

ЭЛЕМЕНТЫ НАУКОМЕТРИИ В ПОЧВОВЕДЕНИИ И ЭКОЛОГИИ (НА ПРИМЕРЕ ФАКУЛЬТЕТА ПОЧВОВЕДЕНИЯ МГУ)

Глаголев М.В., Янин М.В.
m_glagolev@mail.ru

Что такое наукометрия и зачем она нужна?

Под названием «наукометрия» оформилась область статистического изучения динамики информационных массивов науки и потоков научной информации. Восходящая к трудам Прайса и его школы, она представляет собой применение методов математической статистики к анализу потока научных публикаций, ссылочного аппарата, роста научных кадров, финансовых затрат [Кохановский с соавт., 2007]. В данной работе (на примере, главным образом, факультета Почвоведения МГУ им. М.В. Ломоносова) мы рассмотрим лишь отдельные элементы наукометрии, связанные с анализом публикаций и ссылочного аппарата.

Какова значимость ученого и его воздействие на научный процесс? Как мы можем измерить престиж журнала или конференции? Оценивание работы ученого, а также качества журнала или конференции уже давно привлекает внимание исследователей. Этот значительный интерес обусловлен теми преимуществами, которые может дать справедливый и открытый критерий такого оценивания [Sidiropoulos et al., 2006]. Пристальное внимание ученых и администраторов к научному цитированию объясняется тем, что этот процесс – практически единственный «видимый» след научной коммуникации и механизма рождения нового знания, предоставляющий, таким образом, исследователю пусть не исчерпывающие, зато объективные показатели, характеризующие и позволяющие изучать данный механизм [Писляков, 2005]. По образному выражению Р.Г. Касимовой [2004], «цена» отдельной работы, автора или издания – это индексы цитирования или импакт-факторы. Причем в последнее время эта виртуальная «цена» все чаще обретает материальное воплощение.

Действительно, например, статистики баз данных⁵ американского Института научной информации (ISI - Institute for Scientific Information,

⁵ Иногда можно на эти базы данных ссылаются как на JCR и SCI. Однако указатель цитируемости журналов Journal Citation Reports (JCR) и индекс цитирования Science Citation Index (SCI или его интернет-версия Web of Sciences – WOS) – это, скорее, результат обработки БД.

Филадельфия, США) широко используются как характеристики количества и качества научного продукта [Касимова, 2004]. В частности, новая модель финансирования исследований, запланированная Сектором высшего образования Норвегии и по его просьбе разработанная в течение 2002-2004 гг. Университетом г. Осло, была введена в 2005 г. Модель предназначена для шести университетов Норвегии и почти сорока других учреждений высшего образования (приблизительно 8000 публикаций/год). Идея, которая лежит в основе этой новой модели финансирования, заключается в том, что научное качество, зарегистрированное главным образом через публикации, должно быть экономически вознаграждено. Некоторые американские университеты кроме публикаций также вознаграждают гранты [Vibe and Jakobsson, 2005]. Статистики ISI начали применяться и государственными органами России при распределении федерального ресурса, т.е. стали инструментом официальной политики [Касимова, 2004], но пока это касается (в противоположность норвежской модели) лишь научной-исследовательских институтов РАН, а не учреждений высшего образования.

Что такое импакт-факторы и индексы цитирования?

Понятие импакт-фактора родилось в 1975 г. [Васьковский, 1993]. Классический (синхронный, Гарфильдовский) импакт-фактор (ИФ) для некоторого журнала J за год Y представляет собой отношение числа появившихся во всем массиве журналов за год Y ссылок на статьи журнала J, вышедших в годах Y-1 и Y-2, к суммарному числу статей, вышедших в журнале J за годы Y-1 и Y-2:

SCI - поистине удивительное библиографическое издание, выпускаемое Институтом научной информации с 1963 г. В 1975 г. в качестве последнего выпуска годового комплекта "SCI" стало выходить, по существу, совершенно новое издание - JCR. Оба издания тесно связаны между собой: у них общий "изобретатель" - создатель филадельфийского Института научной информации д-р Юджин Гарфилд - и делаются они на основе общей базы данных (тем не менее, между ними есть принципиальное отличие: SCI отражает взаимосвязь отдельных научных публикаций, а JCR - отдельных журналов). База данных, используемая для JCR и SCI, сформирована с помощью экспертного и наукометрического анализа. Она охватывает более 90% важнейшей информации в области естественных наук, техники, медицины и сельского хозяйства. Уже в начале 90-х гг. круг обрабатываемых журналов заметно расширился: кроме указанной базы данных стали использовать еще и базу данных по социальным наукам - "Social Sciences Citation Index", а также 300 медицинских журналов из других баз данных института. К сожалению, эта информация была не слишком доступна российскому читателю. Например, на рубеже 80-90-х гг. XX столетия в Россию поступало всего шесть комплектов "SCI", частью которого был "JCR": четыре в Москву - в Российскую государственную библиотеку, ГПНТБ, ВИНТИ, Библиотеку по естественным наукам РАН, и по одному экземпляру в Ленинград и во Владивосток (причем комплект за 1991 г. появился только в начале 93-го) [Васьковский, 1993].

