделение сопоставляемых записей (если это возможно) на меньшие блоки данных, содержащие общие характеристики, использовать семантические ограничения [6–8] и т. д.

Более поздние работы, например [9], посвящены использованию алгоритмов машинного обучения для повышения эффективности решения задачи связывания данных. Для оценки качества связывания данных в большинстве работ используется так называемая F-мера ($F_{measure}$), являющаяся гармоническим средним между точностью (*Precision*) и полнотой (*Recall*).

В настоящей работе представлен алгоритм связывания данных как одно из средств решения задачи связывания данных. В основе этого алгоритма лежит вариант адаптивной меры подобия записей (АМПЗ), оптимизация которого для решения задачи связывания данных осуществляется алгоритмом машинного обучения, описанным в работе [10]. Этот алгоритм продемонстрировал свою эффективность при решении задачи поиска глобального экстремума на различных функциях многих переменных, а также имеет параллельную реализацию, что позволяет эффективно использовать имеющиеся вычислительные ресурсы. Подробное описание, тестирование и проверка эффективности параллельной реализации алгоритма оптимизации представлены в работе [11].

РЕАЛИЗАЦИЯ АЛГОРИТМА СВЯЗЫВАНИЯ ДАННЫХ

В настоящей работе подход к реализации алгоритма связывания данных основан на следующих соображениях. Так, для определения схожести записей при координации данных в информационных хранилищах необходимо опираться на различные поля, с целью обеспечения максимальной корректности сопоставления. Например, для сопоставления физических лиц помимо поля ФИО следует использовать дополнительную информацию о субъектах сопоставления. С другой стороны, как было отмечено в работе [12], данные каждого из полей обладают своими синтаксическими особенностями и это необходимо учитывать при выборе мер подобия записей. Кроме того, для повышения вклада «важных» слов и уменьшения вклада общеупотребительных необходимо предусмотреть возможность учета статистической информации.

В работе в качестве мер подобия при реализации АМПЗ предлагается использовать наиболее эффективные и широко употребимые меры: мера, основанная на расстоянии Левенштейна, мера, основанная на сходстве Джаро-Винклера, а также мера, основанная на коэффициенте сходства Джаккарда (далее мера Левенштейна, мера Джаро-Винклера, мера Джаккарда).

Тогда, имея в виду [5, 13], представим меру подобия пары слов одного поля пары записей:

$$sim\left(word_{i_1jk_1}, word_{i_2jk_2}\right) = \begin{pmatrix} w_{Lev\,j} sim_{Lev}\left(word_{i_1jk_1}, word_{i_2jk_2}\right) + \\ w_{JW\,j} sim_{JW}\left(word_{i_1jk_1}, word_{i_2jk_2}\right) + \\ w_{Jaccard\,j} sim_{Jaccard}\left(word_{i_1jk_1}, word_{i_2jk_2}\right) \end{pmatrix} \cdot \log\left(\alpha + \frac{\beta}{f\left(word_{i_1jk_1}\right)}\right) \cdot \log\left(\alpha + \frac{\beta}{f\left(word_{i_2jk_2}\right)}\right), \quad (1)$$

где sim_{Lev} – мера Левенштейна, $w_{Lev\,j}$ – соответствующий ей весовой коэффициент для поля j;

sim_{JW} – мера Джаро – Винклера, $w_{JW\,j}$ – соответствующий ей весовой коэффициент для поля j;

$sim_{Jaccard}$ – мера Джаккарда, $w_{Jaccard\,j}$ – соответствующий ей весовой коэффициент для поля j;

$f\left(word_{i_1jk_1}\right)$ – частота слова $word_{k1}$ поля j записи i_1 в общем массиве данных;

α, β – масштабирующие коэффициенты, необходимые для коррекции вклада статистической информации при расчете АМПЗ.

Тогда мера подобия одного поля пары записей имеет вид:

$$sim\left(field_{i_1j}, field_{i_2j}\right) = \frac{1}{2n_{i_1j}} \sum_{k_1=1}^{n_{i_1j}} \max_{k_2} sim\left(word_{i_1jk_1}, word_{i_2jk_2}\right) + \frac{1}{2n_{i_2j}} \sum_{k_2=1}^{n_{i_2j}} \max_{k_1} sim\left(word_{i_1jk_1}, word_{i_2jk_2}\right) \quad (2)$$

Поскольку, как было сказано ранее, записи могут состоять из нескольких полей, получим формулу АМПЗ:

$$sim\left(record_{i_1}, record_{i_2}\right) = \frac{\sum_{j=1}^{F} \partial_j sim\left(field_{i_1j}, field_{i_2j}\right)}{\sum_{j=1}^{F} \partial_j} \quad (3)$$

где ∂_j – весового коэффициента поля j.

Как видно, формула (3) содержит множество весовых коэффициентов. Задача этих коэффициентов – усиливать (ослаблять) вклады каждой из мер подобия, входящих в АМПЗ, в зависимости от конкретных полей имеющегося набора данных, а также увеличивать (уменьшать) важность того или иного поля набора данных при принятии решения об эквивалентности (различии) пар записей.

