



ПОЛИТ.РУ



Владимир Писляков
Фото: Наташа Четверикова

Наука

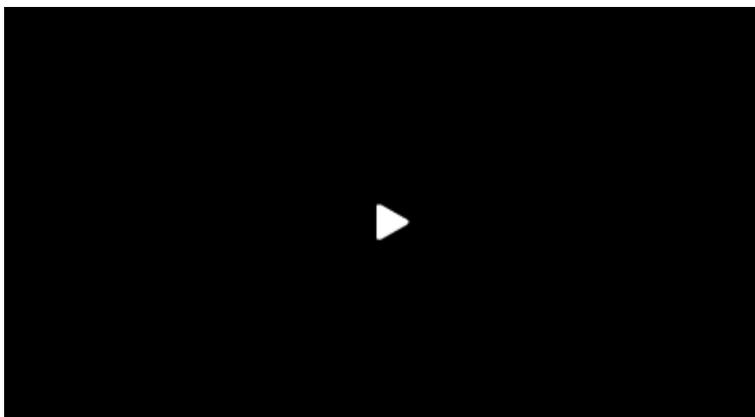
Наука через призму статей

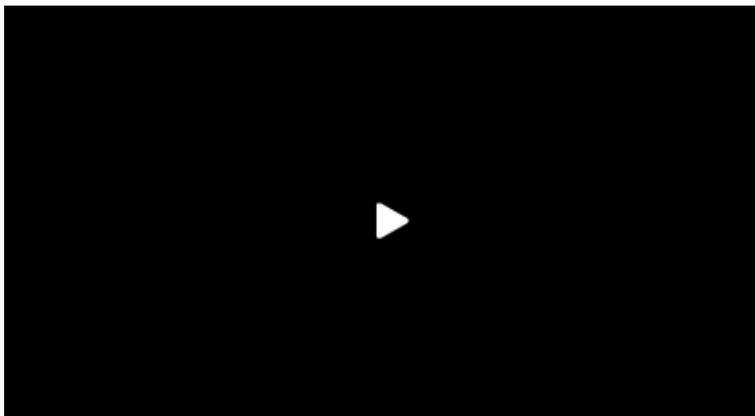
21 декабря 2011, 10:29

Адрес → http://polit.ru/article/2011/12/21/pislyakov_2011/

*Мы публикуем текст лекции кандидата физико-математических наук, заместителя директора по управлению электронными ресурсами Библиотеки Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ) **Владимира Владимировича Пислякова** "Высокоцитируемые статьи: что это такое и кто их пишет в России", прочитанной 30 июня 2011 года в Политехническом музее в рамках проекта «Публичные лекции Полит.ру».*

«Публичные лекции "Полит.ру"» проводятся при поддержке:





Текст лекции



Владимир Писляков
(фото Наташи Четвериковой)

Здравствуйте, коллеги. Я рад видеть вас здесь, причем в таком большом количестве: все-таки конец июня. Мы завершаем сезон лекций 2010-2011 года.

Меня зовут Владимир Писляков, я библиотекарь и работаю в библиотеке Высшей школы экономики. Помимо этого, как часто бывает с библиотекарями, занимаюсь рядом библиометрических исследований по России, по миру, по месту России в мире. Сегодня мы с вами поговорим о «научных шедеврах» среди журнальных публикаций, о так называемых высокоцитируемых статьях. У лекции будут две части. В первой части, достаточно обширной, я расскажу о том, что такое высокоцитируемая статья, какие инструменты используются в библиометрии, как правильно использовать термины и что каждый термин означает. Я прошу прощения у тех, кто более-менее профессионально занимается библиометрией, потому что все это будет вам известно. Но в связи с тем что сегодня я выступаю перед достаточно широкой аудиторией, я эту часть освещу практически с нуля. Как мне кажется, это будет половина нашей с вами лекции, после нее я сделаю паузу и вы зададите вопросы по вводной части. Потом мы перейдем к результатам того исследования, которое ведется Высшей школой экономики, в этой части тоже будет два раздела, которые, однако, пойдут слитно. После этого будет время для финальных вопросов. Такой план.

Итак, терминология, определения и инструменты, которые используются в библиометрии, с плавным подводом к термину «высокоцитируемая статья». Начнем с терминологии, где встречается немало достаточно двусмысленных терминов. Дело в том, что библиометрия, по сравнению с Европой, с Соединенными штатами, с Канадой, у нас в России не очень распространенная наука. У нас не так много людей

этим занимается, соответственно есть сложности с определениями. Первое из них вы сейчас видите, это разница между словами «цитирование» и «ссылка». На самом деле, даже когда говорят «высокоцитируемая статья», в этом термине есть некоторая неточность, потому что речь идет не о цитировании, а о ссылке на статью. Я долго думал «неужели кто-то не понимает? что тут сложного? неужели непонятно, что цитирование и есть ссылка?», пока на занятиях с преподавателями Высшей школы экономики мне регулярно не стали задавать вопрос: «Мы нашли, что эта

статья цитирует другую, а где сама выдержка из статьи, где цитата?». Тогда я понял, что это разделение надо выдерживать, что правильно все время говорить о ссылках. То есть когда говорят «одна статья цитирует другую», это всего лишь означает, что в списке цитируемой литературы первой статьи содержится вторая статья. Тем не менее, по терминологической скудости, я буду в качестве синонимов использовать слова «ссылка» и «цитирование» — после такой вводной ремарки, что на самом деле это не совсем правильно.

Далее. Самое известное словосочетание, термин из мира библиометрии. Термин, о котором, как о политике и как о футболе, говорят все, как только заходит речь о каких-то библиометрических исследованиях. Это «индекс цитирования». Проблема в том, что никакого такого понятия нет. Когда, скажем, из руководящих органов Российской Федерации приходит распоряжение посчитать «индекс цитирования вуза», или какого-то ученого, будет корректно спросить: «А что имеется в виду? за какой отрезок времени это считается? считается суммарный показатель или это показатель в расчете на одну статью?..». Исходно это пошло от того, что произошла некорректная метонимия и термин «citation index», который в английском языке означает исключительно базу данных, стали воспринимать как показатель. «Citation index» по-английски это только база данных, вы не найдете ни одного показателя, который бы назывался «citation index». И если возвращаться к предыдущему пункту, то базу данных, когда мы говорим «citation index», совсем уж правильно и по-русски переводить как «указатель ссылок». Указатель ссылок в том значении, как вы в конце книги находите указатель по словам или по именам — где, на какой странице находится тот или иной ученый, упоминавшийся в книге.

Чаще всего под «индексом цитирования», особенно ученого, имеют в виду суммарное число полученных ссылок, по умолчанию чаще всего так. Но не совсем понятно, почему: все-таки мы как правило не называем индексом суммарный показатель, чаще всего это какое-то отношение. Впрочем, это возражение более спорное: скажем, *h*-индекс, индекс Хирша, который употребляется в библиометрии, — общепринятое обозначение, хотя это фактически подсчет числа статей с определенными свойствами. Поэтому я предлагаю не использовать термин «индекс цитирования». А говорить мы будем, в зависимости от того что оцениваем, о цитируемости ученого или о цитируемости статьи, — то, что в английском называется *citedness*. Мы будем говорить «суммарное число ссылок», сколько ссылок получила статья. Когда же мы говорим о журналах, то самый распространенный показатель — это «импакт-фактор», цитируемость в расчете на одну статью с некоторыми определенным образом заданными хронологическими промежутками, учетом времени выхода статьи и времени, когда считаются цитирования. К импакт-фактору мы сегодня еще вернемся.

Теперь об инструментах, которые используются в библиометрии и, в частности, в том исследовании, о котором пойдет речь во второй части лекции. Есть две международно признанные базы данных цитирования — это база данных Scopus компании Elsevier, молодая база, которая возникла в 2004 году, и та линейка инструментов, которые мы использовали в своем исследовании. Это более древний продукт, с более глубокой историей, — линейка продуктов компании Thomson Reuters. Исходно это был Институт научной информации, основанный Гарфилдом в Филадельфии, который потом был куплен компанией Thomson. Затем компания Thomson купила новостное агентство Reuters, и сейчас она называется Thomson Reuters. Ей принадлежат эти продукты.

Что это за продукты? Прежде всего, это исходные данные в продукте Web of Science, на следующем слайде будет подробно о нем написано. Это база данных цитирования, «указатель ссылок», если говорить правильно. Это citation index, «индекс цитирования» в значении «база данных цитирования». Она состоит из нескольких частей. Есть часть по естественным, точным наукам и медицине — база Science Citation Index Expanded. Есть база данных по общественным наукам — Social Sciences Citation Index. Есть база данных по гуманитарным наукам — Arts &

Humanities Citation Index. Есть дополнения по трудам конференций — Conference Proceedings Citation Index. Всё это исходные данные. И есть ряд надстроек над базой данных Web of Science — это агрегированные, собранные данные по той или иной «сущности» из исходной базы. Например, база данных Journal Citation Reports, которая публикует импакт-факторы журналов, — она собирает данные по цитируемости журналов, то есть для нее «сущности» — журналы. База данных Essential Science Indicators также является агрегирующей надстройкой над Web of Science, она собирает данные по журналам, ученым, странам и по научным организациям.

Немного подробнее: исходные данные базы Web of Science — 4 блока, о которых я вам вкратце рассказал, а здесь я привел число журналов, которые расписываются в каждой из этих баз и число отечественных журналов среди них. Как раз два года назад произошло серьезное пополнение базы данных Web of Science отечественными журналами, было добавлено еще 33 журнала. Любопытно, что все эти журналы попали в первую базу Science Citation Index, то есть ни одного общественно-научного или гуманитарного журнала добавлено не было.

Social Sciences Citation Index — 2500 журналов и всего 3 отечественных. Это «Психологический журнал», «Вопросы психологии» и «Социологические исследования» — это все общественно-научные журналы из России, которые есть в Web of Science. Далее, гуманитарная часть, но о ней мы сегодня не будем говорить по некоторым причинам. Там поменьше журналов. И база по трудам конференций.

Что это за базы, что они собирают? Они берут каждый журнал, скажем, из 7 тысяч, которые входят в первую базу, берется каждая статья из этого журнала и в базу данных Web of Science помещаются:

— выходные данные статьи (автор, название статьи, журнал, том, выпуск, страницы...);

— аннотация статьи, если она была в исходном журнале. В случае с естественнонаучными журналами в подавляющем большинстве случаев это так, она есть, и аннотация помещается в базу данных;

— ключевые слова, причем в двух изводах: ключевые слова, приписанные самим автором, и ключевые слова (они называются keywords plus), которые навязала сама база данных, присвоила этой статье;

— адреса работы авторов (affiliations), место их работы, для нас это будет важно;

— и, что делает просто реферативную базу данных базой данных цитирования, — помещается список цитируемой литературы для каждой статьи.

