

Технология создания комплексированных индексов публикационной активности

© Обухова О.Л.

© Заикин М.Ю.

© Соловьев И.В.

Институт проблем информатики РАН
Москва

obuhova@amsd.com

maxim@amsd.com

vanya@kakoe-to.ru

Аннотация

В статье дан анализ требований к формированию показателей, на основе которых производится оценка результативности научных организаций в соответствии с Приказом Министерства образования и науки № 406 от 14 октября 2009 г.

Для автоматизации процесса получения данных о публикационной активности сотрудников научно-исследовательского института, как одного из критериев научного вклада организации, разработана технология, заложенная в основу информационно – поисковой системы, разрабатываемой в ИПИ РАН.

В качестве одного из прикладных результатов работы системы приведена статистика цитирований в статьях, опубликованных в трудах конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (RCDL) за последние 4 года.

1 Введение

Одним из показателей научного вклада интеллектуальных работников является публикационная активность, определяемая количеством публикаций автора и числом их цитирования, для подсчета которого используются библиографические ссылки научных статей. Пристатейные списки используемой литературы – «это кодовый язык, позволяющий представлять публикации в компактной форме» [1], и, в общем случае, являются объектом библиометрического исследования. Наблюдение за цитированием научных работ позволяет проследить за развитием той или иной идеи во времени, за проникновением ее в смежные области. Д. Прайс, один из основоположников библиометрии, отметил, что цитирование образует «сеть, связывающую все работы в единый комплекс» [2]. Ю. Гарфилд, основатель Института

научной информации (Institute of Scientific Information) США, в статье «Science Citation Index-A New Dimension in Indexing» [3] вводит понятие *индекса цитирования*: «...это упорядоченный список цитированных статей, каждая из которых сопровождается списком цитирующих статей. Цитирующая статья определяется как источник, а цитируемая – как ссылочная. Любая статья-источник может впоследствии стать ссылочной».

В настоящее время понятие «*индекса цитирования*» используется, в широком смысле, для обозначения специализированной информационной программной системы, в которой собирается и обрабатывается полная библиографическая информация о журнальных статьях и списках цитируемой литературы. Такая база данных позволяет находить как публикации, цитируемые в отдельно взятой статье, так и публикации, цитирующие эту статью. Анализ библиографических ссылок позволяет получить количественные показатели цитирования, к примеру, суммарный объем цитирования или индекс Хирша. *h-индекс*, или индекс Хирша — библиометрический показатель, предложенный в 2005 г. американским физиком Хорхе Хиршем [5] из университета Сан-Диего, Калифорния. Индекс Хирша является количественной характеристикой продуктивности учёного, основанной на количестве его публикаций и количестве цитирований этих публикаций. Индекс вычисляется на основе распределения цитирований работ данного исследователя. Хирш пишет: «Учёный имеет индекс *h*, если *h* из его *N_p* статей цитируются как минимум *h* раз каждая, в то время как оставшиеся (*N_p — h*) статей цитируются не более, чем *h* раз каждая».

Индекс цитирования, в узком смысле, позволяет судить о научной производительности автора и является одним из показателей результативности научной организации, алгоритм вычисления которой определен Приказом Министерства образования и науки Российской Федерации от «14» октября 2009 г. № 406 об утверждении:

«Типового положения о комиссии по оценке результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения, и Типовой методики оценки результативности деятельности научных организаций, выполняющих научно-исследователь-

Труды 14-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL-2012, Переславль-Залесский, Россия, 15-18 октября 2012 г.

ские, опытно-конструкторские и технологические работы гражданского назначения» [4].

Составной частью методики оценки научной организации является пункт «Публикационная активность» раздела «Научный потенциал и эффективность научных исследований». Для оценки публикационной активности методикой определяется ряд показателей: количество публикаций, цитируемость по каждой публикации и импакт-фактор журнала, где опубликована работа. Учитываются те журнальные статьи, которые зафиксированы в информационных ресурсах:

- реферативно-библиографическая база данных научного цитирования Web of Science, разработанная Институтом научной информации США;

- Российский индекс научного цитирования (РИНЦ), функционирующий на платформе www.e-library.ru

2 Международные и отечественные индексы цитирования

На мировом рынке научной информации наиболее известны и широко представлены два международных индекса цитирования: Web of Science компании Thomson Reuters и Scopus компании Elsevier. Оба эти проекта являются коммерческими, доступ к ним осуществляется на платной основе.