Задача, которую необходимо решить – оптимизация величин указанных весовых коэффициентов для получения адекватной АМПЗ на имеющемся наборе данных. Кроме того, предлагается автоматизировать подбор пороговых значений для оценки адекватности указанной меры. Для этого введем два параметра оценки адекватности АМПЗ – θ_{low} и θ_{high}. Их смысл заключается в следующем. При сопоставлении пар записей мера вычисляет значение, которое необходимо сравнить со значениями этих параметров. Таким образом, для того, чтобы можно было сделать вывод об эквивалентности (различии) пары записей, необходимо вычислить значение АМПЗ и далее сравнить его с параметрами θ_{low} и θ_{high}. При значении АМПЗ, превышающем величину θ_{high}, делается вывод о том, что записи эквивалентны. Если значение меньше величины θ_{low}, делается вывод о том, что записи различны. Если значение меры попадает в промежуток между величинами параметров – никакие выводы не делаются. Такой подход обусловлен желанием выделить интервал значений, попадая в

который при сопоставлении пары записей АМПЗ не может адекватно оценить их эквивалентность (различие). В этой ситуации решение предлагается принимать Пользователю.

Таким образом

$$sim\left(record_{i_1}, record_{i_2}\right) = \begin{cases} > \theta_{high} - \text{записи эквивалентны} \\ < \theta_{low} - \text{записи уникальны} \\ \geq \theta_{low}\ u \leq \theta_{high} - \text{невозможно определить} \end{cases} \quad (4)$$

Корректную настройку параметров θ_{low} и θ_{high} в процессе оптимизации АМПЗ, как было сказано выше, также предлагается автоматизировать, а именно производить их настройку при помощи алгоритма оптимизации, совместно с весовыми коэффициентами формулы (3).

Для обеспечения работы алгоритма оптимизации по обучению АМПЗ необходима функция приспособленности (фитнесс-функция), с помощью которой можно было бы оценивать корректность оптимизации АМПЗ и параметров ее оценки. Эта функция для обеспечения условия (4) должна, с одной стороны, учитывать корректность определения пар эквивалентных записей при обучении меры, с другой — учитывать корректность определения пар уникальных записей. Кроме того, существенный интерес вызывает обеспечение возможности влиять на процесс обучения АМПЗ параметрами указанной функции, определяемыми Пользователем до начала обучения. Эти параметры должны указывать насколько АМПЗ «самостоятельна» в принятии решений и, как следствие, насколько широким должен быть интервал между параметрами оценки адекватности АМПЗ – θ_{low} и θ_{high}.

Для этого предлагается модифицировать $F_{measure}$. Модифицированная $F_{measure'}$ или $F^{\sim}_{measure'}$ в таком случае будет представлять собой взвешенное гармоническое среднее четырех метрик – полноты определения пар эквивалентных записей, полноты определения пар уникальных записей, точности определения пар эквивалентных записей и точности определения пар уникальных записей, т.е., если

$$Recall_{duplicates} = \frac{N_{correctDuplicatesPredictions}}{N_{duplicates}},$$

$$Recall_{uniques} = \frac{N_{correctUniquesPredictions}}{N_{uniques}},$$

$$Precision_{duplicates} = \frac{N_{correctDuplicatesPredictions}}{N_{duplicatesPredictions}},$$

$$Precision_{uniques} = \frac{N_{correctUniquesPredictions}}{N_{uniquesPredictions}},$$

где $N_{duplicates}$, $N_{uniques}$ – количество пар эквивалентных и количество пар уникальных записей согласно данным из обучающей выборки;

$N_{duplicatesPredictions}$, $N_{uniquesPredictions}$ – количество пар записей, определенных АМПЗ как эквивалентные, и количество пар записей, определенных АМПЗ как уникальные;

$N_{correctDuplicatesPredictions}$, $N_{correctUniquesPredictions}$ – количество пар записей, верно распознанных как эквивалентные, среди всех пар записей, определенных АМПЗ как эквивалентные, и количество пар записей, верно распознанных как уникальные, среди всех пар записей, определенных АМПЗ как уникальные.

Тогда

$$F^{\sim}_{measure} = \frac{2\left(w_{recall} + w_{precision}\right)}{w_{recall}\left(Recall_{duplicates}^{-1} + Recall_{uniques}^{-1}\right) + w_{precision}\left(Precision_{duplicates}^{-1} + Precision_{uniques}^{-1}\right)} \quad (5)$$

Здесь w_{recall} – параметр, отвечающий за «вклад» полнот в общую величину $F^{\sim}_{measure}$, $w_{precision}$, аналогично, отвечает за «вклад» точностей.

На основании вышесказанного, в процессе обучения на каждой итерации алгоритма оптимизации АМПЗ оцениваются все возможные пары записей из заранее сформированной обучающей выборки. Основываясь на имеющейся информации относительно того, какие в обучающей выборке пары записей эквивалентные, а какие различные и на результатах производимого мерой оценивания высчитывается значение $F^{\sim}_{measure}$. Задача алгоритма оптимизации – максимизировать $F^{\sim}_{measure}$ путем варьирования весовых коэффициентов и параметров оценки АМПЗ. При повышении адекватности меры в процессе ее обучения и, как следствие, увеличении величины $F^{\sim}_{measure}$, оценивающей эту адекватность, значения параметров θ_{low} и θ_{high} будут приближаться друг к другу, сокращая интервал значений, характеризующий невозможность автоматического принятия решения об эквивалентности (различии) пар записей.