$$\text{ИФ} = [\text{CIT}_J(Y, Y-1) + \text{CIT}_J(Y, Y-2)] / [\text{PUB}_J(Y-1) + \text{PUB}_J(Y-2)], \quad (1)$$

где $\text{CIT}_J(X, Z)$ – количество ссылок, появившихся во всем массиве журналов за год X на статьи журнала J , вышедшие в году Z ; $\text{PUB}_J(Z)$ – число статей, вышедших в журнале J за год Z [Писляков, 2005]. Например, в JCR за 1988 г. помещена такая информация о журнале "Nature": в 1987 г. в нем было напечатано 1210 статей, в 1986 – 1165, всего за два года – 2375 статей; в 1988 г. на его публикации 1987 года было в общей сложности 18953 ссылки, на публикации 1986 года – 18462, а всего на публикации за эти два года – 37425 ссылок. Если теперь разделить суммарное число ссылок на статьи 1986-1987 гг. (37425) на число самих этих статей (2375), получается 15.578. Это и есть импакт-фактор, то есть среднее число ссылок, приходящихся на одну статью этого журнала [Васьковский, 1993]. Использование импакт-фактора в качестве критерия для оценки журнала основывается на том естественном предположении, что журнал, публикующий значительное число статей, на которые активно ссылаются другие ученые, заслуживает особого внимания. При этом подразумевается, что чем выше значение импакт-фактора, тем выше научная ценность, авторитетность журнала.

При этом следует сделать два замечания. Во-первых, в действительности ISI, который на практике и занимается подсчетом ИФ, учитывает не полное число появившихся на свет публикаций, а только исследовательские статьи и научные обзоры, оставляя за скобками редакционные заметки, письма в редакцию, новости, отчеты о конференциях и др. Во-вторых, невозможно посчитать все существующие в мире статьи, которые ссылались на публикации журнала J , вышедшие в течение определенного периода. Поэтому каждая база данных, каждый индекс цитирования ограничивается некоторым фиксированным массивом журналов [Писляков, 2005]. Никто точно не знает, сколько научно-технических журналов выходит сейчас во всем мире. Одни говорят – 30-40 тысяч, другие – даже 100 тысяч. Известно, что для "Chemical Abstracts" обрабатывают более 12 тысяч журналов [Васьковский, 1993]. Например, в 1991, 2000, 2003 и 2006 гг. ISI и Национальный научный фонд (NSF) анализировали публикации, соответственно, в 4461⁶, 5684⁷, 5315⁸ и 6164 ведущих научных журналах мира. В связи с этим следует остановиться на

⁶ Данные из [Васьковский, 1993].

⁷ Данные с сайта <http://www.mtas.ru/second.php?ID=238>

⁸ Данные из [Арутюнов, 2007].

одной ошибке, часто совершаемой, когда речь идет об ИФ. Например, даже такой уважаемый автор, как д.б.н. В.Д. Сонькин [2005] пишет: «Во всем мире реферируемых журналов с ненулевым импакт-фактором насчитывается почти 6000. Среди современных российских научных журналов только 100 (1,66 процента от мирового количества) имеют более-менее значимый импакт-фактор, остальные никто в мире не читает. Не значит ли это, что 90 процентов наших научных журналов можно было бы и не издавать?». Мысль сформулирована так, что неискушенный читатель неизбежно приходит к выводу: у всех остальных журналов (кроме «почти 6000», в частности, у всех российских научных журналов, кроме «только 100») ИФ = 0. Как видно из (#1), если журнал имеет ИФ = 0, то это действительно означает, что такой журнал никто не читает (если быть точным, то не «никто не читает», а никто не цитирует; но полное отсутствие цитирований, вероятно, означает именно то, что действительно никто не читает данный журнал). Но на самом деле для журналов, не входящих в эти «почти 6000» ***ИФ не нулевой, а он просто не известен.*** Т.е. для остальных журналов ИФ не рассчитан. При чем год от года список «почти 6000» журналов изменяется, из него выбывают одни журналы и в него включаются другие. При этом может оказаться, что вновь включенный в список журнал имеет весьма солидный ИФ. Скорее всего, близкое качество статьи журнала имели и годом раньше, когда, согласно В.Д. Сонькину, их ИФ был нулевым. Он не был нулевым! Он просто не был измерен!! Абсурдно говорить, что еще год назад журнал никто не читал⁹, а вот теперь у него появилось множество читателей!!! На самом деле: год назад число читателей не было известно, теперь же оно вычислено и выражено посредством ИФ¹⁰. Следовательно, то, что ИФ не был измерен для 90% отечественных журналов – это вовсе не повод к их ликвидации (а, скорее, повод к тому, чтобы, все-таки, измерить ИФ и для этих журналов, что, в конце концов, было сделано в рамках