База данных научного цитирования Web of Science (WOS) [6] является одним из продуктов, входящим в состав информационно-поисковой системы компании Thomson Reuters ISI Web of Knowledge [7]. WOS предоставляет исследователям и специалистам информацию по всем отраслям знания среди свыше 12 000 журналов и 148 000 материалов конференций в области естественных, общественных, гуманитарных наук и искусства.

Индекс цитирования Scopus [8] разрабатывается компанией Elsevier. Представляет собой крупнейшую в мире единую реферативную базу данных, которая индексирует более 18000 наименований научно-технических и медицинских журналов порядка 5000 международных издательств, а также, в отличие от WOS, Интернет - источников. Scopus обеспечивает программную поддержку в поиске научных публикаций и предлагает ссылки на все вышедшие рефераты из всего объема доступных статей.

Для специалистов в области информатики определенный интерес представляют такие библиографические сервисы, как DBLP [9], ACM Digital Library [10], CiteSeerX [11], ориентированные на индексацию и учет цитирования публикаций по компьютерным технологиям, а также Google Scholar [12]. Но они «непрозрачны» в отношении того, как в них собирается и обрабатывается библиографическая информация. Поэтому нельзя поставить их в один ряд с вышеупомянутыми реферативными базами научного цитирования.

Задача индексирования публикаций российских ученых и учет их цитирования в достаточном представительном объеме, реализующий задачи Web of Science в отечественной среде, возложена на Российский индекс научного цитирования РИНЦ, база данных которого доступна на платформе eLibrary.ru[13].

Целью РИНЦ является систематическое продвижение в сети Интернет российских научных изданий посредством реализации следующих задач:

- создание многоцелевой поисковой системы по публикациям российских ученых, включающей статьи из научных журналов;

- разработка механизмов и инструментария для статистического анализа отечественной науки;

- создание и формирование Единого реестра публикаций российских ученых, авторитетную базу данных, представляющую максимально полную и достоверную информацию о публикационном потоке российских ученых, независимо от источника, времени, места и типа публикации;

- создание эффективной системы навигации в массиве научной информации и обеспечение доступа российских пользователей к полным текстам публикаций через механизмы унифицированного доступа;

- доступность по сравнению с западными аналогами.

Достоинства и недостатки каждой из этих реферативных баз научного цитирования авторы описали в своей статье [14]. Но именно индексы цитирования WOS и РИНЦ, в соответствии с Приказом Министерства образования и науки №406 от 14 октября 2009 г, являются источником для получения сведений о публикационной активности научных работников.

В рамках исполнения данного Приказа по поручению Президиума Академии Наук России была разработана и внедрена автоматизированная система учета результатов интеллектуальной деятельности Российской академии наук.

3. Автоматизированная система учета результатов интеллектуальной деятельности Российской академии наук АСУ РИД

АСУ РИД РАН [15] – система, предназначенная для автоматизации процессов учета результатов интеллектуальной деятельности Российской академии наук.

В данной статье не ставится задача полного описания данной системы. Поэтому рассмотрим ту её часть, которая относится к показателям публикационной активности. Поскольку сбор и предоставление статистики о публикационной жизни научно-исследовательских институтов РАН является обязательным ежегодным действием, то дадим пояснения по некоторым показателям, расшифровывающие их смысл и значение.

2.4.1. Число научных публикаций работников научной организации, опубликованных в отчетном году

Пояснение. Данный показатель определяется числом научных публикаций работников научной организации за каждый год в любых научных журналах, в тематических сборниках, в иных видах печатных изданий, а также на официальных интернет-сайтах признанных научных изданий. Каждая публикация учитывается один раз для организации не зависимо от того, в скольких изданиях она вышла и сколько соавторов участвовало в ее написании. Переводы статей не учитываются

2.4.2. Число научных публикаций в зарубежных научно-технических изданиях работников научной организации, опубликованных в отчетном году

2.4.3. Число научных публикаций работников научной организации в отечественных изданиях, включённых в перечень ВАК Минобрнауки России

2.4.4. Число публикаций работников научной организации в Российском индексе научного цитирования (РИНЦ)

2.4.5. Цитируемость работников научной организации в РИНЦ

Пояснение. Данный показатель определяется общим числом ссылок, появившихся в отчетном году, на все публикации работников научной организации в РИНЦ

2.4.6. Число публикаций работников научной организации в базе Web of Science

2.4.7. Число публикаций работников научной организации других базах реферативной информации

Пояснение. Данный показатель определяется общим числом публикаций (научных статей) работников научной организации в базах реферативной информации, за исключением баз РИНЦ и Web of Science.

2.4.8. Цитируемость работников научной организации в указанных базах реферативной информации

Пояснение. Данный показатель определяется общим числом ссылок, появившихся в отчетном году на все публикации работников научной организации в реферативных базах (ВКЛЮЧАЯ Web of Science и НЕ ВКЛЮЧАЯ РИНЦ).