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ АЛГОРИТМА СВЯЗЫВАНИЯ ДАННЫХ

Блок-схему обучения адаптивной меры подобия записей, лежащую в основе алгоритма связывания данных, покажем на рис. 1. Программная реализация этой схемы строится следующим образом. В качестве алгоритма обучения предлагается использовать представленный в работах [10, 11] алгоритм, который имеет параллельную реализацию, основанную на фреймворке распределенных вычислений *Apache Spark*, для функционирования в распределенной вычислительной среде. Поскольку каждая глобальная итерация алгоритма обучения подразумевает много-

кратное вычисление функции приспособленности, а для каждого вычисления последней, в свою очередь, необходимо иметь значения всех мер подобия записей, используемых в АМПЗ от каждой из пар записей, входящих в обучающую выборку, требуется обеспечить оперативность функционирования алгоритма. Очевидно, необходимо заранее (до начала непосредственного выполнения обучения) произвести вычисление мер подобия, используемых АМПЗ, представленной выше, для каждой из пар записей обучающей выборки, обойдя в дальнейшем необходимость их вычисления. Необходимо иметь в виду, что коллекция значений от вычислений мер подобия может быть велика (меры подобия вычисляются от всех возможных пар слов каждого поля каждой пары записей). В этой связи для корректной работы алгоритма нужно предусмотреть организацию доступа вычислительных узлов кластера к коллекции для возможности вычисления функции приспособленности. В настоящей работе в качестве решения этой проблемы предлагается использовать *lookup*-таблицу в базе данных, в которой будет содержаться вся коллекция указанных значений. В процессе обучения на каждой глобальной итерации алгоритма узлы кластера будут обращаться к этой таблице с целью получения необходимой информации из базы данных. В таком случае временные затраты на расчет АМПЗ и, как следствие, функции приспособленности значительно сократятся. Кроме того, такой подход исключает необходимость постоянной передачи больших объемов данных от менеджера кластера к узлам и связанные с этим накладные расходы, например при использовании так называемых *broadcast*-переменных.

Таким образом, в соответствии с алгоритмом оптимизации [11] инициализируется исходная популяция поисковых агентов (частиц), после чего происходит ее разделение на ряд субпопуляций. Каждый вычислительный узел получает данные обо всех величинах мер подобия, используемых АМПЗ, путем обращения к *lookup*-таблице. После чего узлы, используя полученные данные и оперируя «своими» субпопуляциями, производят решение задачи оптимизации указанным алгоритмом, т. е. задачи по оптимизации АМПЗ и параметров оценки ее адекватности. После определенного числа итераций, выполненных на каждой из субпопуляций, происходит сбор последних программой-драйвером с каждого из узлов с последующим слиянием собранных субпопуляций в общую популяцию. По завершении обмена данными в виде переразбиения общей популяции другим способом разбиения на субпопуляции, начинается следующая глобальная итерация алгоритма оптимизации. В итоге, по истечении заданного числа глобальных итераций полученное наилучшее с точки зрения величины функции приспособленности решение представляет собой набор весовых коэффициентов обученной АМПЗ и параметров оценки ее адекватности.

