

Методы и программные средства для ввода и верификации наукометрических данных

*Зензинов Андрей Александрович
НИИ механики МГУ имени М.В.Ломоносова
119192, Москва, Мичуринский пр-т, 1
andrey.zenzinov@gmail.com*

*Занчурич Максим Анатольевич
НИИ механики МГУ имени М.В.Ломоносова
119192, Москва, Мичуринский пр-т, 1*

*Афонин Сергей Александрович
к.ф.-м.н.,
НИИ механики МГУ имени М.В.Ломоносова
119192, Москва, Мичуринский пр-т, 1*

Аннотация: Для поддержки принятия административных решений в сфере науки и образования необходимо основываться на верифицированных данных по результатам деятельности отдельных исследователей и преподавателей. В статье предлагается описание подхода к организации процессов ввода данных по принципам «однократного» их сбора «снизу вверх» (от отдельных исследователей к научным коллективам) с использованием многоуровневой верификации: автоматической верификации данных с использованием внешних информационных ресурсов; верификации пользователем, соавторами и уполномоченными представителями организации.

Ключевые слова: ввод данных, верификация данных, наукометрия, библиометрия, многоуровневая верификация

Acquisition and Verification of Scientometric Data. Methods and Software Tools

*Zenzinov Andrey Alexandrovich
Lomonosov Moscow State University
andrey.zenzinov@gmail.com*

*Zanchurin Maxim Anatolyevich
Lomonosov Moscow State University*

*Afonin Sergey Alexandrovich,
Ph.D.
Lomonosov Moscow State University*

Abstract: Administrative decision support in the field of science and education should be based on verified scientometric data on the individual researchers and teachers activities results. The article proposes an approach to data acquisition processes organization based on the principles of “one-time” “bottom-up” data collection (from individual researchers to research teams) using multi-level verification: automatic data verification using external information resources; verification by the user, co-authors and authorized organization representatives.

The Abstract paragraph should be indented 1.25 inch on both the left and right-hand margins. Abstract must be width alignment, and in point size 12. Two line spaces precede the Abstract. The Abstract must be limited to one paragraph.

Keywords: data acquisition, data verification, scientometrics, bibliometrics, multi-level verification

1 Введение

В сфере науки и высшего образования для поддержки принятия административных и управленческих решений, как правило, используются наукометрические системы, именуемые также CRIS-системами (Current Research Information Systems). Такие системы аккумулируют данные о научной и педагогической деятельности не только организации в целом, но и о результатах научно-инновационной и педагогической деятельности отдельных её работников. При решении некоторых вопросов административного характера руководители опираются на данные по организациям в целом (формирование отчётных материалов, рейтинговых показателей и т.п.), для других задач используются данные по структурным подразделениям организации и по отдельным работникам (конкурсное избрание). Имея в виду социальные льготы, статус и авторитет отдельного работника (учёного и/или педагога) в коллективе, на которые прямо влияют персональные оценки, такие данные должны быть предельно корректными и актуальными. Они должны быть легко проверяемыми и не вызывать сомнений.

Наукометрические данные, значительная часть которых представляет собой описания результатов научно-инновационной и педагогической деятельности, можно охарактеризовать следующим образом. Такого сорта данные в CRIS-системах имеют сложные и разнообразные модели описания. Они связаны с большим количеством объектов разных типов (работники, подразделения, журналы, НИР и НИОКР, диссертации и др.), могут быть аффилированы с несколькими пользователями, и, как следствие, обладать сложной моделью разграничения доступа. Право на редактирование информации о результате выдаётся не только пользователю, осуществившему ввод результата в систему, но и его соавторам, а также *ответственным* — пользователям, которым организация делегировала полномочия на сопровождение информации по организации или соответствующему структурному подразделению.

В наиболее популярных существующих CRIS-системах (Pure¹, Converis²) источником данных являются библиометрические системы (Web of Science (WoS), Scopus, РИНЦ), которые, в свою очередь, аккумулируют данные, полученные от издательств, выпускающих научные издания. Таким данным можно доверять на уровне отдельных изданий и выходных данных публикаций. Однако привязка (аффилиация) отдельных публикаций к конкретным авторам не верифицируется на должном уровне, что часто приводит к ошибкам, например, в случае работающих в одной организации однофамильцев. В этой связи проецирование полученных таким образом («сверху вниз») данных на деятельность отдельного учёного не может в полной мере служить характеристикой его деятельности, а значит — не может стимулировать его к более эффективной работе.

Примером действующей в России системы для поддержки принятия административных решений в организациях науки и образования является информационно-аналитическая система (ИАС) «ИСТИНА» (Интеллектуальная Система Тематического Исследования Наукометрических данных) [1], которая в течение нескольких лет используется в режиме промышленной эксплуатации в МГУ имени М.В. Ломоносова и в ряде организаций РАН и Минздрава РФ.

1 Pure, Elsevier, <https://www.elsevier.com/solutions/pure>

2 Converis, Clarivate Analytics, <https://clarivate.com/products/converis/>

2 Постановка задачи

Для эффективного использования наукометрических данных в решении описанных выше административных и управленческих задач необходимо, чтобы было сформировано единое, легко верифицируемое информационное пространство. Такое пространство должно характеризоваться не только тем, что данные верифицированы с использованием всех доступных современных цифровых технологий и административных механизмов, но и тем, что они актуальны и полны в том смысле, что они всесторонне отражают научно-инновационную и педагогическую деятельность субъектов на всех уровнях (организации, коллектива или работника).