2.4.9. Средний импакт-фактор публикаций работников научной организации в Web of Science

Пояснение. Для расчета данного показателя необходимо общее количество статей, опубликованных в отчетном году в каждом журнале, зарегистрированном в Web of Science, умножить на импакт-фактор соответствующих журналов в соответствующем году и сумму этих произведений разделить на общее количество статей вышедших в отчетном году.

Если данные об импакт-факторах журналов за отчетный год еще не опубликованы, то следует

взять импакт-факторы соответствующих журналов за предыдущий отчетный год.

$$I_p = \text{Сумма}(n_j * I_j) / N, j=1, m.$$

Здесь

I_p - импакт-фактор публикаций работников научной организации в Web of Science;

n_j - количество опубликованных в отчетном году статей в j -ом журнале, зарегистрированном в Web of Science;

I_j - импакт-фактор j -ого журнала в соответствующем году;

N - общее число статей, опубликованных научной организацией за отчетный год в журналах, индексируемых в Web of Science.

2.4.10. Число докладов, тезисов докладов, представленных работниками научной организации на наиболее значимых конференциях, симпозиумах и чтениях

2.4.11. Количество опубликованных монографий, учебников, научно-справочных изданий, словарей, атласов, энциклопедий и проч.

Для автоматизации работы по генерации данных показателей в ИПИ РАН разрабатывается программная система «Публикационная деятельность (ПД)», целью которой является формирование указанных показателей и создание комплексированных индексов научного цитирования.

4. Цели и задачи программной системы «Публикационная деятельность»

Программная система предназначена для сбора, хранения и предоставления web-доступа к информации, представляющей собой:

- полные списки научных трудов авторов;
- списки цитирований научных трудов авторов;

Дополнительно публикационная активность описывается:

- вычисляемым индексом Хирша автора;
- импакт-факторы журналов, в которых публикуется автор.

Научными трудами являются журнальные статьи в отечественных и зарубежных изданиях, монографии, учебники и учебные пособия, темы НИР, объекты интеллектуальной собственности, диссертации, доклады и тезисы на конференциях

Задача сбора и занесения в базу данных системы «ПД» данных о публикациях автора является чисто технической, на ней мы останавливаться не будем. Интерес вызывает проблема учета цитируемости научных публикаций. Поскольку создание программной системы продиктовано Приказом министерства образования и науки, в котором указываются источники информации Web of Science и РИНЦ на платформе eLIBRARY.ru, то, по определению, данные о цитированиях научных статей должны браться из этих источников. Но вопрос получения данных цитирования из предоставляемых отчетов eLIBRARY.ru является

нетривиальным. По ряду причин статистические отчеты данного ресурса представляют списки публикаций и их цитирований в двух видах: привязанные и непривязанные. Идеологи и разработчики eLIBRARY.ru признаются, что «привязанные» публикации не формируют полного списка публикаций автора, хранящихся в базе данных данного ресурса. И уж, тем более, не охватывают весь объем публикационной работы автора. Существенная часть публикаций авторов оказывается в списках «непривязанных» публикаций, и как следствие, полные списки цитирований автора должны извлекаться из той части системы, которая позиционируется разработчиками РИНЦ как «непривязанная».

Таким образом, перед авторами статьи, которые участвуют в разработке программной системы «Публикационная деятельность», стоит задача извлечения и объединения данных о списках публикаций и их цитирования из различных источников, имеющих разную природу и принципы организации. Объединение данных не может быть формальным суммированием, а требует комплексной обработки, в частности, для получения комплексированных индексов цитирования.

На первом этапе для решения данной задачи необходимо разработать концептуальную модель базы данных, чтобы определить все информационные объекты и принципы построения связей между ними.

5 Концептуальная модель базы данных

Для построения концептуальной модели в данной статье применяются принципы объектно-ориентированного программирования, при которых традиционно используются два вида организации данных: классы и типы.

Типы данных являются спецификациями структуры и поведения информационных объектов (ИО).

Классы специфицируют множество однотипных объектов. Каждому классу сопоставляется тип, описывающий структуру экземпляров класса.

В данной статье для описания типов и классов используются принципы фасетной классификации. Типы информационных объектов, используя методы, базирующихся на фасетной классификации и ранее разработанных авторами [16], представим в виде *фасетной формулы объектов (ФФО)* [17]. ФФО представлены в виде множества совокупностей: <фасетный признак : список значений> для всех ИО. ФФО являются аксиомами для предлагаемой концептуальной модели.