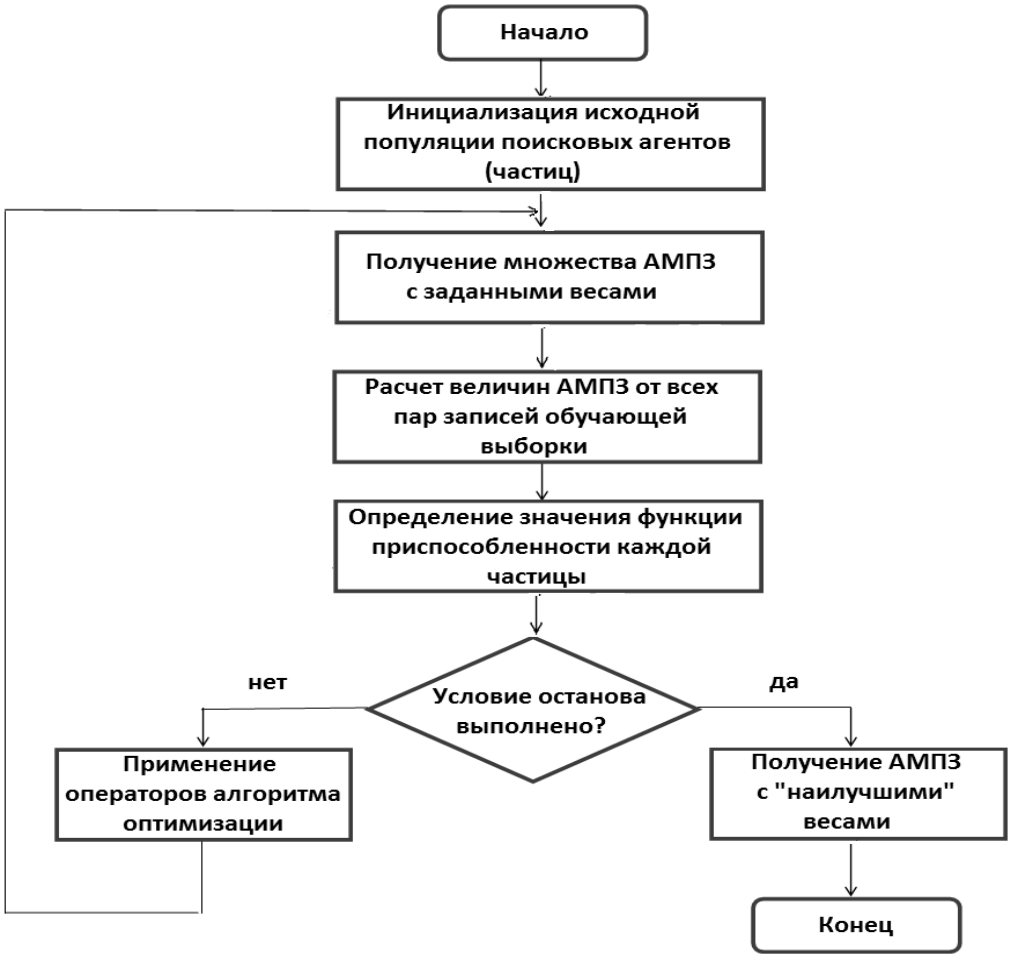

Рис. 1. Блок-схема обучения адаптивной меры подобия записей, лежащей в основе алгоритма связывания данных

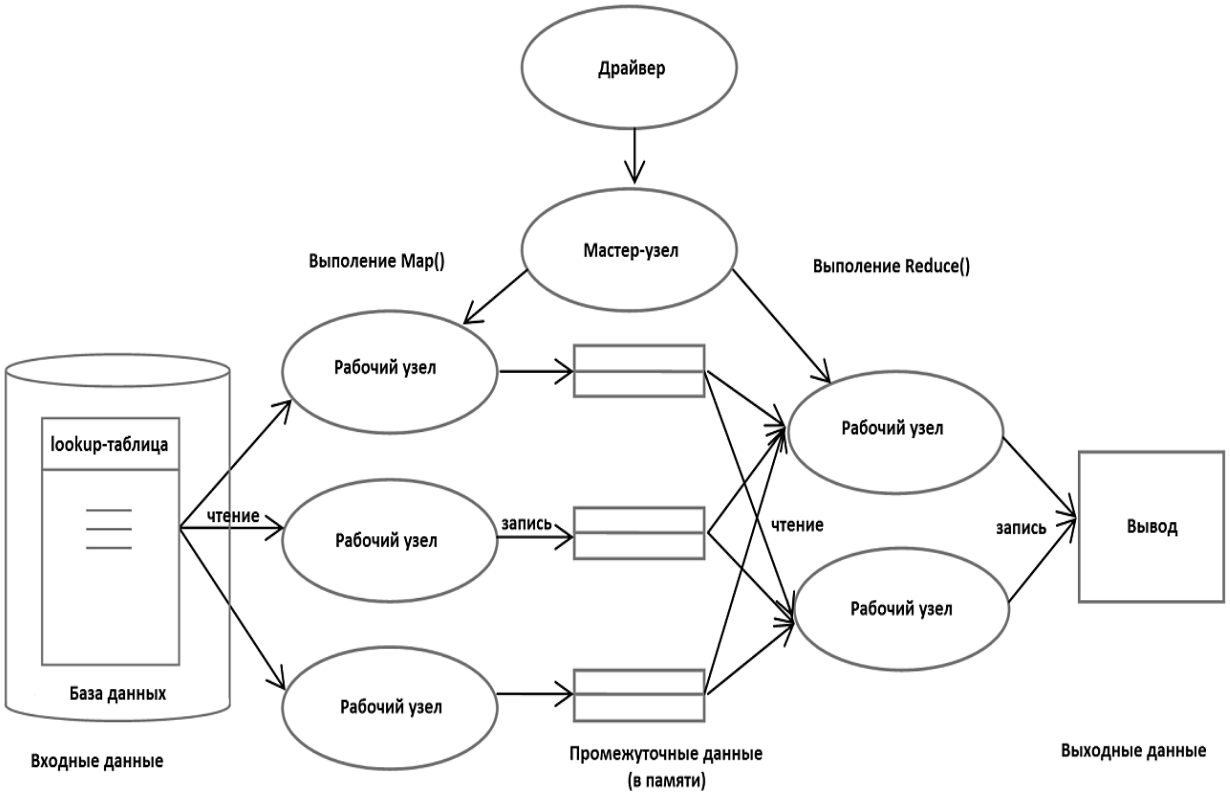

Рис. 2. Схематическое представление глобальной итерации процесса обучения адаптивной меры подобия записей с помощью фреймворка *Apache Spark*