В Web of Science единица исследования, с которой вы работаете, — это одна статья, с теми данными из нее, которые я перечислил. Полного текста нет.

Essential Science Indicators — при помощи этой базы я буду сегодня рассказывать, как исследуются высокоцитируемые статьи, поэтому о ней тоже подробнее стоит сказать. Essential Science Indicators — это десятилетняя база данных. Она собирает данные по последним десяти годам, плюс текущий год, «ten rolling years» — так это там называется. И сейчас в ней данные с 2001 года по, кажется, май^[1] текущего года. В Essential Science Indicators все публикации разделены на 22 тематические области. Это для нас тоже очень важно, потому что мы будем говорить об этих областях.

Как статья попадает в ту или иную тематическую область? Это широкие области: если вся наука разделена на 22 области, это означает, что каждая область будет достаточно широкой. Например, целая область «физика», целая область «химия», без деления; «биология/биохимия» совместно — это одна рубрика, «науки о

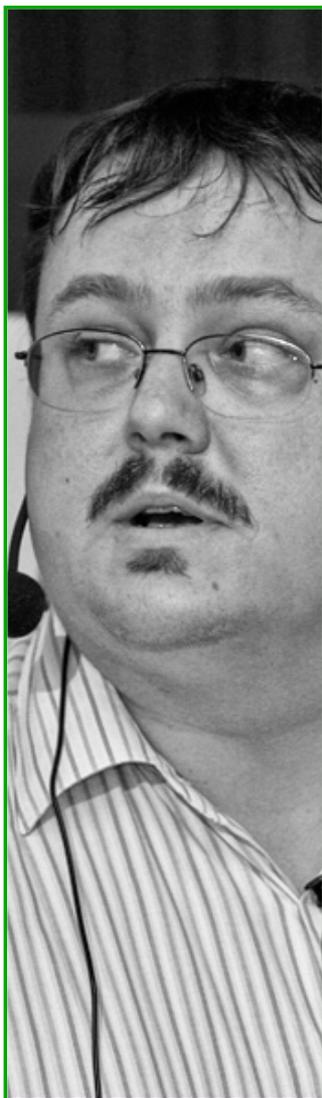
Земле»... 21 тематическая область и еще одна, 22-я, «multidisciplinary» — мультидисциплинарные статьи. Как статья попадает в ту или иную область? К сожалению, и с этим ничего не поделаешь, статье рубрика присваивается на основании того журнала, где она опубликована. То есть по этим рубрикам классифицируются журналы, а дальше каждая статья, опубликованная в журнале, попадает в соответствующую область журнала. Исключение составляет как раз рубрика «multidisciplinary», и это интересно, ведь там одни из важнейших статей, потому что там Nature и Science, например. В Nature и Science может быть опубликована статья по любой тематике. Вот там база данных прodelывает дополнительную работу и раскидывает практически все статьи по тематическим рубрикам. Как она это делает? Она исследует списки цитируемой литературы: кого цитирует та или иная статья. И по этим спискам, в зависимости от того, из какой области цитируются журналы, идет приписка статьи к той или иной рубрике. С самыми популярными журналами Nature и Science такая работа проводится, как и с остальными журналами, их немного в разделе «multidisciplinary».

Так устроена система классификации. Обратите внимание, что в Essential Science Indicators нет гуманитарной части — гуманитарных журналов, гуманитарных статей там просто нет. Поэтому о гуманитарной составляющей мы сегодня говорить не будем. Почему так? Дело в том, что и в других базах, за исключением Web of Science, сторонятся считать показатели цитируемости для гуманитарных журналов, потому что в этой области, во-первых, очень много значат книги; во-вторых, показатели цитируемости не являются таким серьезным индикатором значимости статей, как, в первую очередь, в естественных, точных науках и, в чуть меньшей степени, в науках общественных.

В базу данных Essential Science Indicators попадают статьи только типов Article и Review. Article — это научная, исследовательская статья, а Review — это обзорная статья. Здесь небольшая тонкость: на самом деле еще там есть Proceedings Paper — это тоже статья, которая опубликована в журнале, но просто она родилась из доклада на конференции. Это дополнительный раздел, и я его включил в Article, они ничем не отличаются. Различные письма в редакцию... — это все в Essential Science Indicators не включается и эти статьи считаются вторичными для оценки тех сущностей, которые оценивает Essential Science Indicators: журналы, организации, авторы и страны.

Итак, в Essential Science Indicators есть 4 раздела: по авторам, организациям, странам и журналам. По ним собирается публикационная активность, сколько тот или иной автор опубликовал статей в таком-то году, или сколько страна получила за определенный период ссылок на свои публикации. То есть это публикационные показатели и показатели цитируемости. Кроме этого, отдельным разделом в этой базе есть «Highly cited papers» — то, о чем мы будем с вами говорить, «высокоцитируемые статьи». Мы сейчас перейдем к определению высокоцитируемых статей. Этот раздел существует отдельно от остальной базы данных и в нем собираются, скажем так, «научные шедевры». Еще есть раздел «Hot papers», это то же самое, что «highly cited», только там промежуток, за который смотрятся статьи, — последние два года. То есть это статьи, которые не просто получили много ссылок, но получили их быстро. Тема их исследования популярна в настоящий момент. Есть также данные по всей науке, без деления на страны, это раздел «Baselines». Там показатели, например, «сколько во всем мире за какое-то время было опубликовано статей по физике и сколько эти статьи получили ссылок».

Перейдем к определению высокоцитируемой статьи. Что это такое, как «highly cited paper» можно было бы определить и как ее определяет компания Thomson в своем продукте? Для того чтобы понять, что статья высокоцитируемая, что она является «шедевром», ее нужно сравнить на фоне таких же статей как она. Потому что нельзя сравнивать статьи, вышедшие, скажем, 10 лет назад и вышедшие 2 года назад, поскольку одна из статей дольше существовала. Их цитируемость сравнивать нельзя, потому что одна статья могла за счет своего возраста получить больше



Владимир Писляков
(фото Наташи Четвериковой)

ссылок. Статью надо сравнивать внутри того окружения, которое максимально похоже на нее. Что это в данном случае значит? Надо фиксировать две вещи. Это то, что я уже привел в качестве примера, год выхода статьи. Статью надо сравнивать по цитируемости со статьями, которые вышли в тот же самый год, что и она. И второе, статьи должны принадлежать к одной и той же тематической области, хотя бы в рамках широких 22 категорий, о которых мы говорили. Почему это важно? Дело в том, что в различных областях уровень активности цитирования разный. Я сейчас вам покажу сравнение импакт-факторов из разных тематических категорий. Я в данном случае отвлекся: мы изучаем отдельные статьи, а я вам покажу данные по журналу. Но это как раз импакт-фактор, цитируемость средней статьи, и по разнице этих значений мы увидим вообще уровни цитирования в различных научных областях. Это медианные импакт-факторы журналов: до черты — это естественнонаучная база, тематика «STM», «science, technology, medicine», то, что берется из Science Citation Index Expanded, а после черты — это общественные науки. Посмотрите, как различаются медианные импакт-факторы, — если не на порядки, то в разы. Если вы занимаетесь биологией клетки и вы опубликовались в журнале с импакт-фактором 2, то вы попали в третью квартиль, в журнал так себе, ни то, ни се. А если вы математик и опубликовались в журнале с импакт-фактором 2, то это будет журнал, который будет входить в 10 самых цитируемых журналов по математике.

Почему такое распределение? Означает ли это, что цитология гораздо лучше математики? Нет, конечно. Просто сложились разные, как говорят, «культуры цитирования». В реальности это означает среднюю длину списка цитируемой литературы в каждой области и, кроме того, насколько важны книжные источники, а насколько — журнальные. Потому что если книжные

источники для вашей науки важны, то ссылки на них уходят вовне базы и они, таким образом, не поднимают средний уровень цитируемости внутри области. Кроме этого, играет роль еще один показатель, который тоже очень важен. Это показатель хронологии распределения ссылок, потому что импакт-фактор учитывает ссылки на статьи, которые вышли за два предыдущих года. И если у вас в области в основном ссылаются на статьи пятилетней давности, то понятно, что средние импакты там будут меньше, чем в случае, скажем, медицины, где наоборот, все ссылки валятся в первые два года. Такие факторы влияют на это, а нам важно понять, что дисциплины разные и сравнивать статьи из разных дисциплин, для того чтобы определить, какая статья высокоцитируемая, невозможно.

После черты мы видим, что, например, психология сильно отличается от истории, в несколько раз. Хочу обратить ваше внимание, что данные приведены за 2010 год, эти импакт-факторы появились в ночь с позавчера на вчера, так что вы получаете самую свежую информацию. Вышло новое издание: ежегодно, в июне, Thomson Reuters публикует новые импакты. А главное что там? Главное, что Nature опять победил Science по импакт-фактору. Так уже два года подряд, а до этого Science вел несколько лет[2].

Посмотрите хронологию, исследование которой провели Гленцель с Мутом[3], это венгерский и голландский ученые. Они взяли журнал по социологии «American Sociological Review» и «Lancet», журнал по медицине. Взяли статьи, которые были

опубликованы в 1980 году и стали отходить от 1980 года к 1990 году и смотреть, сколько в среднем ссылок на статью получил один журнал и другой. Подчеркиваю, для одних и тех же статей, которые вышли в 1980 году. Если взять первый год, то среднее число ссылок на статью в три раза больше в медицинском журнале по сравнению с социологическим, если отойти на 2-3 года, показатели практически одинаковые, если посмотреть с десятилетнего расстояния, то окажется, что в среднем социологическая статья получила ссылок больше, чем медицинская. Здесь мы видим, насколько по-разному устроены области науки с точки зрения временного распределения ссылок.

Таким образом, для того чтобы найти высокоцитируемые статьи, мы должны зафиксировать год выхода статей и тематику, научную область, в которой они опубликованы. Взять все статьи, которые вышли в этом году в этой тематике и дальше каким-то образом отрезать верхушку.

Еще вопрос: а всё ли это, во всё ли мы уравнили статьи? Надо честно сказать, что не совсем, поскольку, например, известно, что Article (научная статья) и Review (обзор) получают разное количество ссылок. Обзорная статья, как правило, получает больше ссылок, чем обычная научно-исследовательская статья. Этого разделения в базе данных Essential Science Indicators нет, мы видели, что Articles и Reviews там вместе. Есть еще тонкости — скажем, статья теоретической направленности обычно получает больше ссылок, чем статья, описывающая эксперимент. Но это более слабые детали. Основное и самое главное зафиксировано.