Цель настоящей работы — сформулировать общие требования к механизмам ввода и верификации наукометрических данных в CRIS-системах и предложить методы верификации такого рода данных. Далее Системой будет называться абстрактная CRIS-система, а примеры реализации таких механизмов будут даны со ссылкой на ИАС «ИСТИНА». С учётом того обстоятельства, что ИАС «ИСТИНА» — действующая система, будет проверено, соответствует ли она сформулированным требованиям к механизмам ввода и верификации наукометрических данных, а также по каким направлениям возможно дальнейшее улучшение.

Описание наукометрического результата будем считать *корректным*, если выполнены следующие условия: *корректность сведений* (все указанные сведения совпадают с указанными в оригинале результата, например, журнале); *корректность связей с другими сущностями в системе* (например, к статье привязан профиль верного автора Иванова И.И., а не его однофамильца); *единственность* (оригиналу результата соответствует единственный объект описания результата в Системе); *целостность* (описание результата соответствует единственному оригиналу). С учётом данного определения под *верификацией* будем понимать процесс проверки корректности описания наукометрического результата. Более подробно про различные виды верификации будет изложено далее в разделе "Верификация результатов с использованием источников".

Обеспечение полноты и актуальности данных требует наличия постоянных стимулов для ввода этих данных в систему, а также непосредственно на проведение исследований и получение результатов. Среди примеров таких стимулирующих факторов можно отметить следующие: конкурсное переизбрание; распределение поощрительных надбавок на конкурсной основе. В основе существующих стимулирующих процедур, которые используются в организациях, подключенных к ИАС «ИСТИНА», лежат механизмы рейтинговых оценок. Для того, чтобы такие оценки были объективны и корректны, используемые для их расчёта данные должны быть верифицированы и актуальны. Следует отметить, что рейтинговые оценки носят предварительный характер и не заменяют проведение экспертизы. Оценки, напротив, должны дополнять и усиливать экспертное рецензирование.

Для формирования верифицированного информационного пространства необходимо, чтобы в этом пространстве была налажена эффективная работа связки обозначенных на рис. 1 пяти компонент.

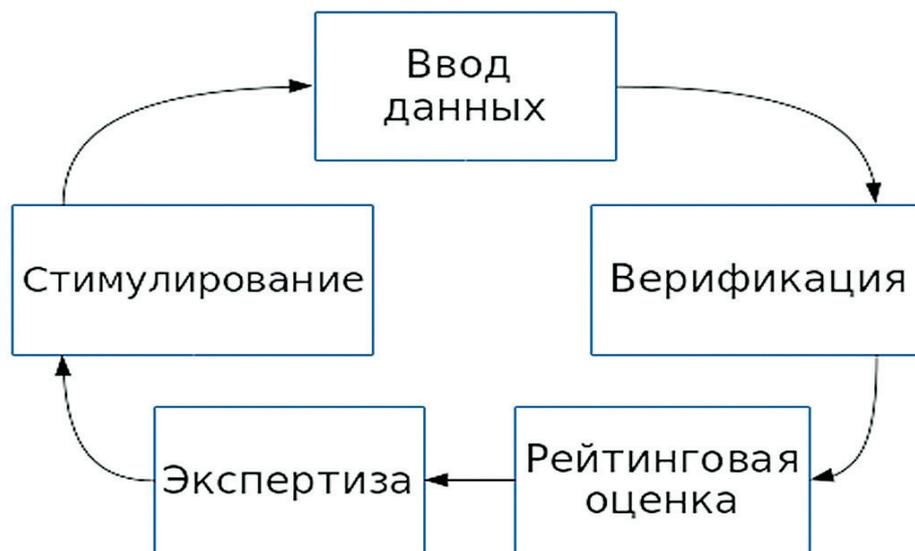


Рисунок 1 – схема формирования верифицированного информационного пространства

Требования к механизмам ввода и верификации

В качестве базовых принципов использования наукометрии для предварительной оценки научной, инновационной и образовательной деятельности, научное сообщество в 2014 году приняло Лейденский манифест о наукометрии [2]. Из этого манифеста можно выделить те принципы, которые имеют непосредственное отношение к процессам ввода и верификации наукометрических данных, они приведены ниже в переводе и с сохранением нумерации.

4. Процессы сбора данных и их анализа должны быть открытыми, прозрачными и простыми для понимания и возможности их воспроизведения.
5. Работникам, деятельность которых оценивается, должна быть предоставлена возможность проверять и анализировать данные, которые положены в основу оценки.

В дополнение к принципам Лейденского манифеста и с учётом реалий развития наукометрии в России, разработчики ИАС «ИСТИНА» сформулировали ряд других методологических принципов [1], которыми руководствуются при создании этой системы. Ниже приводятся только те принципы, которые имеют непосредственное отношение к процессам ввода и верификации наукометрических данных, как это было сделано для Лейденского манифеста.