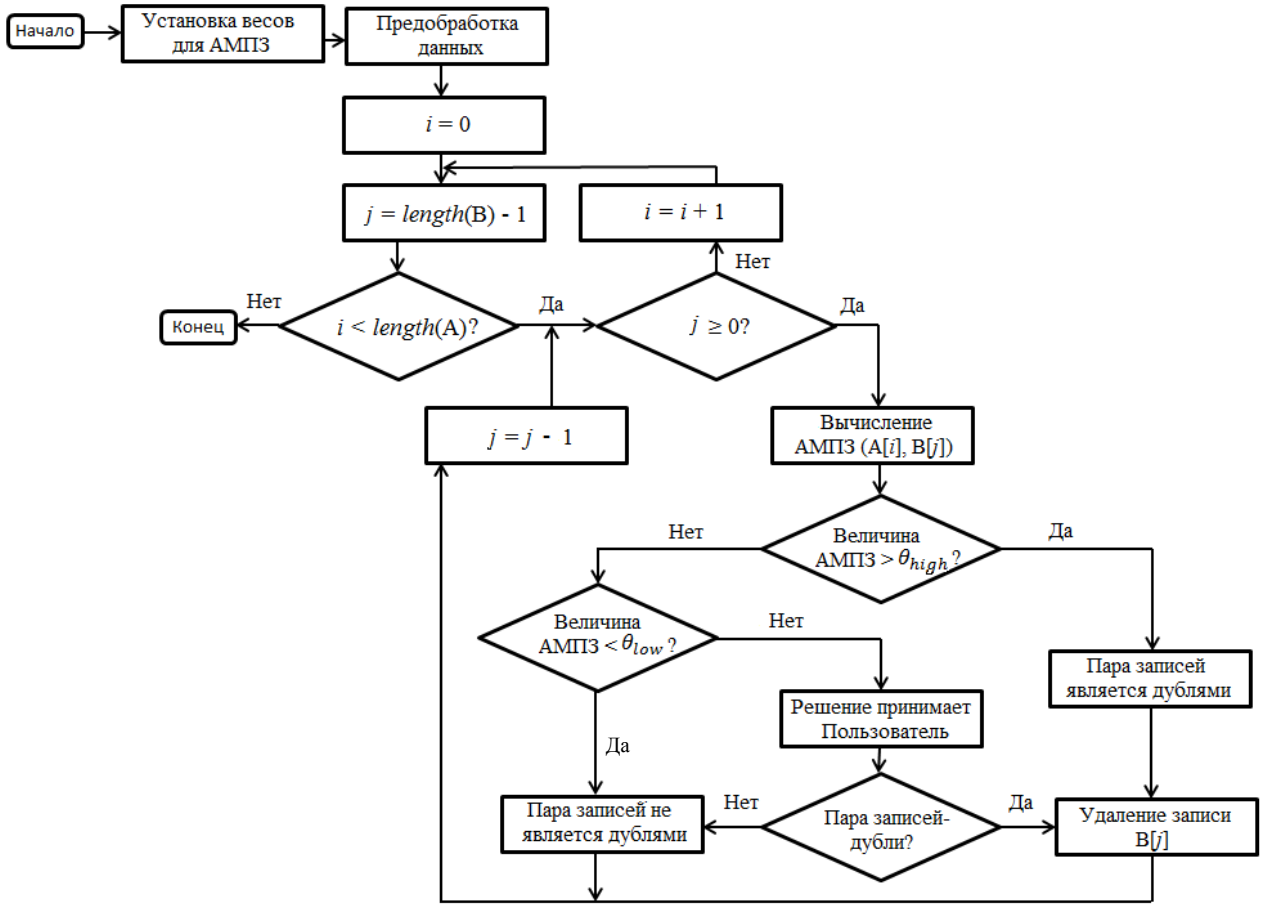

Рис. 3. Блок-схема функционирования алгоритма связывания данных

Схематическое представление одной глобальной итерации процесса обучения АМПЗ с помощью фреймворка *Apache Spark*, основанного на вышесказанных соображениях, представлено на рис. 2.

Как было отмечено выше, результатом обучения АМПЗ является набор весов, необходимых для корректной работы меры на имеющемся источнике данных. Для решения задачи связывания данных полученные веса подставляются в представленную выше АМПЗ.

Блок-схему функционирования алгоритма связывания данных, использующего АМПЗ для сравнения пар записей, покажем на рис. 3. Здесь A и B массивы, содержащие наборы записей для дальнейшего связывания.

ПРОВЕРКА ЭФФЕКТИВНОСТИ АЛГОРИТМА СВЯЗЫВАНИЯ ДАННЫХ

Представленный алгоритм связывания данных использовался при координации архивных данных при разработке и реализации Единой государственной информационной системы учета научно-исследовательских, опытно-конструкторских и технологических работ. Приведем пример такого использования.

ЕГИСУ НИОКТР содержит в своем хранилище информацию, в частности, о защитах открытых диссертационных работ. С момента запуска системы в эксплуатацию (январь 2014 г.) ввод данных о таких работах осуществляется через формы системы. Поля и компоненты нормативно-справочной информации этих форм были утверждены Приказом Министерства образования и науки РФ[1]. Одним из таких справочников стал справочник организаций, основанный на базе Единого государственного реестра юридических лиц (ЕГРЮЛ). Так, например, данные об организациях, в советах которых проходили защиты, заполняются путем выбора соответствующих организаций через выпадающий список этого справочника. Данные о диссертационных работах, зарегистрированных до января 2014 г., при переносе из архива информационной системы, предшествующей системе ЕГИСУ НИОКТР — Единой федеральной базы данных НИОКР (ЕФБД НИОКР), также надо было сопоставлять со справочником организаций ЕГРЮЛ. Основной проблемой при подобном сопоставлении и дальнейшей координации данных было то, что в ЕФБД НИОКР информация о подобных организациях хранилась в простых текстовых полях. Несмотря на то, что данные при регистрации работ тщательно проверялись службой эксплуатации ЕФБД НИОКР, работал корректорский отдел, выпускались информационные издания о выполненных работах, задача сопоставления данных со справочником организаций являлась нетривиальной. Для ее выполнения было принято решение использовать полученный алгоритм связывания данных.