Использование фасетного подхода позволяет унифицировать «три пространства спецификаций: агрегация (структура объектов), обобщение/специализация и классификация» [18].

При спецификации типов данных мы принимаем во внимание, что значением свойств объектов одного класса являются объекты других классов.

Введем определения классов для предлагаемой модели базы данных.

SpecializedDataEntity – класс, экземплярами которого являются научные публикации. Значений однородных свойств объекта «публикация», которые являются экземплярами других классов, сгруппированы в классы $\text{Facet}_1, \dots, \text{Facet}_n$.

Определение 1. Множество

$$X = \{ X_j^i \mid i=0, \dots, n, j=1, \dots, n_i \},$$

где индекс i определяет номер класса, j определяет номер экземпляра данного класса, определяется множество всех ИО. Примем следующее обозначение:

X_m^o - экземпляры класса **SpecializedDataEntity**,

$m=1, \dots, m_0$, m_0 - количество экземпляров данного класса;

X_k^f - экземпляры классов **Facet_f**,

$f=1, \dots, n$, $k=1, \dots, k_f$, k_f – количество экземпляров соответствующего класса.

Определение 2. Двуместный предикат $S(X_m^o, X_k^f)$

обозначает, что объект X_m^o имеет свойство X_k^f , где $m \in \{1, \dots, m_0\}$, $f \in \{1, \dots, n\}$, $k \in \{1, \dots, k_f\}$.

Как выше было сказано, для представленной модели значением свойства объекта является объект другого класса, с которым установлено отношение данного объекта.

Определим множество $S' = \{ S(X_m^o, X_k^f) \}$ для таких

$m \in \{1, \dots, m_0\}$, $f \in \{1, \dots, n\}$, $k \in \{1, \dots, k_f\}$, для которых $S(X_m^o, X_k^f) = \text{true}$ - множество свойств всех ИО представленной модели предметной области.

Поскольку понятия «бинарное отношение» и «признак (свойство)» идентичны в случае, если значением свойства является объект другого класса, то определим набор свойств объектов классов **Facet_f**, $f=1, \dots, n$, $k=1, \dots, k_f$, следующим образом.

Определение 3. Двуместный предикат

$R_{f_1 f_2}^2 (X_{i_1}^{f_1}, X_{i_2}^{f_2})$ обозначает бинарное отношение двух объектов $X_{i_1}^{f_1}$ и $X_{i_2}^{f_2}$,

где $X_{i_1}^{f_1}, X_{i_2}^{f_2}$ - объекты классов **Facet_{f₁}** и **Facet_{f₂}**, $f_1 \neq f_2$, $i_1 \in \{1, \dots, k_{f_1}\}$, $i_2 \in \{1, \dots, k_{f_2}\}$, k_{f_1}, k_{f_2} - количество объектов соответствующего класса, и принимает значение из множества $\{\text{true}, \text{false}\}$ в соответствии со следующими условиями:

$R_{f_1 f_2}^2 (X_{i_1}^{f_1}, X_{i_2}^{f_2})$ принимает значение **true**,

если \exists объект X_m^o (экземпляр класса **SpecializedDataEntity**, $m=1, \dots, m_0$) такой, что

$\exists S(X_m^o, X_{i_1}^{f_1}) \in S'$ и $\exists S(X_m^o, X_{i_2}^{f_2}) \in S'$

иначе $R_{f_1 f_2}^2 (X_{i_1}^{f_1}, X_{i_2}^{f_2})$ принимает значение **false**.

Определим множество

$S'' = \{ R_{f_1 f_2}^2 (X_{i_1}^{f_1}, X_{i_2}^{f_2}) \text{ для таких } f_1, f_2, i_1, i_2, \text{ для которых } R_{f_1 f_2}^2 (X_{i_1}^{f_1}, X_{i_2}^{f_2}) = \text{true} \}$ - множество свойств объектов классов **Facet_f**, $f=1, \dots, n$, $k=1, \dots, k_f$

Таким образом, $S = S' \cup S''$ - множество свойств объектов представленное модели.

Множество свойств S является исходными данными для построения N-арных отношений объектов разных классов. Собственно, построение отношений и есть основная тема данного раздела. N-арные отношения (предикаты), связывающие три и более объектов разных классов **Facet_f**, $f=1, \dots, n$, определяются рекурсивно.