1. При сборе и верификации данных в больших наукометрических коллекциях предпочтение следует отдавать восходящим потокам их поступления, основанным на персональном или коллективном интересе источника данных.
2. Процессы сбора и верификации данных должны рационально сочетать не только приоритетные восходящие «снизу вверх», но и нисходящие «сверху вниз» (из агрегированных данных к составляющих их индивидуальным) потоки.
3. Процессы сбора и верификации должны основываться на рациональном (сбалансированном) сочетании для отдельных работников и для коллективов стимулов как принудительного (приказного, обязывающего) характера, так и стимулов, основанных на персональном и/или коллективном интересе.

С учётом изложенных выше методологических принципов можно сформулировать следующие общие требования к подсистеме ввода и верификации:

- авторы результатов могут и должны участвовать в процессах ввода и верификации наукометрических данных (ввод данных по принципу "снизу вверх");
- данные о каждом отдельном результате достаточно ввести один раз (принцип "однократного ввода");
- для верификации данных, полученных по принципу "снизу вверх" должны использоваться данные, полученные по принципу "сверху вниз";
- подсистема должна позволять вводить результаты как научной, инновационной, так и деятельности по подготовке кадров.

2.1 Требования к интерфейсу

Для того, чтобы результаты были корректными (в определённом выше смысле), необходимо не только правильно организовать процессы верификации и корректировки данных, но и реализовать интерфейс для ввода и верификации наукометрических данных таким образом, чтобы снизить вероятность появления некорректных данных.

Можно сформулировать следующие требования к интерфейсу:

- интерфейс должен позволять автору вводить информацию о результате;
- связанные с результатом пользователи (соавторы) должны иметь возможность проверить и скорректировать введённую информацию;
- ответственные по местам работы автора или соавторов при проверке и подтверждении должны иметь возможность скорректировать информацию о результате самостоятельно или попросить об этом автора;
- при добавлении результата должна проводиться проверка на то, что он ещё не добавлен в Систему;
- если модель результата предусматривает связь с другими объектами Системы, то интерфейс должен позволять определить, существуют ли необходимые объекты или нет;
- при наличии технической возможности интерфейс должен позволять загрузить данные для формы из внешних систем по заданному идентификатору результата в этих системах, что, однако, не отменяет необходимости проведения проверки на дубликаты и сопоставления авторов;
- интерфейс должен быть понятным и непротиворечивым;
- механизмы ввода и верификации должны препятствовать "злоупотреблениям", например ошибочному приписыванию результату более высокого статуса или завышения показателей (количество страниц), которые могут учитываться в индикаторах рейтинговой оценки;
- интерфейс ввода должен обеспечивать валидацию данных, для которых существуют заранее заданные правила, например корректность ISSN можно проверять по последней цифре — проверочному разряду.

2.2 Верификация результатов с использованием источников

В зависимости от типа информации, её источника и технических возможностей, процесс верификации может быть *автоматическим* (без участия пользователя), *автоматизированным* (пользователь видит результаты выполнения алгоритмов и принимает самостоятельное решение) и *ручным* (без использования алгоритмов). При участии пользователя и наличии у него прав на редактирование, возможно также осуществлять корректировку результата. Автоматическая корректировка данных без участия пользователя и с учётом того обстоятельства, что в данных у источников могут быть ошибки, может привести к распространению этих ошибок.

Верификация может осуществляться как для уже добавленных в систему результатов, так и непосредственно во время их ввода в систему. Примерами такой проверки по вносимым данным могут быть: автоматизированная проверка на то, что такой результат не был внесён в систему; автоматизированное сопоставление авторов результата с профилями авторов в системе с использованием информации о предыдущих совместных результатах; загрузка части информации о результате по идентификаторам (например, DOI) с использованием API внешних систем.

Для того, чтобы понять, как именно можно верифицировать сведения о результатах, опишем модель информации о результате в общем виде:

- сведения об авторах результата;
- аффилиации авторов результата с организациями и связи результата с другими сущностями (гранты, диссертации, оборудование и т.п.);
- выходные данные для результата;
- содержательное описание результата (название, аннотации, ключевые слова, полный текст и др);
- вторичные аналитические данные о результате, которые появились после того, как результат был зарегистрирован (показатели цитируемости, доля некорректных заимствований, иные аналитические данные).

Результаты научно-инновационной и педагогической деятельности, как правило, зарегистрированы в изданиях (публикации), специализированных базах данных (патенты, авторские свидетельства) или официальных документах (чтение учебных курсов отражается в учебном плане). В таких источниках регистрируются основные сведения о результате (обычно в текстовом виде), а также может быть присвоен регистрационный номер. Такого типа источники могут использоваться при проверке корректности сведений, в первую очередь, о выходных данных. Те виды результатов, которые не регистрируются, не могут являться верифицируемыми, а значит — не могут лежать в основе оценки деятельности.

Для некоторых типов результатов могут существовать дополнительные источники данных, которые могут быть получены путём обработки первичных данных о результате. Примером такого источника может являться библиометрическая индексирующая система, которая получает данные непосредственно от издательств, индексирует их и производит новую информацию — импакт-факторы журналов, данные о цитированиях публикаций и другие. В качестве другого примера источника данных пост-обработки результатов можно выделить систему поиска некорректных заимствований "Антиплагиат" и её аналоги.