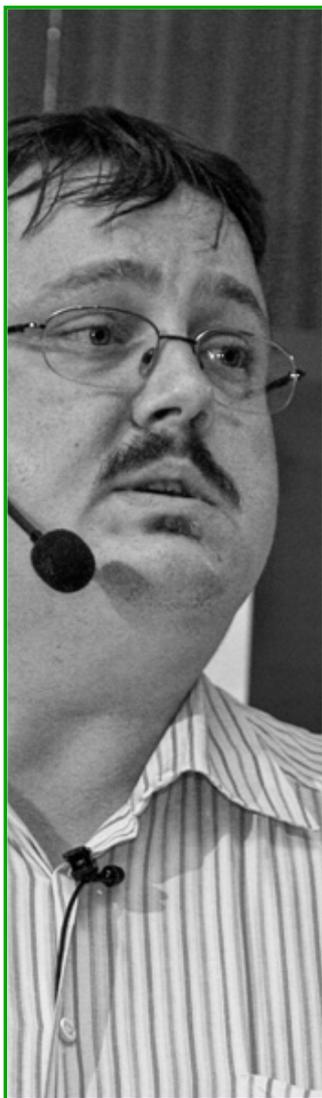
Вот это фрагмент раздела Baselines в базе Essential Science Indicators. Теперь можно понять, сколько ссылок должна получить статья, вышедшая в таком-то году и в такой-то области, для того чтобы попасть в некий верхний срез. Сколько она должна получить ссылок, чтобы попасть в 0,01% самых цитируемых статей, или в 1% самых цитируемых статей, или просто — сколько она должна получить, чтобы быть на среднем уровне в своей области и для своего года выхода. В разделе «Highly Cited Papers» базы данных Essential Science Indicators считается, что высокоцитируемая статья — это та статья, которая при фиксированном годе выхода и тематике попала в 1% наиболее цитируемых статей. Поэтому нам интересны вот эти две строчки. Обратите внимание: это статьи по сельскому хозяйству, сельскохозяйственные журналы, а это — биология и биохимия. Статьям, вышедшим в 2003 году, чтобы стать высокоцитируемыми (то есть попасть в верхний 1%), в сельском хозяйстве надо набрать 85 ссылок, а в биологии и биохимии — 187. Это та самая разница в цитируемости. В биологии гораздо больше уровень цитирования, поэтому надо получить больше ста ссылок, для того чтобы попасть в верхний 1% самых хороших статей.

На этом закончилась первая, вводная часть. Мы определили, что такое высокоцитируемая статья, и если есть вопросы по этой части, я прошу вас их задать, чтобы потом мы спокойно двигались и к highly cited papers отсылали как к чему-то нам хорошо известному и понятному.

Обсуждение лекции

Борис Долгин: я начну с вопроса, который был задан заранее. У нас к анонсам лекции читатели могут задавать вопросы, и я думаю, что это скорее к первой части, чем ко второй. Как при оценке цитируемости статьи вычленяется «негативное» и «позитивное» цитирование? Приведен пример, скажем, «в своей статье Петрик пренебрегает законами физики и демонстрирует качество псевдоученого».

Владимир Писляков: хороший пример. Разумеется, никак. То есть этого разделения в базах данных нет. Я объясню почему, с технической точки зрения. Потому что для того чтобы отсоединить такого рода ссылки, надо читать все 12



Владимир Писляков
(фото Наташи Четвериковой)

тысяч журналов, которые есть. Компания Thomson на это неспособна, а если бы была способна, то она бы продавала свой продукт по такой цене, что мы бы его не увидели. Он и так достаточно дорогой.

Вообще такого массивированного эффекта здесь нет. Можно находить какие-то хорошие, любопытные, забавные примеры, но если мы оцениваем показатели, например, «как страна выглядит в мире», разумеется, этих эффектов нет. Не столько Петриков здесь, чтобы....

Борис Долгин: на уровне страны — да, на уровне автора это вполне может сыграть.

Владимир Писляков: на уровне автора да, но это будут такие эффекты, что их можно будет разобрать, почему этот человек получил столько ссылок. Сейчас немного не про это, но подобный эффект в единичных случаях бывает. Сейчас по импакт-фактору первый в мире журнал «A Cancer Journal for Clinicians», у него запредельный импакт 90. А на втором месте два года идет журнал по кристаллографии, кажется «Acta Crystallographica A», второй в мире импакт-фактор 53. Если посмотреть на то, как менялся его импакт-фактор, мы увидим, что он все время был в районе 1–2, а два последних года он очень высокий. Оказывается, это цитируют одну статью! Благодаря большой цитируемости одной статьи весь журнал получил такого рода преференцию. Сейчас пройдут два года, в импакт она перестанет попадать, и у него снова упадет импакт-фактор. Конечно, существуют какие-то дополнительные средства, например тот же индекс Хирша, который отказывается считать средние значения, а требует, чтобы было много статей, которые много цитируют. Тут уже вряд ли такого рода эффекты возможны. Но в нашем случае не так, в базе Essential Science Indicators мы будем считать средние показатели.

Борис Долгин: еще вопрос, почему такие грубые классификации. Почему тематика идет по тематике журнала, а не делается более или менее универсальный классификатор с пожеланием, если журналы или авторы хотят входить в индексы, чтобы они просто размечали свои статьи подобным образом и это будет гораздо более точный инструмент, мне кажется. Систематизация на уровне статей.

Владимир Писляков: здесь следующая проблема, я думаю. Я не могу отвечать за компанию Thomson, но у меня есть предположение, что они боятся это дело выпустить из своих рук. Как только мы отдаем на аутсорсинг какой-то действие с моей базой данных, которую я собираю, то я не могу за нее всецело отвечать. Поэтому такого рода классификация там не принята. Надо сказать, что сейчас в библиометрии есть инструменты, которые приближаются к такому — не к привлечению авторов, но к классификации статей на уровне статей. В базе данных Scopus один из принятых недавно показателей — SNIP, «source normalized impact per paper». Он как раз пытается определить «окружение» журнала, потому что для журнала считается SNIP на основе конкретных статей, а не журналов. Определяются статьи, которые хотя бы раз в течение 10 лет цитировали этот журнал. Соответственно, наверное, они из той же области. Но все это делается в автоматическом режиме, просматривать вручную или отдавать вовне, я думаю, на это не решатся, это все-таки не википедия...

Кстати, почему 22 рубрики, поделить же мы можем журналы помельче. В исходной базе данных Web of Science сделано более тонкое деление, там этих рубрик по одним естественным наукам 150, плюс еще по общественным в районе 70. Но там встает проблема, что многим журналам приписывается несколько мелких рубрик. Это более точно их характеризует в каком-то смысле, но тогда у нас нет однозначности. Здесь каждый журнал попадает в одну из широких 22 рубрик.

Борис Долгин: а почему такой странный состав, в какой степени, если они не готовы выпускать из рук, в какой степени они отвечают за такой странный подбор отечественной гуманитарной и социальной научной периодики. Они готовы отвечать за такой подбор, они осознают, что это несколько спорный и не представительный подбор?

Владимир Писляков: что касается общественных наук (с гуманитарными я меньше сталкивался, а эти три журнала знаю наизусть), конечно, это плохой подбор. Во-первых, он нерепрезентативный, во-вторых, там можно заменить журналы. Даже если они хотят, чтобы было три журнала, то пусть это будут другие три. Я не буду сейчас говорить, о каком журнале речь, но можно что-то заменить на «Вопросы экономики». Буквально сейчас в «Экономическом журнале Высшей школы экономики» выйдет заметка, мы рассчитали, какой был бы импакт-фактор у «Вопросов экономики» (это сейчас лидирующий журнал по цитируемости в общественных науках у нас в России, это в базе данных Российского индекса научного цитирования видно хорошо), какой будет его импакт-фактор. И оказалось...

Борис Долгин: ну, к РИНЦ есть свои претензии, скажем честно.

Владимир Писляков: да, да... Так или иначе, импакт-фактор «Вопросов экономики» получается 0,25. При этом он будет выше, чем импакт-фактор всех трех журналов, но все равно это, конечно, слабый уровень. Я хочу сказать, что с общественно-научными журналами это во многом и наша вина, потому что мы не подаем журналы. «Вопросы экономики» сейчас удовлетворяют уже всем требованиям Web of Science, но я не знаю, подавали они или нет. У меня в редакции есть знакомые, но и они не уверены.

Даниил Наумов: скажите, почему учитывается страна, но не язык.

Владимир Писляков: в Web of Science есть поле «язык». Но ведь есть масса журналов, где публикуются авторы из России на английском языке. У нас половина статей идет на английском языке, половина на русском — из тех, которые присутствуют в Web of Science. Таким образом, мы бы теряли половину статей. Видимо в этом дело, учитывается страна. Как раз на следующем слайде я хотел об этом сказать, но давайте сейчас скажу. Разумеется, страновая принадлежность определяется по местам работы авторов, то есть если в адресе места работы есть слово «Russia», статью мы будем относить — и база данных Essential Science Indicators относит — к России. Если нет, то нет.

Голос из зала: скажите, сейчас ситуация такова, что если статья А цитирует статью Б и в теле статьи один раз ее упоминает, то ее цитирование увеличивается на единицу. А если статья целиком посвящена разбору этой статьи, она там цитируется 30 раз, все равно увеличивается на единицу. Есть ли в библиометрии такая перспектива или намерение библиометристов сделать так, чтобы считался не только факт упоминания, а количество.

Владимир Писляков: я боюсь соврать, но мне одна из таких баз данных встречалась, какая-то из национальных баз данных цитирования. По-моему, это тайваньская база по гуманитарным журналам, у них есть своя, как и у нас есть своя. Там меня удивило, что считают количество включений ссылки на соответствующую публикацию в тело статьи. Я не готов ответить точно, мне надо проверить^[4]. Но в

целом — расписываются, включаются в индексы списки цитируемой литературы, а в противном случае пришлось бы работать с полным текстом.

Константин Иванович: хотел бы задать такой вопрос, рассматриваются ли пути, которые потом станут мошенническими, есть ли такие уже сейчас и в будущем, в перспективе. Спасибо.

Борис Долгин: в каком смысле мошенническими, не понял.

Константин Иванович: очень просто. Создается какая-то команда, которая будет цитировать друг друга по кругу.

Борис Долгин: понятно.