Определение 4. Трехместный предикат

$R_{f_1 f_2 f_3}^3 (X_{i_1}^{f_1}, X_{i_2}^{f_2}, X_{i_3}^{f_3})$, где $f_1, f_2, f_3 \in \{1, \dots, n\}$, $i_1 \in \{1, \dots, k_{f_1}\}$, $i_2 \in \{1, \dots, k_{f_2}\}$, $i_3 \in \{1, \dots, k_{f_3}\}$, $k_{f_1}, k_{f_2}, k_{f_3}$ - количество объектов соответствующего класса, выражающий отношение трех объектов, принимает значение true, если $R_{f_1 f_2}^2 (X_{i_1}^{f_1}, X_{i_2}^{f_2}) = \text{true}$ и \exists объект $X_{i_3}^{f_3}$ такой, что $R_{f_1 f_3}^2 (X_{i_1}^{f_1}, X_{i_3}^{f_3}) = \text{true}$ и

$R_{f_2 f_3}^2 (X_{i_2}^{f_2}, X_{i_3}^{f_3}) = \text{true}$,

иначе $R_{f_1 f_2 f_3}^3 (X_{i_1}^{f_1}, X_{i_2}^{f_2}, X_{i_3}^{f_3})$ принимает значение false.

Определение 5. N-местный предикат

$R_{f_1 f_2 \dots f_N}^N (X_{i_1}^{f_1}, \dots, X_{i_N}^{f_N})$, где $f_1, f_2, \dots, f_N \in \{1, \dots, n\}$, $i_1 \in \{1, \dots, k_{f_1}\}$, $i_2 \in \{1, \dots, k_{f_2}\}, \dots, i_N \in \{1, \dots, k_{f_N}\}$, $k_{f_1}, k_{f_2}, \dots, k_{f_N}$ - количество объектов соответствующего класса, выражающий отношение N объектов, принимает значение true, если $R_{f_1 \dots f_{N-1}}^{N-1} (X_{i_1}^{f_1}, \dots, X_{i_{N-1}}^{f_{N-1}}) = \text{true}$ и \exists объект $X_{i_N}^{f_N}$ такой, что для $\forall X_{i_j}^{f_j}, j \in \{1, \dots, N-1\}$ справедливо: $R_{f_j f_N}^2 (X_{i_j}^{f_j}, X_{i_N}^{f_N}) = \text{true}$

иначе $R_{f_1 f_2 \dots f_N}^N (X_{i_1}^{f_1}, \dots, X_{i_N}^{f_N})$ принимает значение false.

Таким образом, множество $R = \{ R^k | k=1, \dots, n \}$ - множество k-местных предикатов, принимающих значение true и выражающих отношения между объектами представленной модели предметной области.

Модель, базирующаяся на принципах фасетной классификации, позволяет строить связи между объектами по произвольно выбранному сценарию, причем последовательность их выбора не должна отражаться как на разработке программной поддержки, так и для организации пользовательского интерфейса. Программа поддержки взаимосвязей

объектов инварианта к выбору пути выстраивания цепочки связей.

6 Спецификация концептуальной модели

6.1 Классы объектов

Для поставленной задачи зафиксированы следующие классы объектов:

1. Список авторов
2. Автор
3. Публикация
4. Издание
5. Признак индексации в реферативных базах
- индекса цитирования
6. Тип публикации
7. Литература
8. Организация
9. Отдел
10. Должность
11. Ученая степень
12. Ученое звание
13. Направления научной деятельности ИПИ РАН
14. Тематика
15. Период времени
16. Объект связывания «Автор-Публикация»
17. Объект связывания «Публикация-Публикация»

Для задачи, рассматриваемой в данной статье, интерес представляет частный случай механизма построения связи объектов, а именно, метод выявления связей между цитирующей и цитируемой публикацией. При занесении новой публикации в базу данных в фоновом режиме запускается программа, реализующая метод построения отношений объектов. Выявленные связи сохраняются в таблицах связи и служат в дальнейшем данными для формирования статистических отчетов.

6.2 Построение связей между цитируемой и цитирующей публикациями

Рассмотрим специальное подмножество всех отношений, а именно отношения цитируемая – цитирующая публикация.

Введем обозначения:

X_m^0 - публикация с индексом m, где $m \in \{1, \dots, m_0\}$, m_0 - общее количество публикаций.