Для описанных выше типов источников могут быть доступны программные механизмы для поиска и проверки данных. Если есть подобный API и имеется доступ к этому интерфейсу, то возможна автоматическая или автоматизированная проверка данных. Если такого интерфейса нет, то возможна ручная проверка (проверка по документам, проверка по официальному сайту источника).

Второй тип источника сведений о результате — его авторы, которые непосредственно принимали участие в получении результата и осуществлять проверку связей этого результата с профилями его соавторов. В большинстве случаев авторы также принимают участие в процессе верификации во время ввода, о которой говорилось выше.

Третьим типом источника может служить сама организация или структурное подразделение этой организации в лице уполномоченного представителя, являющегося пользователем системы, который сможет проверить, насколько полученный результат соотносится с исследованиями подразделения или проверить наличие в статье указанной аффилиации автора с организацией. Кроме того, таким уполномоченным пользователям могут быть доступны сведения о данных всей организации или подразделения, и они могут проводить дополнительные проверки и корректировку данных с представлением об общей картине деятельности.

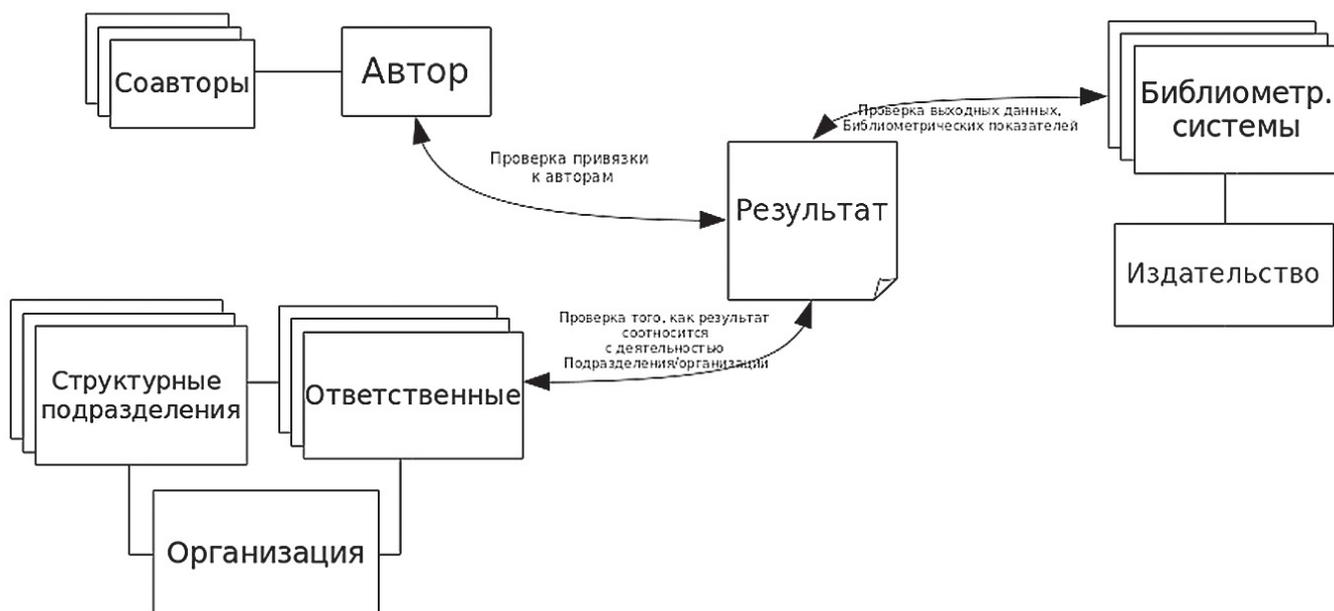


Рисунок 2 – схема верификации результата с использованием источников

Каждый из обозначенных выше источников обладает уникальной информацией о результате, которой нет у остальных или они не являются её первоисточником. На рис.2 изображена схема проверки данных с использованием различных источников. Поскольку информация, полученная от разных источников может различаться или содержать ошибки, то механизмы ввода и верификации должны иметь механизмы для разрешения возможных несоответствий. Если все источники верифицировали результат и несоответствий нет, то можно предположить, что итоговый результат корректен.

3 Реализация механизмов ввода и верификации наукометрических данных в ИАС «ИСТИНА»

В настоящем разделе описана реализация описанных методов и механизмов для ввода и верификации информации о научных публикациях в ИАС «ИСТИНА». В зависимости от типа результата и технических возможностей по проверке данных во внешних источниках конкретный набор действий может отличаться. Описание процесса ввода информации о публикациях в Систему представлено на рис. 3 в формате BPMN.