Владимир Писляков: да, понятно, и понятно, что этому вопросу десятилетия. И здесь можно лишь повторить то, что я говорил: это можно делать в каких-то единичных случаях, но массированно это сделать нельзя. Есть яркий пример, это показатели самоцитируемости журналов. Журнал может начать сам себя побольше цитировать, и у него должен расти импакт-фактор. Когда мы проводили такое исследование по нашим отечественным журналам, мы выстраивали журналы в порядке импакт-фактора с учетом самоцитируемости (а в импакт-факторе, который считает Thomson, самоцитируемость учитывается, не исключается), а после этого вручную выкидывали самоцитаты и смотрели, как журналы располагаются после этого, в каком порядке. И оказалось, что общий принцип такой: 80-й журнал с 90-м могут поменяться местами, но 1-й со 2-м — никогда. То есть невозможно за счет этого из грязи в князи, можно так, уровень грязи немного сбросить. Ну начнут они цитировать друг друга, ну перейдут с 90-го места на 80-е... Кроме этого, сейчас такого рода механизмы тоже внедряются, потому что и в Thomson'e считается показатель Eigenfactor и в базе данных Scopus считается показатель SJR, SCImago Journal Rank — это показатели для журналов, но в принципе то же самое можно применять и для авторов, — где рассматривается вся сеть цитирования, и блоки, которые находятся сами по себе, у них сразу этот показатель начинает падать. Там считается взвешенный показатель, то есть учитывается авторитетность того источника, который процитировал соответствующий журнал, и это получается анализ всей сети и такого рода эффекты там нивелируются.

Борис Долгин: в продолжение последнего вопроса я бы спросил не о мошенничестве, а о влиянии наличия тех закономерностей рейтингования, о которых вы говорили, на структуру тематической, жанровой организации статей. Не исследовалось ли, каким образом именно эти способы подсчета и те закономерности, то о чем вы говорили, что обзоры цитируются больше чем статьи, что теоретические больше чем... и так далее. В какой степени это смещает общее поле.

Владимир Писляков: я не готов ответить на это в целом. Единственное: ведь почему обзорные статьи цитируются чаще? В первую очередь потому что у них самих большой список литературы. И работа над списками цитируемой литературы, конечно, ведется, и, кстати, хорошо, что она ведется. По крайней мере, в случае наших журналов по общественным наукам сейчас в некоторых изданиях есть такая установка: давайте больше цитировать. Потому что если уровень, культура цитирования к нам придет, культура ссылок на коллег, на работы друг друга, то это приведет и к возрастанию показателей и, соответственно, ценности для включения в базу данных Thomson'a, где это учитывается при решении о включении журнала — решение принимается в том числе и на основании цитируемости.

Коростылев: скажите пожалуйста, какое влияние оказывает факт наличия того или иного индекса, точнее ссылок, на саму науку и на действия самого ученого. У одного больше, у другого меньше, какое это оказывает влияние на само научное исследование.

Борис Долгин: как это используют при управлении наукой.

Коростылев: да, зарплата там больше, как?.. Вот в Высшей школе экономики, я знаю, от этого зависит зарплата конкретного ученого.

Владимир Писляков: Вы знаете, нет, у нас не так и, насколько мне известно, к счастью в широкое использование такого рода механизмы не пошли еще. Наверное пока. Но пока надо избавиться от формулировок «индекс цитирования» в указивках, присылаемых в организации. Вот когда это научатся делать, тогда наверное можно будет. Что касается Высшей школы экономики, у нас учитываются публикации, но не их цитируемость. У нас есть программа академических надбавок, которая учитывает именно публикации, в том числе там есть раздел надбавок за публикации в журналах из этих баз данных. Но сам факт цитируемости у нас не рассчитывается, по отдельным людям. А по университету да, потому что это тоже надо поставлять.

Борис Долгин: спасибо. Я думаю, может, мы ко второй части перейдем.

Текст лекции. Часть 2

Владимир Писляков: Давайте продолжим. Сейчас мы с вами поговорим об общих количественных показателях высокоцитируемых статей по России. Без персоналий и без мест работы. Просто: как Россия выглядит в целом и как Россия выглядит в отдельных дисциплинарных областях. А потом мы более пристально посмотрим на конкретные организации, в первую очередь, ну и на людей тоже.

Я говорил, что база данных рассматривает 10 последних лет, плюс текущий год. Сбор данных проводился летом и осенью прошлого года, и мы решили ограничиться только теми статьями, которые выходили с 2003 по 2007 год. Во-первых, 5 лет это такой принятый в библиометрии промежуток времени, он часто встречается, в том числе и в этих базах: показатели считаются пятилетними кусками — скажем, по странам. Кроме того, мы обрезали два хвоста. Один хвост — это старые статьи, которые не очень показывают что у нас сейчас творится. Статьи, вышедшие с 2000 по 2002 год, мы не стали брать как достаточно древние статьи. И мы избавились от самых свежих статей, по той причине что если мы рассматриваем последние годы, то критерий для включения в высокоцитируемые статьи, этот порог, получается очень маленький. Для сельского хозяйства, например, — а есть даже менее цитируемые области, — сейчас для того чтобы стать высокоцитируемой, статье, которая вышла в 2010 году в сельском хозяйстве, достаточно набрать 4 сссылки. Это может быть какой-то случайный эффект, в том числе и эффект какой-то совместной «работы» над своим показателем. Поэтому мы хотели взять те годы, где уже достаточно большие пороги и можно корректно рассматривать высокоцитируемые статьи. Поэтому промежуток 2003–2007. Это будет всегда: когда я буду говорить о процентах, о показателях разных стран, это все будут статьи с 2003 по 2007 год выхода.

Всего в мире высокоцитируемых статей, вышедших в эти годы, 47 тысяч, в России из них 511 публикаций, это получается чуть больше 1%. Для сравнения, например: больше чем каждая вторая высокоцитируемая статья имеет автора из США, из Германии — 11%, Франции — 7%. Почему в этом порядке (и в таком же порядке эти страны будут встречаться дальше)? Здесь мы сравнивали показатели тех стран, которые для нас самые близкие научные партнеры, то есть те страны, с которыми мы чаще всего пишем статьи — не высокоцитируемые, а все статьи. С кем у нас развита коллаборация? У нас больше всего она развита с Германией, дальше США и Франция. Я уже рассказал, что считается статьей английской, американской, немецкой — это статьи, в которых указана соответствующая страна в одном из адресов места работы. Например, когда статья считается российской? Когда хотя бы один из соавторов указал в качестве одного из своих мест работы организацию, адрес которой содержит слово Russia, то есть она размещена в России. Здесь есть некоторые тонкости. Понятно, что эти цифры не аддитивны, то есть статья может

принадлежать сразу нескольким странам по количеству стран, участвовавших в ее создании. И еще бывают такие случаи, надо иметь это в виду: статья может быть за авторством одного человека без соавторов, но при этом принадлежать двум, а то и трем странам, если он одновременно работает в организациях, размещенных в разных странах.

Эти цифры сами по себе может быть не очень интересны, но их интересно сравнить вот с чем. Посмотрите, эта колонка показывает, какая доля у соответствующей страны во всех публикациях, а эта колонка показывает, какая доля в высокоцитируемых статьях. Оказывается, что каждая 40 статья в мире пишется с участием российского ученого, то есть ученого, который в качестве одного из мест работы указал организацию из России. А если мы возьмем блок только с высокоцитируемыми статьями, там наша доля меньше. Там наша доля чуть выше 1%. Это плохо: получается, что мы пишем статей больше плохих, чем хороших. Что у нас с другими странами, например с США? В мире каждая третья статья (сейчас Китай немного подвигает, скоро будет каждая 4 статья) написана при участии американского автора. А если мы возьмем высокоцитируемые статьи, то будет каждая вторая и даже чаще. Это хорошо, это значит, что больше высококачественных, «шедевральных» статей пишется. У Германии и Франции показатель тоже возрастает, если мы переходим от массива просто статей к статьям высокоцитируемым.

Распределение по рубрикам, по тематическим областям. Вы видите, что подавляющее большинство наших статей написано по одной единственной рубрике — это физика. Более-менее измеримые, статистически значимые показатели у нас только по физике, потому что есть 281 статья по физике, а дальше скачок до 40 статей по инженерии (обычно переводят «технические науки»). Дальше химия, клиническая медицина и так далее. Обратите внимание: вообще в мире больше всего статей пишется по клинической медицине, соответственно и высокоцитируемых статей в мире больше всего по клинической медицине, потому что это 1% от всех статей. У нас наука устроена немного по другому, у нас физика и химия на первых местах по числу публикаций. Дальше науки о Земле, биология/биохимия, науки о космосе, материаловедение, математика (всего 9 статей, несмотря на то что математикой мы так гордимся). Дальше идут рубрики, где у нас есть немного статей: есть по одной статье в компьютерных (вычислительных) науках, в нейронауках и в психиатрии/психологии; заметьте, в общественных науках у нас есть 3 статьи, хотя наших журналов там мало, — но, разумеется, не из наших журналов, не из тех трех, там высокоцитируемые статьи. И есть 4 области, где за эти 5 лет не было опубликовано ни одной высокоцитируемой статьи с автором из России. Это иммунология, «multidisciplinary» (статьи, которые не смогли классифицировать), фармакология и по экономике/бизнесу тоже ни одной статьи. Вот так выглядит у нас дисциплинарная структура нашей науки, если посмотреть на высокоцитируемые статьи.

Посмотрите, здесь каждый столбик это какая-то дисциплина. Сами столбики — это абсолютное число высокоцитируемых статей, опубликованных по соответствующей области нашими учеными — тех, где мы соавторы (или авторы, если нет соавторов из других стран). Отрезки с квадратиками на конце — это процент из всех опубликованных нами в этой области статей тех, которые становятся высокоцитируемыми. Я напому, что в мире в среднем процент статей, которые становятся высокоцитируемыми, — это 1%. Это среднемировой уровень. Больше того, и в физике в среднем в мире каждая сотая статья, то есть 1%, становится высокоцитируемой, и в химии... В каждой дисциплине это тоже средний уровень. Посмотрите: первый столбик, физика, самый у нас большой, где 280 статей. Несмотря на то что там самое большое число статей, все равно лишь 0,8% из всех статей, которые мы пишем по физике, становятся высокоцитируемыми. То есть мы реже пишем высокоцитируемые статьи даже в физике, в нашей самой сильной области. Реже, чем в среднем по миру, даже здесь не дотягиваем до уровня 1. Дальше идут технические науки и т. д. А вот, например, клиническая медицина — она сама

по себе невелика, но на втором «относительном» месте. То есть мы немного пишем статей по клинической медицине, поэтому у нас получается и мало высокоцитируемых статей, но относительный показатель на фоне других дисциплин — второй: вероятность, с которой статья, написанная нами по клинической медицине, становится высокоцитируемой, — 0,58%.