S_k^0 - свойство объекта X_m^0 , обозначает, что X_m^0 ссылается на объект X_k^0 , $k \in \{1, \dots, m_0\}$

R_m^0 - свойство объекта, обозначает, что на объект с данным свойством есть ссылка от объекта X_m^0

Аксиома. Если $S(X_m^0, C_k^0) = \text{true}$, т.е. для публикации X_k^0 определена ссылка на публикацию X_m^0 , то $S(X_k^0, R_m^0) = \text{true}$, т.е. для публикации X_k^0 существует цитирующая публикация X_m^0

7 Программная реализация

7.1 Задачи программной системы

Создание программной системы индексации публикаций сотрудников ИПИ РАН «Публикационная деятельность» позволит обеспечить полноту наполнения базы данных с учетом всех авторов и полного списка публикаций, включающего не только журнальные статьи, но и монографии, учебники, книги, патенты, доклады на конференциях.. Программная система реализуется в виде серверного приложения, которое выполняет следующие функции:

1. Поиск научных трудов:

- по названию,
- по автору, в том числе среди соавторов,
- по соотносению с направлением научных исследований ИПИ РАН,
- по соотносению с тематиками фундаментальных исследований,
- по году издания,
- по подразделению, в котором работает автор

2. Предоставление информации о публикационной активности сотрудников, включающей следующие сведения :

- полные списки научных трудов автора;
- списки цитирований научных трудов автора;
- индекс Хирша автора;
- импакт-факторы журналов, в которых публикуется автор.

3. Формирование статистических отчетов распределения научных трудов по направлениям научных исследований института, по тематикам, а также по подразделениям и годам.

7.2 Реализационная модель

Создание макета реализационной модели авторы предполагают завершить к сентябрю 2012 г. Используемые программные средства:

- среда разработки web-приложений Yii PHP framework (Yiiframework 1.1.10) [19];
- язык разработки PHP 5.3.13 [20]
- веб-сервер: Apache HTTP Server 2.2.22 [21]
- СУБД: SQLite DB 3.7.12 [22]



Рис. 1 Соотношение цитирований отечественных и зарубежных источников

7.3 Результаты статистической обработки библиографических ссылок в статьях, опубликованных в трудах конференции RCDL

Для тестирования разрабатываемой программной системы «ПД» мы использовали материалы из сборника трудов конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL за последние 4 года. Объектом интереса была статистика библиографических ссылок.

Первые две группы, на которые были поделены все ссылки на используемую литературу, составляли ссылки на отечественные и зарубежные источники. Диаграмма распределения ссылок по годам представлена на рисунке 1

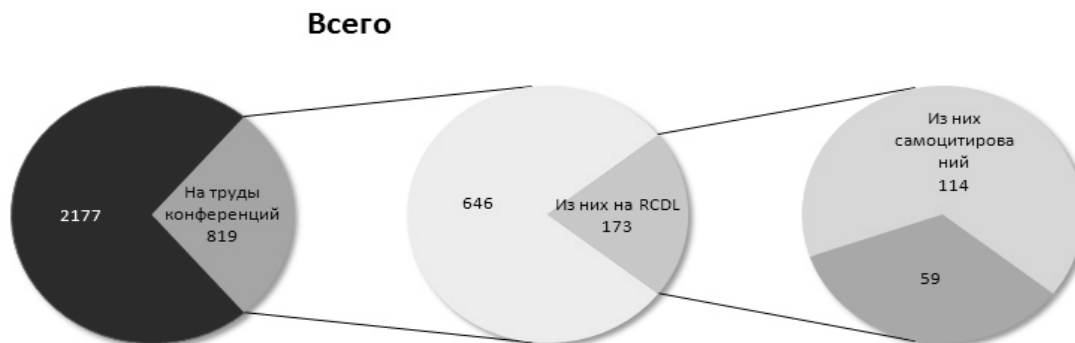
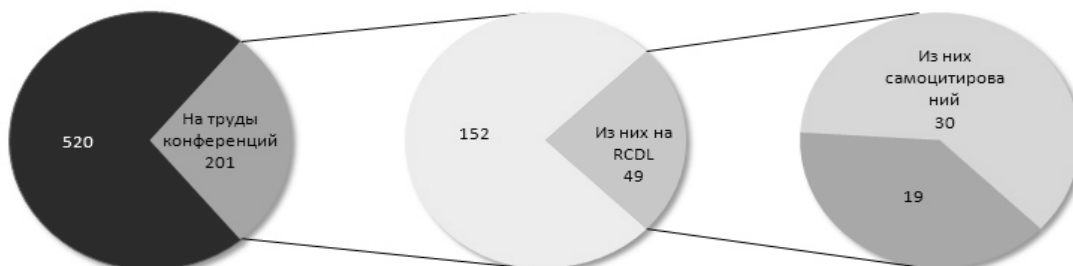


Рис. 2 Сводная статистика библиографических ссылок на материалы конференций.