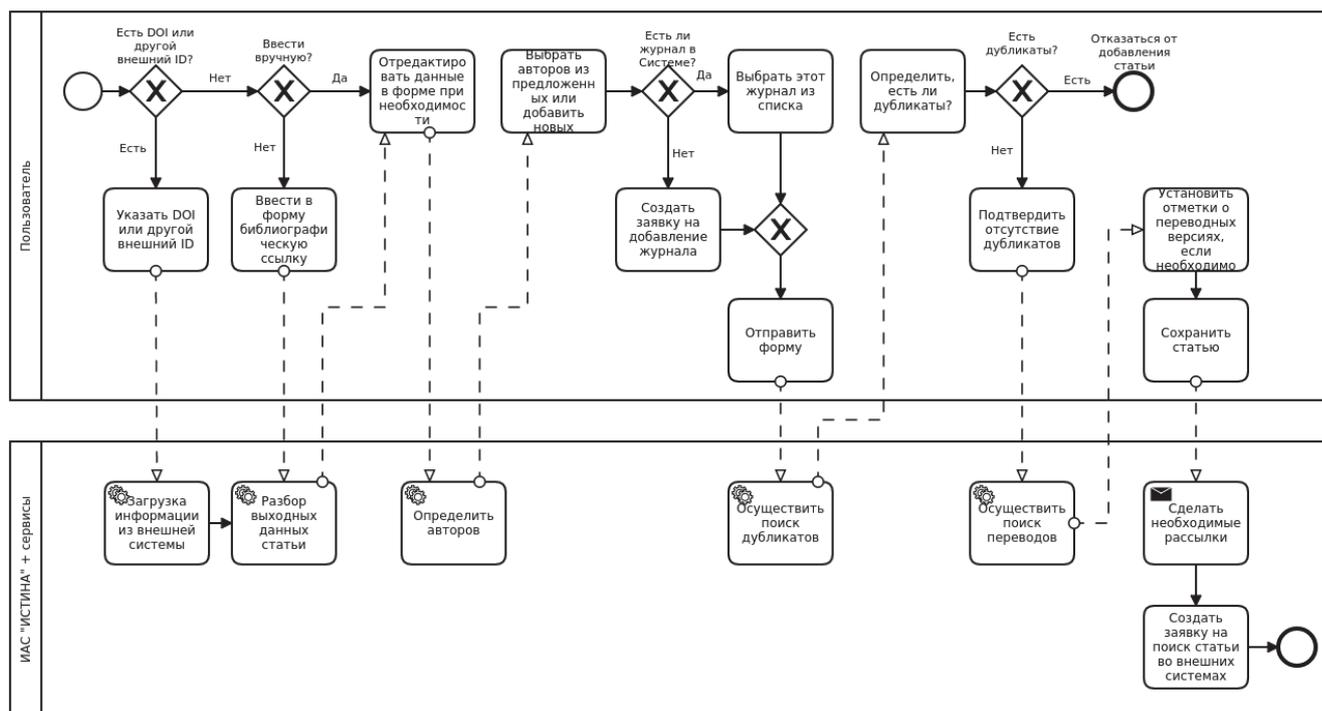


Рисунок 3 – схема бизнес-процесса ввода информации о публикации в ИАС «ИСТИНА» в формате BPMN

3.1 Ввод результата

Процесс ввода результата в Систему можно представить в виде определённой последовательности действий. Некоторые процессы верификации происходят непосредственно во время процесса ввода (сопоставление авторов, проверка наличия дубликатов), а оставшаяся часть реализуется уже после завершения ввода (проверка соавторами, ответственными).

Для ввода информации о публикации в Систему пользователю предлагается заполнить форму добавления публикации. Если у пользователя есть библиографическая ссылка или один из внешних идентификаторов (DOI или идентификатор публикации в Web of Science/Scopus), то поля формы могут быть заполнены автоматически. Разбор библиографических ссылок осуществляется с использованием сервиса FreeCite³, а для внешних идентификаторов используется API соответствующих систем (CrossRef для идентификатора DOI). При необходимости пользователь может скорректировать сведения в форме.

Часть полей формы представляет собой поле для указания ссылки на другие объекты Системы (журнал, сборник трудов конференции), в котором по названию или идентификаторам издания ISSN/ISBN может выбрать нужное издание. Рис. 4 демонстрирует форму ввода научной публикации. Поле "Журнал" предлагает пользователю найти по названию или ISSN журнал, данные о котором уже имеются в Системе. В том случае, если пользователь не смог найти журнал, он может отправить администраторам Системы заявку на добавление данных о журнале. Такая схема обусловлена тем, что библиометрическая информация о журналах оказывает значительное влияние на расчёт рейтинговых баллов, а именно — индексирование журнала в библиометрических базах данных и соответствующие импакт-факторы.

3 FreeCite — open source citation parser, <http://freecite.library.brown.edu>

Авторы	<input type="text" value="Petrov A., Boshirov R."/>	Тип статьи ¹ <input checked="" type="checkbox"/> Исследовательская ста Характеристики публикации: <input type="checkbox"/> Научно-популярная статья <input type="checkbox"/> Электронная публикация
Название статьи	<input type="text" value="is it a felony to visit an an ancient cathedral?"/>	
Журнал	<input type="text" value="Введите часть полного названия журнала"/> Моего журнала нет в списке	
Номер журнала	<input type="text" value="Natio secur"/>	
Том	<input type="text" value="Central European Journal of International and Security Studies"/> <small>Нет дополнительных данных об этом журнале</small>	
Первая страница	<input type="text" value="Comparative labor law journal : a publication of the U.S. National Branch of the International Society for Labor Law and Social Security [and] the Wharton School, and the Law School of the University of Pennsylvania"/> <small>ISSN: 1043-5255</small>	
Последняя страница	<input type="text" value="Integrating Science and Technology for sustainable Energy Development as Basis for secure Future. Proceedings of the International Humboldt Conference,"/> <input type="checkbox"/>	
Год издания	<input type="text"/>	
Объём (п.л.)	<input type="text"/>	
<input type="button" value="Продолжить добавление публикации"/> <input type="button" value="Очистить форму"/>		