Вот такая раскладка, здесь дальше видно. Мелкие столбцы не нужно смотреть, потому что едва мы что-то там публиковали. Посмотрите, например, на провал химии. Химия на третьем месте по общему числу высокоцитируемых статей, но если посмотреть, сколько мы вообще пишем статей по химии, мы должны были бы иметь гораздо больше высокоцитируемых статей.

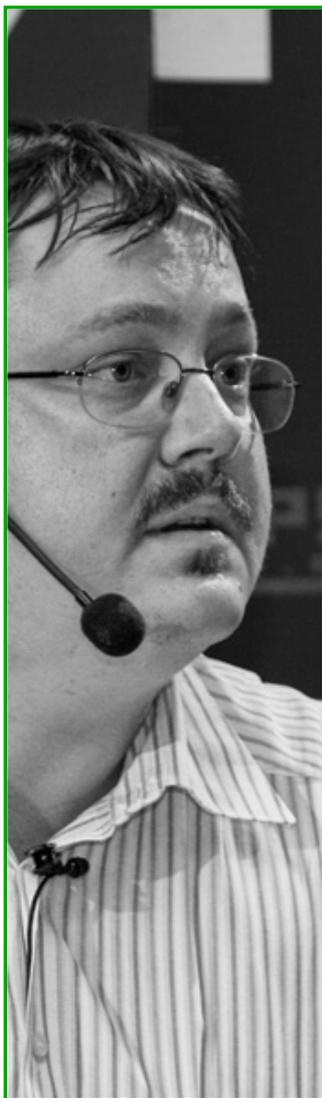
Теперь, чтобы более понятно, практически тот же слайд, только изложенный словами и не целиком, не по всем дисциплинам. В среднем у нас в России становится высокоцитируемой одна статья из 240. Вместо одной статьи из ста, как в среднем по миру, у нас — одна из 240. При этом по физике чуть-чуть не дотягиваем до среднего уровня, одна из 120 становится высокоцитируемой; в клинической медицине — одна из 170, в технических науках и в биологии/биохимии — одна из 230, в науках о Земле — одна из 310, то есть там уже в три раза меньше среднемирового уровня. При этом обратите внимание, что с химией самый удивительный случай. Во-первых, она на третьем месте по числу наших высокоцитируемых статей, во-вторых, она на втором месте по всей публикационной активности в России: если смотреть на число всех публикаций в России, у нас получится, что на первом месте физика, на втором месте химия. А здесь получается, что физика остается на первом месте по проценту попадания в высокоцитируемые статьи, а химия... химия стелется внизу.

Дальше. Основная цель развития нашего исследования — посмотреть, как мы сотрудничаем с разными странами или внутри России с разными организациями. Здесь можно посмотреть, сколько статей мы пишем в соавторстве с учеными, например, из Германии, США и сколько при этом процентов статей из написанных в соавторстве с этими учеными — в высокоцитируемых статьях. Это — массив всех статей, которые мы пишем в России, а это — статьи, которые становятся высокоцитируемыми. В среднем мы пишем примерно каждую десятую статью с соавторами из Германии, а если мы посмотрим высокоцитируемые статьи, то там почти каждая вторая статья у нас будет иметь соавтора из Германии. Здесь можно смотреть, насколько эффективно сотрудничество с теми или иными странами. Например, США для нас на втором месте, мы пишем с ними примерно каждую 11 статью, а если посмотреть на массив всех высокоцитируемых статей, то каждая вторая статья из них имеет соавтора из США.

Раскладка по странам интересна как детальный анализ, но может быть самый любопытный показатель это вот здесь. Вообще у нас в среднем в России в международном сотрудничестве (так называются эти статьи) пишется примерно каждая третья статья, то есть 36% статей с 2003 по 2007 гг. было написано при участии какого-то международного соавтора, неважно из Украины он или из США. А если мы обратимся к высокоцитируемым статьям, окажется, что мы практически не пишем статей, которые бы становились высокоцитируемыми, без международного соавторства. То есть здесь две трети статей без этого соавторства, без коллег из-за рубежа, а в высокоцитируемых статьях — 34 статьи из 511 у нас было написано без участия международного соавтора, представляете?

Голос не в микрофон.

Владимир Писляков: нет, тут по-разному. Паровоз ведь он в обе стороны паровоз, не обязательно нас тянут, бывает, что и мы тянем. Тяжело подсчитать этот показатель — сколько всего высокоцитируемых статей написано в международном соавторстве, — но в принципе они чаще всего пишутся большими международными коллективами. То есть комментарий такой: надо использовать выгоду



Владимир Писляков
(фото Наташи Четвериковой)

международного соавторства, потому что у кого-то установка, кто-то метод изобрел, кто-то писать хорошо умеет или... у него «знакомства» с журналами. Это преимущество надо использовать, и это одна из характеристик высокоцитируемых статей вообще — что они чаще всего пишутся большим коллективом. Необязательно большим, важно, что международным коллективом. Кстати, у высокоцитируемых статей есть еще особенности — скажем, достаточно большое влияние на цитируемость всей страны только высокоцитируемых статей. Было исследование Дага Акснеса[5], норвежского ученого, что если взять только высокоцитируемые статьи, то они по своей цитируемости внесут хороший вклад в весь большой массив. В нашем случае в 200 раз больше массив всех статей, которые мы публикуем.

Да, в принципе считается, что этот показатель (доля статей, написанных в международном соавторстве) нам надо увеличивать. Это, кстати, один из показателей, который учитывается в министерстве, и у них есть плановое движение, и план 40 с чем-то процентов на ближайшую перспективу. К сожалению, оказалось, что в последние несколько лет, буквально в последние два года, этот показатель, наоборот, стал снижаться. Он рос, рос у нас, — и почему-то... Вот здесь уже не осталось препринтов, но у нас на сайте есть[6], — там как раз анализируется ситуация, что в последние буквально два года международное сотрудничество в процентном соотношении остановило рост.

Зададимся другим вопросом. Хорошо, нам выгодно писать с другими странами, это мы поняли. Ну, это разумеется, это можно было бы предположить и так. А странам выгодно с нами писать или нет? Вот достаточно любопытные данные: можно посмотреть процент статей, которые становятся высокоцитируемыми у страны, если они пишутся с Россией, с участием российского ученого,

и есть процент для остальных статей — в которых нас точно нет, — какой процент из них становится высокоцитируемыми. Получается очень радужная картинка: скажем, когда США с нами пишут, то каждая 40-я статья становится у них высокоцитируемой, а когда без нас — каждая 50-я. Картинка, действительно, чуть ли не единственная из очень позитивных у меня, но на самом деле это, конечно, другой эффект, и его тоже можно было ожидать. Когда мы выкидываем Россию, мы автоматически выкидываем и статьи, которые написаны очень большим авторским коллективом, где кто-то из России тоже есть, и эти статьи очень часто получают много цитирований. И когда мы лишаем страну и говорим: нет страна, давай мы рассмотрим твои статьи без России, — все эти статьи отсюда выпадают и получается, что процент высокоцитируемых падает. То есть это скорее всего не наш личный вклад, а вклад того, что есть большие коллективы, куда мы тоже входим.

Теперь я перехожу к части, которую называю «персональный состав». Мы посмотрим, какие организации и, в конце, какие авторы работают и пишут высокоцитируемые статьи. Надо сказать, что это самая опасная, самая зыбкая часть не просто выступления, а исследования, потому что всё хорошо в библиометрии, пока вы работаете с общими показателями, например, со странами. Как только вы приходите и пытаетесь рассмотреть в лупу человека, это, во-первых, крайне трудозатратно и, во-вторых, это часто дает какие-то сбои. Потому что, например, встает человек из зала и говорит: «Да я знаю его лично, он давно уже там не работает!». То есть когда мы переходим от библиометрии к биографии, всё

становится сложнее. Найти только имена и отчества у тех людей, которые будут показаны (здесь, правда, инициалами), отняло рабочую неделю.

Дальше мы не можем исследовать просто так этот массив из 511 статей. Он содержит такие статьи, что там заисследуешься, невозможно. Это статьи, написанные 200 авторами, где-то бывает 300 авторов и, соответственно, мест работы тоже достаточно много. Я не люблю этот термин, но здесь пришлось для наглядности использовать: когда такие коллективы, на научном сленге это называется «братская могила». Да, там еще неизвестно, участвовал человек в статье или нет, потому что иногда пишут «исследовательская группа такая-то» — и всё, что кто-то пишет из этой исследовательской группы, будет приписано к всему коллективу. Если мы переходим на персональный уровень и на уровень организации, мы решили избавиться от этих статей и дали определение этим странным статьям, где непонятна роль конкретного человека. У нас должно быть не больше 100 авторов; у нас должно быть не больше 30 мест работы в списке адресов мест работы всех авторов; и у нас статья не должна занимать больше 100 страниц, даже если там два автора и одно место работы. Я поясню, в чем дело. Есть некоторые обзорные статьи, которые занимают, например, 150 страниц, фактически это уже главы книг, если не целые книги. Это обзорные статьи, скажем, «Review of Particle Physics». Все-таки большой процент ссылок из-за того что такая большая статья, за счет объема она побеждает всех. Такие статьи тоже нами удалялись. Еще плюс здравый смысл, без него нам здесь не обойтись, хотя это самое сложное для формализации. Например, идут две такие статьи, одна статья 120 страниц, вторая 89 страниц, они абсолютно одинаковые, просто посвящены соседним разделам, почему мы должны ту выкидывать, а эту оставлять? В этом случае такого рода главы из книг тоже элиминировались. Из 511 статей осталась 391 статья, которые мы уже будем рассматривать под микроскопом. На самом деле их оказалось не так много, выброшенных статей. 391 статья, где вполне ощутимо влияние авторов.