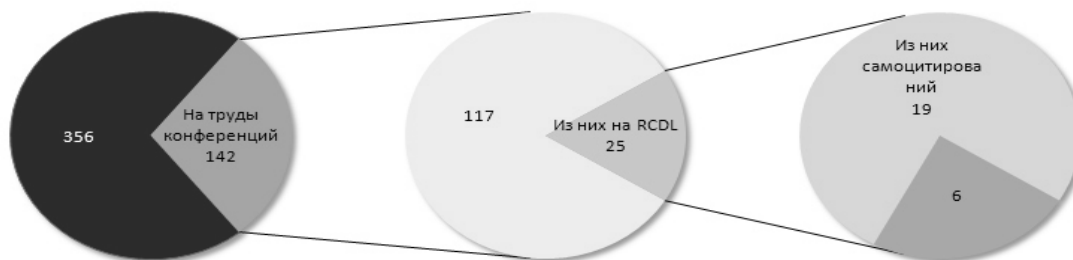
Далее были проанализированы соотношения общего количества ссылок к цитированиям материалов конференций, из которых были выделены ссылки на материалы трудов конференций RCDL. Особый интерес у нас вызвало соотношение самоцитирований и ссылок на работы других авторов, выступающих с докладами на конференции. Обобщенная по 4 годам диаграмма анализа

цитирования трудов, опубликованных в трудах конференций, представлены на рисунке 2. Далее приведены более детализированные диаграммы по каждому году для конференций, проведенных в городах Дубна, Петрозаводск, Казань и Воронеж. Соответствующие диаграммы представлены на рисунках 3-7.

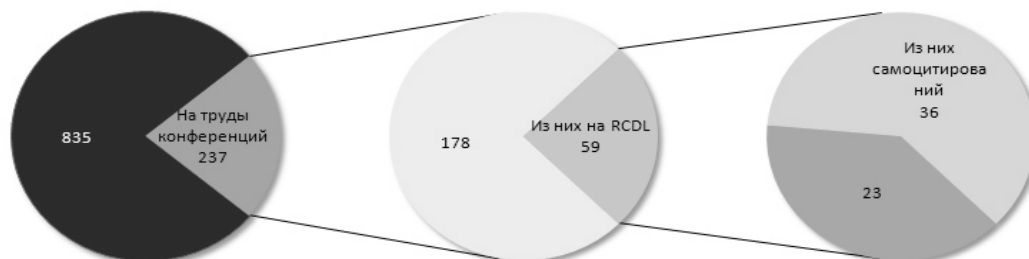
RCDL 2008: Дубна



RCDL 2009: Петрозаводск



RCDL 2010: Казань



RCDL 2011: Воронеж

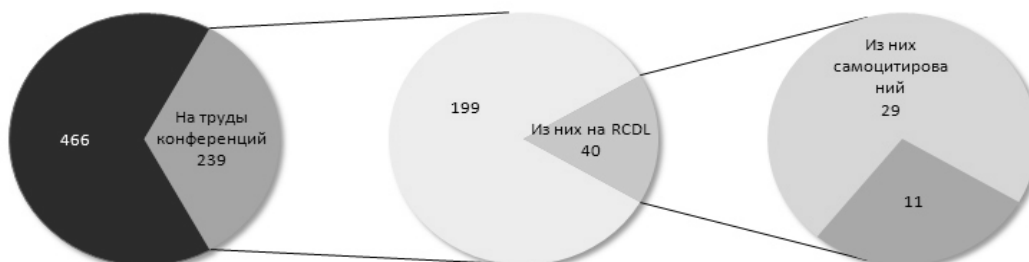


Рис. 3-7 Статистика библиографических ссылок на материалы конференций

Все цифры приведены в нижеследующей таблице.

Город	Число работ	Количество ссылок					
		всего	на зарубежные источники	на отечественные источники	на труды конференций	из них на RCDL	из них самоцитирований
RCDL2008: Дубна	55	721	355	366	201	49	30
RCDL 2009: Петрозаводск	39	498	311	187	142	25	19
RCDL 2010: Казань	80	1072	637	435	237	59	36
RCDL 2011: Воронеж	55	705	417	288	239	40	29

Заключение

Создание программной системы индексации публикаций сотрудников ИПИ РАН позволит обеспечить полноту наполнения базы данных с точки зрения учета всех авторов и полного списка публикаций, включающего не только журнальные статьи, но и доклады на конференциях, монографии, учебники, книги, патенты. Система администрирования предоставляет возможность постоянной актуализации списка работ и ссылок на них, используя ресурсы не только баз индексирования WoS и eLibrary, но и, к примеру, реферативную базу Scopus, базу данных издательство Springer [23], Общероссийский математический портал Math-Net.ru [24] и других источников. Для актуализации БД можно использовать возможности поисковых систем Google, Yandex. и т.д., но, правда, с учетом их неполной достоверности, требующей дополнительной проверки.