Рисунок 4 – интерфейс формы ввода публикации в ИАС «ИСТИНА»

Подтверждение данных о публикации

dsdfsdsdgsd, International Journal of Foundations of Computer Science, 1(1), 1--2, 1996

Внимание: в базе найдены похожие публикации

Пожалуйста, проверьте, что среди приведенных ниже публикаций нет той, которую Вы сейчас добавляете. Если дубликат есть, но не отображается в списке Ваших работ, то возможно Вам нужно указать свое авторство. Для этого перейдите по ссылке редактирования соответствующей публикации.

[1] Шестаков О. В. Регуляризация метода реконструкции функции по ее сферическому преобразованию Радона // *Вестник Московского университета. Серия 15. Вычислительная математика и кибернетика.* — 2012. — № 1. — С. 23–27.
[редактировать](#)

[1] Зотов А. Ф., Смирнова Н. М. Метафизика свободы (окончание) // *Философские исследования.* — 1994. — № 1. — С. 188–198.
[редактировать](#)

Сведения о журнале

International Journal of Foundations of Computer Science
 Показатели цитирования: **Impact Factor 2013 0,326 5-Year Impact Factor 2012 0,430** [подробнее](#)
 Индексирование: Scopus (24 июня 2002 г.-), JCR (18 апреля 1985 г.-)

Пожалуйста, проверьте правильность определения авторов публикации ¹

Автор публикации	Выбранная запись в базе
afonin	Alexander Ivanov

Рисунок 5 – интерфейс подтверждения ввода публикации в ИАС «ИСТИНА»

После отправки формы происходит проверка на наличие дубликатов публикации в Системе (рис. 5). На этой же странице пользователь может проверить информацию о журнале и проверить правильность автоматически предложенного сопоставления авторов публикации с пользователями Системы и теми авторами, информация о результатах которых уже введена, но они ещё не зарегистрированы. За разрешение неоднозначности при определении привязок авторов отвечает специальный сервис, в основе которого лежит графовый алгоритм [3], учитывающий информацию о соавторстве в ранее введённых в Систему результатах.

Для некоторых видов результатов могут быть доступны дополнительные проверки. Для научных публикаций такой проверкой может служить поиск ранее добавленных в Систему переводных версий публикации с использованием графового алгоритма и графа соавторств [4].

После подтверждения пользователем указанных данных информация о публикации сохраняется в Систему. Указанным в форме соавторам отправляются оповещения о добавленном в Систему результате, автором которого они, возможно, являются. Все соавторы, а также ответственные пользователи по подразделениям, в которых работают авторы, могут корректировать данные о результате, а также указывать дополнительные данные о публикации, такие как аффилиации авторов с организациями, аффилиации публикации с другими результатами

(проектами, диссертациями и другими). Если привязка результата к автору была произведена ошибочно, то соответствующий этому автору пользователь может отказаться от авторства и публикация не будет отображаться в его профиле.

3.2 Верификация во внешних по отношению к Системе источниках

Периодически запускается пакетный поиск публикаций во внешних индексирующих библиометрических системах Web of Science и Scopus с использованием их API. Поисковые механизмы этих систем позволяют формировать запросы в виде различных наборов выходных данных: название публикации, имена авторов, номера страниц, год выхода публикации, DOI, название журнала, ISSN журнала, номер выпуска и т.п. Поскольку в сведениях могут быть ошибки или различия в написании спецсимволов (публикации по математике или химии), то представляется целесообразным осуществлять поиск не только по полному набору имеющихся данных о публикации, но также и по подмножествам этого набора. Точность поиска при уменьшении подмножества снижается, поэтому запросы упорядочены по степени "надёжности" такого подмножества. Подмножества, по которым невозможно однозначно определить публикацию, например год и номера страниц, не используются. Если поисковому механизму удалось найти статью во внешней системе, то осуществляется её "привязка": сохраняется идентификатор этой статьи в той системе, где она была найдена. Подобный алгоритм, использующий подмножества набора данных, используется в алгоритмах сопоставления библиографической записи с описанием статьи в Web of Science в CWTS и iFQ, ознакомиться с принципом работы которых можно в [5]. Следует отметить, что подобный принцип поиска является универсальным и может применяться к другим типам результатов в случае наличия соответствующих поисковых механизмов и доступа к ним.

Для тех публикаций, которым удалось создать "привязки", периодически производится обновление числа цитирований по тем показателям, которые рассчитываются библиометрическими системами. Если статья проиндексирована и в Web of Science, и в Scopus, то у неё появляется два показателя: число цитирований по данным Web of Science; число цитирований по данным Scopus.