Интересный момент: сколько может написать автор высокоцитируемых статей? Насколько это сложно? Оказалось, что из всех 663 авторов подавляющее большинство, 538, написали только одну высокоцитируемую статью. То есть они один раз сумели за 5 лет попасть в пул высокоцитируемых статей. Две статьи написали гораздо меньше, 58 человек, и так далее, вплоть до последнего чемпиона, который написал 10 статей за этот промежуток времени, попав чаще всех в список highly cited papers. Здесь учитывались статьи, в которых указывали российское место работы. Это любопытно, дело в том, что в библиометрии есть информетрические распределения, одно из них называется закон Лотки, это Альфред Лотка. Закон гласит следующее: если у вас в научном коллективе есть 100 человек, которые написали по одной статье за какой-то промежуток времени, то в этом же коллективе будут 25 других человек, которые написали 2 статьи, будет порядка 11 человек, которые написали 3 статьи... — обратная квадратичная зависимость. То есть если 100 написали только по одной статье, то две статьи за то же самое время напишут 25 человек, 100 делить на 2 в квадрате, и так далее, вплоть до того что будет один человек, который напишет 10 статей. Будет сотня, которые выдают всего по одной, и будет один, который десять. Здесь же, вы видите, если взять первую и вторую цифру, видно, что разница не в 4 раза, а практически в 10 раз, и закон Лотки не выполняется. Чтобы это точно проверить, можно построить график в логарифмических координатах, по закону Лотки у нас должна получиться прямая линия. Здесь это число авторов, которые написали вот столько статей, логарифм и там и там. Если бы у нас была, как по закону, строгая квадратичная зависимость, у нас должна была бы быть прямая с тангенсом угла наклона минус 2, угловой коэффициент. А эти точки, которые вы видите, взяты из предыдущего слайда и нанесены на этот график. Видно, что они сильно отклоняются и сильно падают ниже этой прямой. То есть оказывается, что если вообще-статьи пишутся примерно по закону Лотки, то высокоцитируемые статьи ему не подчиняются. И в первую очередь этот завал вниз означает, что вторую высокоцитируемую статью написать гораздо сложнее, имея столько же авторов, которые написали по одной статье. Если

здесь построить регрессию, то коэффициент будет не минус 2, а минус 2,8, то есть прямая идет гораздо круче, если пытаться это приблизить прямой.

Распределение авторов. Статей у нас 391, авторов 663, получается 1,7 автора на статью в среднем. Надо сказать, что мне показалось, что это довольно много. Оказалось, у нас не только зарубежное, но и внутрirosсийское соавторство достаточно развито при написании высокоцитируемых статей, это интересный факт. При этом авторов, которые бы работали одновременно за рубежом, всего 18% — 121 автор, это тоже достаточно неожиданное открытие. И для того чтобы проверить, а как такое, почему так мало, вот следующий слайд. Мы возьмем авторов, которые написали одну статью, две статьи за это время и так далее. А какой будет процент среди этих авторов тех, которые одновременно работают за рубежом? Оказалось, что среди авторов, которые написали всего одну статью, тех пятисот авторов, процент «зарубежников», которые работают и там и там, всего 18% [7]. Дальше, две и три статьи — их уже побольше. А если мы возьмем авторов, которые написали больше четырех статей, их семеро, то здесь почти все, кроме одного человека, будут одновременно работать за рубежом. То есть общий процент людей, которые совмещают, не такой большой, но авторы, которые работают за рубежом, больше пишут высокоцитируемых статей.

Наконец, где работают авторы высокоцитируемых статей? Мы взяли статьи авторов, которые написали 4 или больше высокоцитируемых статей, это достаточно активные авторы, и таких авторов у нас 23. Если по ведомствам, 13 из них из центрального РАН, один человек из Сибирского отделения РАН, один человек в академии медицинских наук. Далее, институты, которые не входят в структуру РАН: Курчатовский институт — 3 человека, Объединенный институт ядерных исследований — 2 человека, ИТЭФ — один человек. Кроме этого есть всего два человека из вузов, это из МГУ Белецкая и из Томского педагогического университета Одинцов, физик, он тоже попал в этот список активных авторов.

Дальше. Города, где работают, где места работы. Это в первую очередь Москва — 13 человек. В Московской области 4 человека — это если считать, что ИТФ имени Ландау, Институт теоретической физики, что он в Черноголовке. Просто еще есть московский офис, они часто подписываются Москвой, а если считать, что он черноголовкинский, то получается 4 в области. В Санкт-Петербурге 5 человек, в Новосибирске 3 и по одному из Томска (педагогический университет) и Уфы (авиатехнический). Это самые активные авторы. Но ведь организация может не иметь активных авторов, но много не очень активных. Вот, соответственно, данные по организации как целому. Здесь уже другая картина, на первом месте, конечно же, МГУ, это можно было предположить, 44 статьи. ОИЯИ уже в два раза меньше, чем МГУ, ФИАН, питерский Физтех им. Иоффе, ИТЭФ, ИТФ им. Ландау, Курчатовский институт и остальные. Это все организации, которые написали не меньше 9 статей, перечисление видно на слайде. Кстати, появилась первая медицинская организация — онкологический центр им. Блохина. Все остальные — физики, химики, материаловеды, биологи.

Теперь о людях, самое сложное, с чем мы работаем. Есть автор из Томского педа Сергей Дмитриевич Одинцов с 10 статьями. Здесь показаны места работы, а если у автора есть зарубежные места работы, то здесь они тоже перечислены. В основном это США и Великобритания, а первый автор работает в Испании — в Барселоне в каталонских исследовательских центрах. И всего один автор, Ирина Петровна Белецкая, химик из МГУ, работает только в России — либо если она работает еще где-то, то она не указывает свое второе место работы. Я на этом практически закончил, хочу вам в конце показать лицо автора, который пишет много высокоцитируемых статей. Это Сергей Дмитриевич Одинцов, он справа. Я нашел фотографию на сайте компании Thomson, потому что с ними даже было интервью, это его соавтор Наджири или Нахири — японская фамилия, непонятно, как правильно читать. А посередине просто интервьюер, это не автор высокоцитируемых статей. Вдвоем они

писали почти каждую из десяти статей, иногда с присоединением других соавторов. Можно посмотреть, как выглядит автор высокоцитируемых статей.

Я закончил, я хочу в конце поблагодарить Екатерину Дьяченко из библиотеки нашего Пермского филиала, которая собрала все исходные данные и провела первичный анализ. Работа была поддержана Научным фондом НИУ ВШЭ, также был грант Минобрнауки. Спасибо вам за внимание. Теперь можно снова задавать вопросы.

Обсуждение лекции. Часть 2

Борис Долгин: спасибо. Один маленький методический вопрос. Зачем, когда вы анализируете персоналии, вы исключаете большие списочные, я понимаю. Но ведь когда вы анализируете организации, то исключать из этого массива большие списки сотрудников этих организаций - кто принимал активное участие, кажется ненужным, потому что это и есть структурные подразделения тех организаций, которые вы анализируете, то если работа высокоцитируемая, то почему бы не включить ее в анализ по организациям.

Владимир Писляков: есть исследовательская группа, состоящая из людей со всего мира, там 200 человек. И любая статья, кто бы из них ни написал, — скажем, вдесятером собрались, — у них такое условие: как только мы пишем статью, мы подписываемся всеми двумястами.

Борис Долгин: то есть это когда поверх конкретной организации исследовательская группа.

Владимир Писляков: да, не одна организация, это именно неформальное объединение людей из разных организаций. Но надо сказать, что все равно в какой-то мере можно сказать: да, а мы внедрили нашего человека в эту группу, в какой-то мере мы его финансировали, в какой-то мере эта статья принадлежит и нашей организации тоже, и не только по библиометрическим данным. Поэтому у нас сейчас будет вторая фаза исследования, мы там будем рассматривать все статьи вообще, и я скажу в чем наша цель. Мы сейчас хотим сконцентрироваться на том, как наши организации взаимодействуют между собой, чтобы получилась высокоцитируемая статья. У каких чаще получается — когда они копулируют, то получается больше высокоцитируемых статей. Это интересно, и у нас там некая визуализация должна быть, диаграмма, которая показывает это. И во второй части мы будем учитывать все статьи.

Борис Долгин: ...вы как-то очищали случаи, когда это международное сотрудничество авторов разных стран или когда это разная аффилиация одного автора?

Владимир Писляков: нет, это все считалось вместе, даже если это один автор с двумя странами, то есть это институциональное сотрудничество.

Борис Долгин: нет ли смысла попробовать еще и внутри эту разницу почувствовать, потому что все-таки это немного разные истории?

Владимир Писляков: да, безусловно, это было бы интересно. Надо подумать, насколько это возможно в плане трудозатрат. Но если брать массив только наших статей, то вполне возможно.

Борис Долгин: да-да, хотя бы для наших. Спасибо.

Андрей: когда вы говорили о структуре науки, ее представленности в этих базах данных, там получилось, что физика, химия и так далее — это науки, по которым много статей и много высокоцитируемых статей. А вот науки общественные и так далее — с ними всё плохо. Не может ли это быть следствием того, что российских

журналов в этих базах данных меньше? Вы сами говорили, что общественно-научных журналов только три. Вопрос такой: насколько можно доверять такому структурному распределению и действительно ли все у нас так плохо с общественными науками?



Владимир Писляков
(фото Наташи Четвериковой)

Владимир Писляков: спасибо, хороший вопрос. Как раз немного прояснить отношения между нашими и не нашими журналами и статьями. Дело в том, что большинство высокоцитируемых статей все равно публикуется в международных журналах. Если бы взяли и добавили еще 20 журналов по общественным наукам — решил бы Thomson добавить, как он 33 недавно добавил из естественных наук, — все равно на числе высокоцитируемых статей это ровно никак не отразится. Потому что статьи, которые входят в 1% наиболее цитируемых, все равно публикуются в международных журналах. Я сейчас не готов говорить со стопроцентной уверенностью, в каких журналах хорошо публиковаться или в каких журналах наши ученые любят публиковаться, чтобы получались высокоцитируемые статьи. Это будет в следующем раунде. Но скорее всего там ни одного нашего журнала нет^[8]. У нас максимальный импакт-фактор журнала, который есть в базе данных, — чуть больше 2. Это «Успехи химии», «Russian Chemical Reviews», обзорный журнал по химии. На втором месте, почти с той же двойкой — не помню, может быть 2,2, сейчас новые импакты вышли, — «Успехи физических наук». Дальше все ниже. Журналы, в которых публикуются высокоцитируемые статьи, у них импакт-факторы, как правило, больше. Даже несмотря на то что мы перейдем в область общественных наук, где цитируемость меньше.

Андрей: это связано с языком или это связано с местом?

Владимир Писляков: да, это связано с местом публикации прежде всего, но и с уровнем статьи, это связанные вещи.

Борис Долгин: тут есть еще один вопрос, который прозвучал или полупрозвучал, в какой степени можно использовать эти данные как инструмент оценки качества социальных и гуманитарных наук в России. Или скорее это говорит не о качестве их самих по себе, а о степени их интегрированности. Или это два взаимосвязанных понятия.