Связывание данных производилось для диссертационных работ, зарегистрированных с января 2007 г. Именно с этого момента информация о результатах научной деятельности в России доступна на информационном портале системы в «горячем доступе». Таких работ оказалось 152 000 единиц. Задача связывания решалась следующим образом. Первоначально была произведена попытка связывания организаций по полному совпадению цифровых кодов, указанных в текстовых полях диссертационных работ, с цифровыми полями организаций из справочника ЕГРЮЛ. На этом этапе удалось сопоставить порядка 100 000 работ с 1486 организациями из справочника ЕГРЮЛ. Оставшуюся треть документов связать со справочником таким образом не удалось – имели место неточности или отсутствие данных в полях цифровых кодов записей о диссертационных работах.

Поскольку справочник ЕГРЮЛ (на момент его внедрения в систему) содержал информацию более чем о 8,5 млн организаций, сопоставление организаций из оставшихся 52398 диссертационных работ с организациями этого справочника явилось бы чрезвычайно трудоемкой в вычислительном плане задачей. Было принято решение действовать из предположения, что оставшиеся работы в большинстве своем с высокой долей вероятности имеют те же места защиты, что и успешно связанные организации, сопоставленные путем полного совпадения. Это можно обосновать тем, что общее число мест защит диссертационных работ в стране относительно невелико по сравнению с общим числом всевозможных организаций из справочника ЕГРЮЛ, что, вероятно, позволяет достаточно эффективно решить поставленную задачу.

Полями, которые имели непустые значения, характеризующие организации, в которых проходила защита диссертационных работ, было Полное наименование и Юридический адрес. Эти поля послужили основой для решения конкретной задачи сопоставления.

На первоначальном этапе производилась коррекция данных. В словарь стоп-слов были добавлены данные о наименованиях организационно-правовых форм с целью уменьшения количества общеупотребительных в данном контексте слов, не несущих существенной смысловой нагрузки. Таким образом, например, организационно-правовые формы “Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования”, “Федеральное государственное автономное научное учреждение” и так далее, а также их сокращения “ФГБОУ ВПО”, “ФГАНУ” и прочие не учитывались. Все слова были автоматически приведены к одному регистру.

Далее была подготовлена обучающая выборка. В нее были добавлены данные о некоторых организациях из справочника ЕГРЮЛ и соответствующие им данные из записей о диссертационных работах. Выборка содержала признаки, указывающие на то,

[1] Приказ Министерства образования и науки РФ от 31.03.2016 N 341 «Об утверждении форм направления сведений о научно-исследовательских, опытно-конструкторских и технологических работах гражданского назначения в целях их учета в единой государственной информационной системе учета научно-исследовательских, опытно-конструкторских и технологических работ гражданского назначения, требований к заполнению указанных форм, порядка подтверждения главными распорядителями бюджетных средств, осуществляющими финансовое обеспечение научно-исследовательских, опытно-конструкторских и технологических работ гражданского назначения, условиям государственных контрактов на выполнение научно-исследовательских, опытно-конструкторских и технологических работ гражданского назначения».

какие организации из справочника и из диссертационных работ эквиваленты между собой. Объем выборки составил порядка 100 записей.

Приведем ее фрагмент:

"Пермская государственная сельскохозяйственная академия"DELIMITER614990, г. Пермь, Коммунистическая, 23DELIMITERrecord1
Пермская государственная сельскохозяйственная академия имени академика Д.Н. Прянишникова DELIMITER614990, Пермь, ул. Коммунистическая, д. 23DELIMITERrecord1
Сибирского отделения РАН Институт теплофизикиDELIMITER630090, г. Новосибирск, просп. Академика Лаврентьева, 1DELIMITERrecord2
Институт теплофизики Сибирского отделения РАНDELIMITER Новосибирск, просп. Академика Лаврентьева, д1, 630090DELIMITERrecord2
...

Первоначально, перед обучением АМПЗ на указанной обучающей выборке для параметров функции приспособленности $F^{\sim}_{measure} - w_{recall}$ и $w_{precision}$ были заданы значения равные 1. Это было сделано для тестирования АМПЗ в условиях, при которых величина $F^{\sim}_{measure}$ в равной степени зависит от величин точностей и полнот в формуле (5). Сравнение достигнутого в результате обучения значения $F^{\sim}_{measure}$ АМПЗ с величинами значений $F^{\sim}_{measure}$, полученными на этой же выборке с помощью других подходов, приведено в табл. 1.

Таблица 1

Величины $F^{\sim}_{measure}$, достигнутые различными мерами подобия записей на обучающей выборке

Мера подобия записей	$F^{\sim}_{measure}$
Левенштейна	0.857
Мондж-Элкан [4] с мерой Джаро-Винклера	0.951
АМПЗ	0.981

Как видно, наибольшее значение величины $F^{\sim}_{measure}$ было достигнуто АМПЗ.

Для проверки адекватности полученной АМПЗ производилось ее тестирование на выборке размером в 1000 записей, отсутствовавших в обучающей выборке. Результирующее значение $F^{\sim}_{measure}$, достигнутое на тестовой выборке, представлено в табл. 2.