В зависимости от функциональных возможностей API внешних систем и ограничений подписки организаций к этим системам, возможна реализация дополнительных верификационных механизмов. Примером такого механизма может служить загрузка аффилиаций авторов с организациями. Для корректного учёта опубликованных организацией статей важно, чтобы в каждой статье были корректно указаны аффилиации авторов с этой организацией, поскольку эти публикации выполнены в рамках работы автора в этой организации. Эта информация предоставляется самими авторами во время подготовки публикации, а затем передаётся в издательство. API Scopus, в отличие от сервиса WoS AMR позволяет выгружать данные об аффилиациях, которые должны быть указаны в публикациях. Для публикаций, которые не индексируются в Scopus, возможен ручной ввод аффилиаций авторами или ответственными (рис. 6).

Редактировать аффилиацию авторов статьи

2018 Структура нанопроволок Cu/Ni, полученных методом шаблонного синтеза

Жигалина О.М., Долуденко И.М., Хмеленин Д.Н., Загорский Д.Л., Бедин С.А., Иванов И.М.

в журнале *Кристаллография*, издательство ФГУП Издательство «Наука» (Москва), том 63, № 3, с. 455-462 DOI

Выберите авторов статьи

- Жигалина Ольга Михайловна (имя в статье - Жигалина О.М.)
- Долуденко И.М. (имя в статье - Долуденко И.М.)
- Khmelinin D.N. (имя в статье - Хмеленин Д.Н.)
- Загорский Дмитрий Львович (имя в статье - Загорский Д.Л.)
- Бедин С.А.
- Иванов И.М.

Выберите организацию, в которой работают авторы статьи

Введите для поиска несколько символов

Выберите подразделение, в котором работают авторы статьи

Введите для поиска несколько символов

Название организации, как указано в статье (если отличается, или не нашли в списке выше)

Институт кристаллографии им. А.В. Шубникова РАН

Страна внешней организации (если не нашли в списке выше)

Город внешней организации (если не нашли в списке выше)

Москва

Сохранить и закрыть

Рисунок 6 – форма для редактирования аффилиаций авторов статьи

Другой пример использования API внешних систем — сопоставление списка проиндексированных публикаций для отдельных авторов. Если автор указал в своём профиле в ИАС «ИСТИНА» свой идентификатор во внешней системе, например, Scopus, то можно сравнить те публикации, для которых осуществлена "привязка" и указано его авторство с теми, которые во внешней системе привязаны к соответствующему автору. Подобная проверка позволяет выявить не только отсутствующие или некорректные привязки в Системе, но и определить ошибки в данных внешней системы.

3.3 Интерфейс для подтверждения результатов

Важной частью верификации результатов является механизм проверки результатов ответственными. Эта верификация является многоуровневой и предполагает, что результаты сначала проверяются ответственными по подразделениям нижнего уровня (кафедры, лаборатории), а затем проверяются ответственными по подразделениям верхнего уровня (факультеты, институты, центры) или по всей организации. Ответственные на нижнем уровне имеют представление о том, какая деятельность ведётся в их подразделениях и могут подтвердить данные о результате, направить результат на исправление информации авторам, или отклонить результат с пометкой, что он не относится к деятельности подразделения.

Ответственные на верхнем уровне получают данные о результатах работников со всех подразделений нижнего уровня (рис. 7), поэтому они должны иметь возможность проверить, подтверждён ли результат ответственным по кафедре, или отклонён, а если отклонён, то по какой причине.

При подсчёте рейтинговых баллов по формуле, принятой в подразделении, может учитываться наличие подтверждения. В зависимости от настроек конкретных формул в расчёт могут попасть те результаты, которые не были отклонены ответственными ("чёрный список"), или же только те результаты, которые были подтверждены ответственными ("белый список").

Выберите год **2018**

Выберите подразделение **Биологический факультет**

Выберите тип результата **Все результаты**

Отображать только результаты с подтверждениями

Применить

[К началу списка](#) [Следующая страница >>](#)