Владимир Писляков: если мы исследуем именно высокоцитируемые статьи, то это абсолютно корректный подход. У нас есть три статьи по общественным наукам, это статьи с участием наших ученых, которые были опубликованы, конечно же, за рубежом и они стали по-настоящему высокоцитируемыми статьями. Наших общественно-научных журналов могло и не быть в базе, это бы никак не сказалось. Другое дело, если мы подходим к общественным наукам не исследуя высокоцитируемые статьи, а просто — «сколько у нас статей в базах данных». Тут мы увидим, конечно, что у нас очень мало статей, и это будет именно потому, что наших журналов там мало. А на высокоцитируемые статьи это как раз не влияет.

По поводу уровня общественных наук. Здесь надо понимать, что общественные науки гораздо хуже переносятся из одной среды в другую, из России в международное сообщество. Потому что если человек изучает экономические

особенности лесной промышленности Тамбовской губернии, то вряд ли это будет интересно в журнале Science, American Journal of Sociology или American Economic Review.

Борис Долгин: на самом деле правда, разговор о социальных и гуманитарных науках и как работать с проблемой их интернационализации с одной стороны и с их уровнем с другой стороны, это отдельный разговор, о котором стоит когда-нибудь в будущем подумать.

Олег Градов: здравствуйте, у меня такой вопрос. Учитываются только отечественные публикации — или публикации, вышедшие на других языках тоже? Потому что известно, что многие отечественные журналы переводятся на том же Springer'e под другими названиями, чем в России. В Китае и Японии то же самое, тот же китайский Journal of Tropical Disease выходит под совершенно другим названием в переводе. Поэтому вопрос такой: если статья опубликована и там и там, то суммируются ли данные по статье, вышедшей под одним названием и под другим названием.

Владимир Писляков: имеется в виду, она опубликована на родном языке, а потом, например, переведена на английский, как в нашем случае МАИКовских журналов?

Олег Градов: да.

Владимир Писляков: если посмотреть базу данных Journal Citation Reports, там это хорошо видно. У тех журналов, где такое суммирование производится, стоит значок «+». Например, «RUSS CHEM REV+», это «Russian Chemical Reviews», «Успехи химии», стоит «+» — это значит, учитываются ссылки как на русскоязычное издание, так и на переведенное издание на английском языке. И в большинстве случаев это так. Если такой «+» не стоит, скорее всего, это журнал, который либо не имеет английской версии, либо, наоборот, выпускается здесь сразу на английском языке и не имеет русскоязычной версии.

Ирина Галанова: скажите, пожалуйста, Владимир, что влияет на продолжительность жизни цитируемости статей, по вашим наблюдениям? Что влияет на жизнь научной мысли?

Борис Долгин: то есть жизненный цикл статьи.

Владимир Писляков: то есть когда ее перестали цитировать, забыли о ней.

Борис Долгин: а потом может быть вспомнили...

Владимир Писляков: а бывает, сначала не заметили, а потом вспомнили, этот эффект называется sleeping beauties, «спящие красавицы»... Понятно, что каждая статья уникальна, но если брать глобально, это опять в первую очередь зависит от области, где она опубликована. Потому что есть области, где знание очень быстро устаревает, — та же медицина, где статьи пятилетней давности это уже старье. Есть области, где наоборот: если взять историю или Philosophy of Science, там наоборот любят ссылаться на классиков. В базе данных по журналам, о которой мы сегодня лишь косвенно говорили, Journal Citation Reports, прямо считается показатель cited half-life — это медиана хронологического распределения ссылок на журнал. Если в дисциплине cited half-life 4 года — это значит, что основной массив ссылок, 50%, приходится на самые свежие публикации. Это свойственно медицинским или естественнонаучным дисциплинам, скажем, биологии. Если же брать менеджмент, его я недавно смотрел, там больше 10 лет среднее время. Он долгоиграющий: долго разгоняются, но потом долго цитируют, если это хорошая публикация. Это если брать общий подход, а так каждая статья конечно уникальна, со своей уникальной историей.

Александр Нотченко, Бауманский институт: хотел узнать насчет библиометрии и наукометрии, часто ли используются такого рода исследования в других странах для обоснования государственной политики управления научными исследованиями, финансирования и так далее. Спасибо.

Владимир Писляков: да, часто, чаще, чем у нас. Если отвечать знаючи, то надо устраивать обзор всех политик и тогда со знанием дела говорить. Я на это не готов и я этим мало занимался. Но я точно знаю о RAЕ, которое было в Великобритании, которое сейчас на REF заменяется, там сбор данных по университетам и в том числе собираются данные по цитируемости, по публикациям в высоких журналах. И это влияет на, по-моему, пятилетний цикл финансирования соответствующих организаций. Опять же насколько я знаю, в США чаще не на уровне государственной политики, а на уровне выдачи конкретных грантов учитываются показатели авторов или коллективов — публикационные, или данные по цитируемости, или и то и другое. Длина резюме всегда учитывается.

Борис Долгин: я хочу напомнить, что по проблемам научной политики у нас были лекции, например, [Михаила Соколова](#), где он в том числе говорил и о том, какие аргументы используются. Лекции висят на сайте в виде видео и в виде текста.

Сергей, МГУ: у меня скорее уточняющий вопрос. Правильно ли я понял, что среди 22 дисциплин, на которые разбиты все издания, есть отдельно биология и биохимия и отдельно молекулярная биология и генетика.

Владимир Писляков: да.

Сергей, МГУ: странно, что целая физика, целая химия — и тут биохимия и молекулярная биология в разных дисциплинах.

Владимир Писляков: вы знаете, вообще можно сказать, что там и по числу журналов несколько более пристальное внимание к life sciences, чем ко всем остальным наукам. Это обусловлено в том числе исторически — тем, что все начиналось с создания (еще не в виде Web-интерфейса, а в виде изданий) подобных индексов для медицинских наук.

Борис Долгин: к тому же, видимо, сейчас это одно из передовых направлений. За счет этого большее количество журналов.

Владимир Писляков: да, совершенно верно, соответственно их надо более мелко делить, чем другие журналы.

Коротков: добрый день! Я был журналистом, преподаю журналистам, занимаюсь связями с общественностью, и я хотел бы понять: если мы знаем технологии раскрутки сайтов в Интернете, то можно ли использовать технологии раскрутки статей. Я предполагаю, что есть механизмы, я знаю, как можно повысить цитируемость какой-то работы, если к этому подключить обычную прессу. То есть заняться популяризацией науки.

Борис Долгин: если я не ошибаюсь, речь шла о цитировании в научных работах.

Владимир Писляков: в научных, да, конечно. Вы знаете, есть какие-то более скромные методы раскрутки. Например, один из методов — правда, сейчас дискутируется, приводит это к повышению цитируемости или нет, — это размещение своей статьи в открытом доступе. Либо в виде author-pay model, когда многие издательства, если им заплатить некоторую сумму, будут открывать статью не для всех подписчиков, а для всех вообще. Либо самоархивирование на своих сайтах. То есть такого рода работа по популяризации своей статьи, по повышению видимости, чтобы больше людей ее могли найти, она ведется. Разместить в популярных архивах, например, в ArXiv'e... Насчет прессы я, честно говоря, не

встречался с такой практикой на Западе. Чтобы брали какую-то статью по дифференциальной геометрии — и в газете «Правда»...

Наталья Демина: спасибо за хорошую лекцию. У меня три вопроса, первый вопрос: мы много слышали о золотом веке советской науки, можно ли что-то сказать о том, что бы было, если бы мы анализировали советскую науку в плане высокоцитируемых статей, или нельзя ничего сказать, потому что не было журналов, которые входили в западные базы данных?

Владимир Писляков: навскидку я могу предположить, что да, этого не было, потому что большинство тогда публиковались в наших журналах, пытаюсь заставить всех выучить русский. Скорее всего, картина не была бы слишком радужной, а так я этим не занимался, к сожалению [9].

Борис Долгин: хотя понятно, что если вспомнить о 20-х годах, условно говоря, то практика публикаций, как гуманитариев, так и естественников в международных журналах была распространена. Когда еще не были разорваны связи, когда еще не сформировался этот самый «золотой век советской науки».

Наталья Демина: второй вопрос, могут ли ваши исследования дать какие-то советы, рекомендации для policy maker'ов, наших чиновников, — что сделать, чтобы было больше высокоцитируемых статей?

Владимир Писляков: этому мы надеемся посвятить второй раунд. Я говорил, когда, я надеюсь, выяснится, что сотрудничество конкретных организаций внутри России (да, с привлечением международных партнеров, причем у каждой организации будут свои) чаще приводит к возникновению высокоцитируемой статьи, это будет что-то более конкретное. Пока, кроме того что работать над цифрой 36, это процент в международном соавторстве, конкретики, пожалуй, тут нельзя предложить.

Наталья Демина: третий вопрос, если вы сочтете корректным ответить. Назовите, пожалуйста, позитивные стороны РИНЦ и негативные стороны российского индекса цитирования. И что бы вы сказали ученым, которые считают, что пока нельзя использовать РИНЦ. В официальных документах требуется, чтобы был какой-то рейтинг в РИНЦе. Имеет ли это смысл на данный момент? Спасибо.

Владимир Писляков: Наташа, это немного не по теме лекции... Из корректности, перед столь широкой аудиторией, мне надо подумать, подобрать слова. РИНЦ, безусловно, полезный продукт, то, что он был создан, это безусловно большой шаг в нашей науке. В первую очередь, как я неоднократно говорил, для повышения видимости различных журналов, которые мы раньше вообще не видели. Человек, написавший на вашу же тему во Владивостоке, — вы бы узнали о его статье (если бы вам посчастливилось) где-нибудь в конце своей жизни. Это очень позитивная составляющая. Что касается аналитических функций, здесь, наверное, более спорно. Но в зале присутствует один из разработчиков РИНЦ, и в присутствии эксперта я не буду соваться в калашный ряд.

Борис Долгин: так или иначе, явно тема важная, интересная, и о ней тоже стоит говорить. И о направлениях совершенствования явно тоже.

Владимир Писляков: безусловно. Кстати, если добавить. Эта тема тоже поднималась, сводя воедино обе эти темы. Особенно полезной оказалась роль РИНЦ — я думаю, что уже оказалась, и еще окажется — в плане повышения нашей внутренней видимости журналов по общественным наукам. Потому что раньше кроме этих трех журналов здесь и по общественным, по-моему, 9 журналов в Scopus, просто нигде наша наука не была доступна. Что касается естественных наук, они уже как-то были, но в общественных науках очень серьезный скачок — появление нашего индекса.

Борис Долгин: да, но жалобы со стороны ученых тоже регулярно встречаются.

Владимир Писляков: но больше на аналитические функции, рейтинги и так далее.