Как видно, значение $F^{\sim}_{measure}$, достигнутое на тестовой выборке, лишь немногим уступает величине $F^{\sim}_{measure}$, полученной в результате обучения АМПЗ на обучающей выборке, что говорит об адекватности предложенного варианта меры для ее использования в алгоритме связывания данных.

Отметим, что при таком подходе (когда w_{recall} и $w_{precision}$ равны 1) полученные в результате обучения величины параметров оценки адекватности АМПЗ – θ_{low} и θ_{high} оказались довольно близки – вероятность привлечения экспертного мнения Пользователя невелика. В этой связи, с целью повышения практичности подхода и уменьшения числа ошибочных предсказаний при связывании пар записей, для параметров функции приспособленности $F^{\sim}_{measure}$, — w_{recall} и $w_{precision}$ перед обучением АМПЗ было решено задать значения 0.3 и 0.7, соответственно. Это позволило обеспечить большее влияние точностей на величину $F^{\sim}_{measure}$, чем полнот в формуле (5), что уменьшает «самостоятельность» АМПЗ в принятии решений по связыванию пар записей и увеличивает вероятность привлечения экспертного мнения Пользователя.

Таблица 2

Величина $F^{\sim}_{measure}$, полученная при тестировании АМПЗ на тестовой выборке

Мера подобия записей	$F^{\sim}_{measure}$
АМПЗ	0.905

Непосредственно сам процесс связывания записей при решении задачи координации данных об организациях, в советах которых производились защиты диссертационных работ, осуществлялся согласно блок-схеме алгоритма связывания данных, представленной на рис. 3. Проводилось сопоставление 1486 записей организаций справочника ЕГРЮЛ с 52398 записями об организациях, в советах которых производились защиты диссертационных работ. В результате удалось привязать 37515 записей из работ к имеющимся 1486 записям об организациях из реестра ЕГРЮЛ (~72%). При этом 89% записей из привязанных записей оказались верно распознаны, как эквивалентные записям из списка ЕГРЮЛ. Оставшиеся 11% либо были помечены как нераспознанные, либо привязаны неверно. Анализ показал, что наиболее частое неверное связывание пар записей происходило по причине слишком большой синтаксической схожести записей, по факту являющихся различными. Скажем, в паре

НИВЦ Московского государственного университета имени М.В. ЛомоносоваDELIMITER119991, Москва, Ленинские горы, 1/4

Московский государственный университет имени М.В. ЛомоносоваDELIMITER119991, Москва, Ленинские горы, 1

записи были ошибочно помечены как дублирующиеся, поскольку и наименования, и адреса записей практически полностью совпадают, хотя, при этом, согласно версии справочника ЕГРЮЛ, используемой на момент эксперимента, формально это разные организации (имеют различные коды ОГРН).

Примером ситуации привлечения экспертного мнения Пользователя (невозможности определить различие или эквивалентность) является следующая пара записей:

Институт теплофизики имени Кутателадзе Сибирского отделения РАНDELIMITER630090, Новосибирск, просп. Академика Лаврентьева, 1

Институт лазерной физики Сибирского отделения РАНDELIMITER630090, г. Новосибирск, просп. Академика Лаврентьева, 13/3.

Как видно, в этом случае, несмотря на большие синтаксические отличия в наименованиях организаций, адреса последних практически полностью совпадают. Здесь АМПЗ алгоритма связывания данных не смогла однозначно построить решающее правило для указанных данных.

Проверка непривязанных записей об организациях, в диссертационных советах которых происходили защиты диссертаций, к указанным ранее записям об организациях из справочника ЕГРЮЛ показала, что организации сравниваемых записей действительно отличались и были верно распознаны как различные. Исключение составили немногочисленные случаи, когда организации на момент попадания в справочник ЕГРЮЛ кардинально меняли свое название и юридический адрес. В этих ситуациях АМПЗ ошибочно пометила данные как различные.

ЗАКЛЮЧЕНИЕ

Представленный в настоящей работе алгоритм является одним из средств решения задачи связывания данных. В качестве этого алгоритма предложен вариант адаптивной меры подобия записей (АМПЗ), оптимизация которого для решения задачи связывания данных была осуществлена с помощью алгоритма машинного обучения, описанного в работах [10, 11]. Кроме того, предложен подход по оценке адекватности АМПЗ при связывании данных за счет введения и автоматизированного подбора значений пороговых величин θ_{low} и θ_{high}, что достигается введенной в работе функцией приспособленности, представляющей собой модификацию $F^{\sim}_{measure}$.

Предложенный алгоритм использовался при решении задачи связывания данных для координации данных об организациях, в диссертационных советах которых происходили защиты диссертаций, с данными об организациях из справочника Единого государственного реестра юридических лиц в Единой государственной информационной системе учета научно-исследовательских, опытно-конструкторских и технологических работ.

Алгоритм продемонстрировал свою эффективность. Предложенный подход может быть использован при координации данных в информационных системах.