Структурное подразделение, на уровне которого будет подтверждён результат	Название результата научной деятельности	Подтверждения ответственными в дочерних структурных подразделениях	Подтверждение
Биологический факультет	<i>Книга</i> 2018 Имобилизованные клетки: биокатализаторы и процессы Ефременко Е.Н., Андриюшина В.А., Балабанова Т.В., Беклемишев А.Б., Варфоломеев С.Д., Водякова М.А., Демаков В.А., Дитченко Т.И., Джавахия В.В., Дроздова М.Г., Завьялова Н.В., Ившина И.Б., Исмаилов А.Д., Карпова Н.В., Коваленко Г.А., Криворучко А.В., Куюкина М.С., Лозинский В.И., Лягин И.В., Максимова Ю.Г., Мамедова Ф.Т., Марквичева Е.А., Маслова О.В., Махлис Т.А., Молчан О.В., Перминова Л.В., Плеханова Ю.В., Решетилов А.Н., Сенько О.В., Степанов Н.А., Стыценко Т.С., Филиппова С.Н., Холстов А.В., Юрин В.М., Ядерец В.В. место издания <i>РИОР Москва</i> , ISBN 978-5-369-02004-3, 500 с. DOI	1 подтверждение	<input checked="" type="checkbox"/> Подтвердить <input checked="" type="checkbox"/> Отклонить
Биологический факультет	<i>Книга</i> 2018 Методы клеточной биологии и цитогенетики Алиева И.Б., Гольшев С.А., Жиронкина О.А., Киреев И.И., Курчашова С.Ю., Стрелкова О.С., Узбеков Р.Э. место издания <i>Издательство "Перо" Москва</i> , ISBN 978-5-00122-831-8, 260 с.	1 подтверждение	<input checked="" type="checkbox"/> Подтвердить <input checked="" type="checkbox"/> Отклонить
Биологический факультет	<i>Преподавание курса</i> 24 декабря 2018 - 24 декабря 2018 Актуальные проблемы гидробиологии МГУ имени М.В. Ломоносова, Биологический факультет обязательная, базовой части, практические занятия, 7 часов	1 подтверждение	<input checked="" type="checkbox"/> Подтвердить <input checked="" type="checkbox"/> Отклонить
Биологический факультет	<i>Преподавание курса</i> 22 декабря 2018 - 22 декабря 2018 Актуальные проблемы гидробиологии МГУ имени М.В. Ломоносова, Биологический факультет, Кафедра гидробиологии обязательная, базовой части, практические занятия, 7 часов	1 подтверждение	<input checked="" type="checkbox"/> Подтвердить <input checked="" type="checkbox"/> Отклонить
	<i>Преподавание курса</i>		

Рисунок 7 – интерфейс ответственного для подтверждения результатов научной деятельности

Если авторы результата работают в нескольких структурных подразделениях, то такая проверка производится по всем этим подразделениям. Результат может быть подтверждён на одном факультете, но отклонён на другом, поскольку он по мнению ответственного не связан с деятельностью этого факультета.

4 Заключение

Сформулированы требования к методам и программным механизмам ввода и верификации наукометрических данных. Механизмы, которые используются в ИАС «ИСТИНА» на сегодняшний день, реализуют многоуровневую верификацию и не противоречат сформулированным требованиям, однако опыт эксплуатации и отзывы пользователей выявили необходимость улучшения эргономичности и понятности интерфейса, что позволит снизить вероятность появления некорректных данных, которые подлежат корректировке. В настоящее время ведётся работа над улучшением интерфейса, а также над новыми механизмами верификации во внешних системах.

Узким местом с точки зрения вероятности появления ошибок является человеческий фактор, влияние которого на процессы ввода и верификации необходимо снижать, сохраняя в то же время соблюдение базового принципа участия авторов в этих процессах. Реализация механизмов импортирования информации о результатах из внешних

систем позволит снизить вероятность появления опечаток, но требует в то же время проведения автоматизированного сопоставления авторов и установки связей с другими сущностями системы, которое затем будет передано на проверку авторам. Проверка результатов ответственными является трудоёмкой и требует повышенного внимания. В периоды "массовой" проверки, например во время проведения стимулирующих конкурсов, ответственным приходится подтверждать большое количество результатов в сжатые сроки. Разработка вспомогательных механизмов, которые могли бы предлагать ответственным обращать внимание на вероятные ошибки или эффективно их устранять, могла бы значительно упростить работу ответственных и снизить вероятность появления ошибок.

Участие систем, реализующих предложенные в настоящей статье методы и механизмы ввода и верификации, в процессе формирования верифицированного информационного пространства наукометрических данных будет более эффективным, если будет налажено обратное взаимодействие с источниками информации. Поскольку информация о результатах проходит многоуровневый верификационный "фильтр", то периодически выявляются ошибки в данных источников, которые могли бы быть скорректированы. Системы, подобные ИАС «ИСТИНА», могли бы и сами выступить в роли источника верифицированных наукометрических данных. Опыт показывает, что при поиске информации о научной деятельности отдельного учёного часто используют эту систему.

Список использованной литературы

- [1] Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. Bibliometrics: The Leiden Manifesto for research metrics // *Nature*. — 2015. DOI: 10.1038/520429a
- [2] Садовничий В. А., Васенин В. А. Интеллектуальная система тематического исследования наукометрических данных: предпосылки создания и методология разработки. Часть 1 // *Программная инженерия*. — 2018. — Т. 9, № 2. — С. 51–58. DOI: 10.17587/prin.9.51-58
- [3] Афонин С. А., Гаспарянц А. Э. Разрешение неоднозначности авторства публикаций при автоматической обработке библиографических данных // *Программная инженерия*. — 2014. — № 1. — С. 25–29.
- [4] Kozitsyn A. S., Afonin S. A., Zenzinov A. A. Linking translated articles using authorship statistics // 20th Conference Scientific Services and Internet, SSI 2018. — Vol. 2260 of CEUR Workshop Proceedings (CEUR-WS.org). — CEUR-WS CEUR Workshop Proceedings, Vol-2260, <http://ceur-ws.org>, 2018. — P. 306–312.
- [5] Olensky, M., Schmidt, M., & van Eck, N. J. Evaluation of the citation matching algorithms of CWTS and iFQ in comparison to the Web of science // *Journal of the Association for Information Science and Technology*. — 2016. — 67(10). — 2550–2564. DOI: 10.1002/asi.23590