Борис Долгин: и это, и то, что необходимо самим доискывать свои статьи, которые оказались невидимы в системе и как-то их включать, если удастся. С другой стороны, наличие инструментов для такого самовосполнения – это тоже хорошее дело.

Павел Арефьев: я хотел выяснить, добить российскую социальную науку вопросом. Эти три наиболее цитируемые российские работы из области общественных наук, к какой конкретно дисциплинарной отрасли они относятся? Я смею предположить, что это или экономика или психология.

Владимир Писляков: я не смотрел по этим трем, надо посмотреть, я виноват, что этого не сделал [\[10\]](#).

Павел Арефьев: скорее всего это либо психология, которая отчасти признается на Западе, некоторые направления, либо экономика, точнее эконометрические методы.

Владимир Писляков: Павел Геннадьевич, можно сразу сказать, что это не экономика, экономика идет в отдельной категории Economics & Business, где у нас нет ни одной статьи.

Дмитрий: скажите, пожалуйста, насколько существует разрыв между научными исследованиями и публикациями об этих исследованиях, если вы можете.

Борис Долгин: разрыв по времени или в чем разрыв?

Дмитрий: нет, вообще, как много исследований проводится, но их не публикуют. И второй вопрос.

Борис Долгин: откуда же мы можем об этом судить, если не публикуются.

Дмитрий: я могу вам сказать, что в российских институтах, даже не закрытых, такая практика есть.

Борис Долгин: исследовать и никому об этом не сообщать?

Дмитрий: сообщать, но узкому кругу, например, делать по заказу и так далее.

Владимир Писляков: у меня нет данных на этот счет. Если брать какие-то организационные проблемы, то на самом деле можно сказать, что да, когда я провел исследования, рассказал на конференции, дальше не всегда легко, даже мне, усадить себя и написать статью, каюсь. Один из примеров вы держите в руках, потому что это препринт, который еще не стал публикацией, хотя ему уже год. Это опять же если брать те вещи, на которые можно ответить. Да, вот этим надо безусловно заниматься.

Борис Долгин: то есть это исследование видимо, но просто не дошло пока до стадии публикации?

Владимир Писляков: да, он в открытом доступе лежит, но финальная работа, которая должна отштамповать, в данном случае не проведена.

Борис Долгин: а второй вопрос?

Дмитрий: проводились ли исследования связи этих публикаций и участия ученых в международных обществах. Например, здесь вспоминали довоенные годы, многие

советские ученые состояли в американских, немецких обществах, а сейчас есть какая-то корреляция и может ли она быть?

Владимир Писляков: здесь я, как и любой из зала, могу ответить: «наверное, да. Я не занимался этим исследованием».

Борис Долгин: наверное тут проблема, чтобы это подсчитать, потому что в отличие от места работы...

Владимир Писляков: ...это не указывается, конечно, это надо брать списочный состав, смотреть и то ли писать письма, то ли искать на сайте, а на сайте тоже могут не указать... Это не простое, хотя очень полезное было бы исследование.

Понимаете, вот этими методами, которыми у нас сделано, можно найти достаточно интересные вещи, как мне кажется, при этом не угрохав на это пять лет жизни. Потому что на таком уровне можно проводить при помощи базы данных, а что-то более детальное, — распределение по возрасту, например, «а в каком возрасте люди пишут высокоцитируемые статьи?», — это сразу на порядок более сложное исследование.

Иветта: я хотела уточнить, сейчас появилась тенденция к публикации журналов только в электронном виде, в том числе потому что это быстрее позволяет публиковать новые исследования. Какой процент в базе Thomson Reuters только электронных публикаций, и какой процент высокоцитируемых статей появился только в электронном виде. Спасибо.

Владимир Писляков: опять большое спасибо за интересную тему для исследования, потому что когда мы сделаем раскладку по журналам, то конечно будет очень интересно узнать, а есть ли там журналы «digitally born». Что могу сказать про Thomson. Процента я не знаю. Но я знаю, что у них на сайте есть такая страничка, — это кстати многих интересует — где можно подать свой журнал в базу данных Thomson. И дальше его будут рассматривать и решат, что «конечно не возьмем». Но страничка есть. И там отдельный параграф, они, по-моему, с него начинают, что если у вас чисто электронный журнал, то вам надо то-то и то-то. А если печатный, то оформить на нас подписку [\[11\]](#). Я не шучу, это одно из условий, оформить подписку на Thomson, чтобы им приходили...

Борис Долгин: То есть, это финансовый ценз.

Владимир Писляков: условия там следующие: журнал должен выходить со строгой периодичностью. Именно это они проверяют при помощи печатной подписки и, по-моему, пока им три номера не придет, они вообще даже не рассматривают этот вопрос. Потому что в такой базе данных, которая индикаторы выпускает, ей очень важно чтобы не было как у нас: то ежеквартальник, а то два номера выпускают, в мае сдвоенный и в декабре сдвоенный, и все. Их это не устраивает. Косвенное свидетельство то, что у них это специально написано, что такие журналы они включают. В Scopus эта цифра есть, это меньше 10%. Там есть журналы открытого доступа, есть чисто электронные журналы, и я не помню, одно из них 3%, другое 8%, в районе 5% изданий чисто электронных [\[12\]](#).

Борис Долгин: на результаты вашего исследования сильно обижались, никто не ругался: как же так, вы продемонстрировали, что наша организация или наши ученые не настолько сильны своими публикациями?

Владимир Писляков: я думаю, что завтрашнего утра начнут.

Борис Долгин: что вас, кстати, натолкнуло на эту тему? Понятно, что для человека, работающего в библиотеке, библиометрия — это естественный хлеб...

Владимир Писляков: на эту работу меня натолкнуло предыдущее исследование, где исследовались показатели цитируемости статей наших ученых, написанных в соавторстве с различными странами. А до этого мы исследовали (вот репринт, кому нужно, оттиск статьи из *Scientometrics*^[13]) так называемый индекс Матфея, это когда публикуются наши ученые в зарубежном журнале и сравнивается средний уровень цитируемости этого журнала с цитируемостью нашей статьи и такая статистика собирается по разным статьям. И оказалось, что в физике у нас на среднем уровне цитируемость, а в химии, если наш человек публикуется за рубежом, относительно средней цитируемости журнала данные получаются гораздо ниже. То есть журнал на более высоком уровне, а наша статья немного выпадает. Когда на все это мы посмотрели, то как раз стало интересным исследовать не все статьи, не средний уровень, а исследовать только шедевры, только самую верхнюю прослойку. Давайте мы забудем об этом болоте, в риторическом смысле слова, там конечно есть масса прекрасных публикаций, давайте возьмем только самые хорошие статьи. И здесь, кстати говоря, — в аннотации к выступлению это было, — специальный подход: когда мы отрезаем при помощи этой базы только 1% самых цитируемых публикаций, то дальше мы можем вообще не смотреть их цитируемость, а изучать, как здесь было, количественные показатели. Ведь мы сейчас считали только число публикаций. Каждый раз число публикаций, нигде нас не интересовало, сколько ссылок получила статья, потому что статья уже попала в 1%. Это другой подход, чем поиски средних, которые были раньше. Наверное, в этом мотивация.

Борис Долгин: спасибо. Вы сказали о следующем этапе своего исследования, собираетесь ли вы потом как-нибудь работать и с РИНЦом как некоторым источником?

Владимир Писляков: в рамках этого исследования нет.

Борис Долгин: нет, не в рамках этого исследования.

Владимир Писляков: это было бы интересно, но пока, наверное, рано.

Борис Долгин: я понимаю, что хотя бы 10-20 лет должно пройти.

Владимир Писляков: в перспективе — конечно. Если база данных РИНЦ пройдет ту же историю, которую прошел Web of Science, то будет очень интересно посмотреть — да, на наши Вестники, на наши Известия разных вузов, ничего страшного, их тоже было бы интересно поизучать, когда появится уже зрелый и доступный инструмент.

Борис Долгин: спасибо большое.

Владимир Писляков: спасибо.

Примечания

[1] На момент лекции — по февраль.

[2] В действительности Nature опережает Science четыре года, а до этого два года вел Science. Впрочем, импакт-факторы обоих журналов велики и близки друг к другу; конечно, во многом это «соперничество» чисто символическое.

[3] Glänzel W., Moed H. F. Journal impact measures in bibliometric research // *Scientometrics*. 2002. Vol. 53, No. 2. P. 171–193.

[4] Это предположение основано на одной нечеткой ремарке в статье про тайваньскую базу данных цитирования гуманитарной литературы. С определенностью сказать, применяется ли там этот механизм, нельзя.

[5] Aksnes D., Sivertsen G. The effect of highly cited papers on national citation indicators // *Scientometrics*. 2004. Vol. 59, No. 2. P. 213–224.

[6] Писляков В. В. Соавторство российских ученых с зарубежными коллегами: публикации и их цитируемость: Препринт WP6/2010/01. М.: ГУ-ВШЭ, 2010. 40 с. (http://www.hse.ru/data/2010/06/01/1219988645/WP6_2010_01fin.pdf)

[7] Неточность, правильно — 15%.

[8] Это не совсем так, например в рассматриваемом массиве есть две статьи из «Писем в ЖЭТФ», одна из «Успехов физических наук» и др.

[9] По состоянию на 1970-е годы положение с высокоцитируемыми статьями советских ученых, по-видимому, было хуже, чем сейчас в России, см.: Narin F., Frame J. D., Carpenter M. P. Highly Cited Soviet Papers: An Exploratory Investigation // *Social Studies of Science*. 1983. Vol. 13, No. 2. P. 307–319.

[10] Это статьи о палеонтологических исследованиях на Дону, связи психосоциальных условий на работе с депрессией в трех городах Восточной Европы (в т.ч. в Новосибирске) и потреблении наркотиков в Тольятти.

[11] Здесь речь идет об оформлении подписки и отправки журнала на адрес компании Thomson, а не о подписке редакции на дорогостоящие продукты самого Thomson'a, что, конечно, было бы серьезным финансовым цензом. Неточность и в описании страницы подачи журнала: она предусматривает подачу не только чисто электронного журнала, но и электронной версии издания, выходящего в печати (как альтернативу отправке собственно печатных номеров).

[12] 10% журналов в Scopus — журналы открытого доступа, с общедоступными полными текстами в Интернете. Данных по числу чисто электронных изданий, не выходящих в печатном виде, найти не удалось.

[13] Pislyakov V., Dyachenko E. Citation expectations: are they realized? Study of the Matthew index for Russian papers published abroad // *Scientometrics*. 2010. Vol. 83, No. 3. P. 739–749.