СПИСОК ЛИТЕРАТУРЫ

1. Пошатаев О.Н., Съедин Д.Ю. Информационная система ЕГИСУ НИОКТР, как инструмент мониторинга и анализа работ в научно-технической сфере // Информатизация и связь. – 2016. – № 4. – С.46–52.
2. Newcombe H., Kennedy J., Axford S.J., James A.P. Automatic linkage of vital records // Science. – 1959. – Vol. 130. – P. 954–959.
3. Fellegi I.P., Sunter A.B. A theory for record linkage // Journal of the American Statistical Association. – 1969. – Vol. 64(328). – P. 1183–1210.
4. Monge A., Elkan C. The field-matching problem: Algorithm and applications. // Proceedings of the 2nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. AAAI Press. 1996. – P. 267–270.
5. Cohen W., Ravikumar P., Fienberg S. A comparison of string distance metrics for name-matching tasks // Proceedings of International Joint Conference on Artificial Intelligence. – 2003. – P. 73–78.
6. Бетин В.Н., Лукьянов С.Э., Супрун А.П. Выделение знаний из текстов на естественном языке в интеллектуальной аналитической системе // Информатизация и связь. – 2011. – № 6. – С. 51–54.
7. Бетин В.Н., Лукьянов С.Э., Супрун А.П. Оптимизация алгоритмов поиска решения в системах поддержки принятия решений, реализованных в формализме функциональных нейронных сетей // Информатизация и связь. – 2016. – № 4. – С. 37–45.
8. Бетин В.Н., Лукьянов С.Э., Супрун А.П. Использование метазнаний в системе поддержки принятия решений, реализованной в формализме сетей функциональных нейронов // Научно-техническая информация. Сер. 2. – 2016. – № 1. – С.16–20; Betin V.N., Lukyanov S.E., Suprun A.P. The Use of Metaknowledge in a Decision-Vaking Support System Implemented // Automatic Documentation and Mathematical Linguistics. –2016. – Vol.50, № 1. – P. 8–13.
9. Nitish A., Kartik A., Paul B. DERI&UPM: Pushing Corpus Based Relatedness to Similarity: Shared Task System Description // First Joint Conference on Lexical and Computational Semantics (*SEM), Montreal, Canada, June 7–8, 2012. Association for Computational Linguistics. P. 643–647.
10. Съедин Д.Ю. Новый стохастический гибридный алгоритм глобальной оптимизации на основе алгоритма М-РСА // Информатизация и связь. – 2017. – №1. – С.143–148.
11. Съедин Д.Ю. Параллельная реализация гибридного стохастического алгоритма глобальной оптимизации, основанного на алгоритме М-РСА // Информатизация и связь. – 2018. – №1. – С.150–156.
12. Recchia G., Louwerse M. A Comparison of String Similarity Measures for Toponym Matching // ACM SIGSPATIAL COMP'13, November 5, 2013. Orlando, FL, USA. – P. 54–61.
13. Camacho D., Huerta R., Elkan C. An Evolutionary Hybrid Distance for Duplicate String Matching. – 2008. – URL: http://arantxa.ii.uam.es/~dcamacho/StringDistance/hybrid-distance.pdf (Дата обращения: 25.02.2018).

Материал поступил в редакцию 19.03.18.

Сведение об авторе

СЪЕДИН Дмитрий Юрьевич – начальник отдела разработки и внедрения, Федеральное государственное автономное научное учреждение "Центр информационных технологий и систем органов исполнительной власти", Москва
e-mail: syedin@inevm.ru

Владимир Андреевич Успенский
(27.11.1930 – 27.06.2018)

Скончался Владимир Андреевич Успенский – один из пионеров отечественной информатики, известный математический логик, доктор физико-математических наук, профессор, заведующий кафедрой математической логики и теории алгоритмов Московского государственного университета им. М.В. Ломоносова, главный научный сотрудник ВИНИТИ РАН, член редколлегии сборника «Научно-техническая информация».

В.А. Успенский – выдающийся ученый, педагог и человек высокой математической и гуманитарной культуры.

Ученик и соратник академика А.Н. Колмогорова, Владимир Андреевич – один из создателей важного раздела математической логики и оснований математики – теории алгоритмов, влияние которой на алгоритмические языки, теорию и практику программирования неоценимо.

В трудное время 1950-х–1960-х гг. неприятия новых идей (кибернетики, математической лингвистики, структурной и прикладной лингвистики, семиотики) В.А. Успенский был одним из главных, умелых и смелых сторонников и пропагандистов прогрессивных веяний в науке. Благодаря его энергии и труду были изданы такие важные книги, как «Введение в метаматематику» С.К. Клини, «Математическая логика» А. Черча, «Введение в кибернетику» У. Эшби. Владимир Андреевич – один из наиболее глубоких знатоков теоремы К. Геделя о неполноте, ему принадлежат результаты, связанные с этой теоремой и выдающиеся по своей простоте и элегантности ее изложения.

Вклад В.А. Успенского в фундаментальную науку захватывает многие гуманитарные дисциплины: логику, лингвистику, филологию в целом, историю, историю науки. При этом его лингвистические статьи, бесспорно, вошли в золотой фонд отечественной науки, а лучшие филологические журналы почитали за честь иметь его в числе авторов.

Огромна роль В.А. Успенского как организатора науки. Он стоял у колыбели теоретической лингвистики нового поколения.

Мы гордимся тем, что много лет работали с этим замечательным человеком.

Благодарная память о Владимире Андреевиче сохранится у его многочисленных учеников, последователей и сотрудников.