Программная система индексации публикаций сотрудников ИПИ РАН, с одной стороны, обеспечит сохранность и полное отражение научных трудов Института в единой базе данных, с другой стороны, предоставит необходимые сведения для более точной оценки публикационной активности каждого научного сотрудника для сверки данных, формируемых, к примеру, в отчетах РИНЦ. Дополнительно можно реализовать сервисы, целью которых является оптимизация информационного обслуживания сотрудников.

Литература

- [1] Налимов В.В., Мульченко З.М. Об использовании статистических методов при управлении развитием науки // Управление, планирование и организация научных и технических исследований. Т. 3. М.: ВИНТИ, 1970. С. 327-342
- [2] Derek J. de Solla Price. Networks of Scientific Papers. "Science", July 30, 1967, pp.510-515.
- [3] Garfield, E. Science Citation Index-A New Dimension in Indexing . "Science". May 8, 1964, Vol. 144, No. 3619, pp 649-6S4
- [4] Приказ Министерства образования и науки Российской Федерации от « 14 » октября 2009 г. N 406 «Об утверждении типового положения о комиссии по оценке Результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские и технологические работы гражданского назначения, и типовой методики оценки результативности деятельности научных организаций, выполняющих научно-исследовательские, опытно-конструкторские технологические работы гражданского назначения». Зарегистрирован в Минюст России от 28 января 2010 г. N 16115. <http://mon.gov.ru/>
- [5] Hirsch, Jorge E., (2005) «An index to quantify an individual's scientific research output». Retrieved from arXiv February 13, 2006.
- [5] Eugene Garfield.// <http://www.garfield.library.upenn.edu/>)
- [6]. Web of Science. // http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science
- [7] ISI Web of Knowledge // <http://isiwebofknowledge.com/>
- [8] Scopus. Библиографическая и реферативная база данных и инструмент для отслеживания цитируемости статей, опубликованных в научных изданиях <http://www.info.sciverse.com/scopus/>
- [9] DBLP <http://www.informatik.unitrier.de/~ley/db/>
- [10] ACM Digital Library <http://portal.acm.org/dl.cfm>
- [11] CiteSeerX <http://citeseerx.ist.psu.edu/>
- [12] Google Scholar <http://scholar.google.com/>
- [13] Российский индекс научного цитирования РИНЦ www.elibrary.ru
- [14] Обухова О.Л., Заикин М.Ю., Соловьев И.В. Предпосылки создания комплексированных индексов цитирования сотрудников научного института // Труды 13-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2011, Воронеж, Россия, 2011, стр. 207 – 215

- [15] Автоматизированная система учета результатов интеллектуальной деятельности Российской академии наук АСУ РИД <http://asurid.ru>
- [16] Обухова О.Л., Соловьев И.В., Бирюкова Т.К., Гершкович М.М., Чочиа А.П. Модель фасетного информационного поиска в коллекции научных материалов. // "Системы и средства информатики", доп. выпуск, Москва, Наука, 2009 стр. 163 – 174
- [17] Обухова О.Л., Бирюкова Т.К., Гершкович М.М., Соловьев И.В., Чочиа А.П. «Метод динамического создания связей между информационными объектами базы знаний» //Труды 11-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2009, Петрозаводск, Россия, 2009, стр. 39 – 45
- [18] R. Hull, R. King. Semantic database modelling: Survey, applications and research issues. // ACM Computing Surveys. – Vol. 19, No. 3. – 1987.
- [19] PHP framework for developing Web 2.0 applications.
<http://www.yiiframework.com/download/>
- [20] Sripting language for Web development.
<http://php.net/>
- [21] The Apache HTTP Server Project .
<https://httpd.apache.org/>
- [22] SQL database engine. SQLite.
<http://www.sqlite.org/>
- [23] Издательство Springer.
<http://www.springerlink.com/>
- [24] Общероссийский математический портал
<http://www.mathnet.ru/>

The Technology of Aggregated Science Citation Indexes Formation

Olga Obuhova, Maxim Zaikin, Ivan Soloviev

The article analyzes the requirements for the indexes defined by the Order of Ministry of education and Science No. 406 on October 14, 2009 according to which the effectiveness of scientific institutes should be evaluated.

The article also describes new technology of collection of data on scientific researchers publishing activity which was developed as the basis of IPI RAN's data retrieval system.

The statistics of citation in the proceeding of the Conference "Digital libraries: advanced method and technologies, digital collections" (RCDL) is given as one of the results of the technology application.