⁹ Утверждение В.Д. Сонькина [2005] о том, что «среди современных российских научных журналов только 100... имеют более-менее значимый импакт-фактор, остальные никто в мире не читает, спорно еще и с другой стороны. Следует ли считать «значимым» ИФ = 0.036 (именно такой ИФ получен для журнала «Почвоведение» в 2006 г.) или даже ИФ = 0.181 (ИФ журнала «Экология» в том же году)? Например, в случае с «Почвоведением» («Eurasian Soil Science») данный ИФ означает, что на каждую сотню работ, опубликованных в этом журнале в 2004-2005 гг., в 2006 г. приходилось менее 4 ссылок!

¹⁰ Чтобы не быть голословными, приведем конкретный пример. В 2000 г. «Int. J. of Systematic and Evolutionary Microbiology» не имел ненулевого ИФ, т.е. согласно В.Д. Сонькину [2005], его никто в мире не читал, а вот буквально на следующий год ИФ этого журнала сразу стал равен 2.004!

государственного контракта по созданию Российского Индекса Научного Цитирования).

Нижний предел значения импакт-фактора, очевидно, составляет 0.000, а верхний предел, т. е. импакт-фактор самого авторитетного и "весомого" журнала, в разные годы разный (рис. 1). Общий рост высоких импакт-факторов объясняется тем, что для "JCR" с каждым годом обрабатывается все больше журналов, и число ссылок в публикуемых ими статьях тоже имеет тенденцию к росту, хотя для отдельных изданий от года к году могут быть большие флуктуации [Васьковский, 1993]. На рис. 1 мы привели (наряду с трендом динамики роста ИФ) еще и ИФ в расчете на 4461 журнал (именно столько журналов обрабатывалось в 1991 г.).

Кроме классического ИФ существует множество иных наукометрических показателей (НМП), с которыми можно познакомиться, например, в статье [Писляков, 2005].

Один из спорных вопросов касается использования ISI-журналов в качестве критерия – должны ли использоваться ИФ журналов в финансировании, основанном на конечном результате исследования.

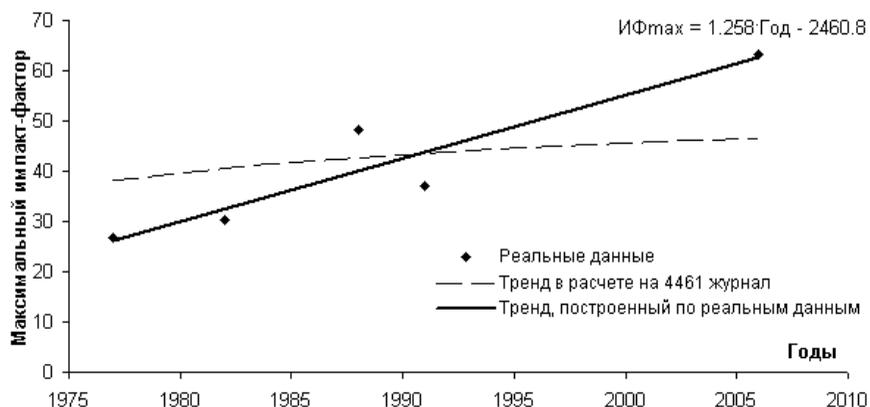


Рис. 1. Динамика роста максимального импакт-фактора.

И некоторые отдельные исследователи, и ряд организаций (в частности, Библиотека Университета Осло и даже сам ISI) считают, что ИФ журнала не может быть использован как критерий качества. Так, классик наукометрии доктор Юджин Гарфильд, основатель ISI, писал еще 10 лет назад: «Большое беспокойство относительно импакт-факторов журналов вызывается их неправильным применением при оценке специалистов... Во многих странах в Европе... импакт-фактор журнала используется как идентификатор для оценки. Я всегда предупреждал

против такого использования. В пределах одного журнала статья статье разнь...» [*Vibe and Jakobsson, 2005*].

К счастью, наряду с такими НМП, существует группа принципиально иных показателей, характеризующих статьи конкретного исследователя, поэтому мы будем называть их личными индексами цитирования (ЛИЦ).

В настоящий момент большинство ученых-научометристов признают, что в общем случае не существует одного-двух количественных параметров, на основании которых можно было бы уверенно и всеохватывающе выстраивать иерархию и рейтинги участников научного процесса. Принятие решений должно осуществляться на базе многоаспектного анализа, принимающего во внимание целый ряд характеристик [*Лисляков, 2005*].

Система "Publish or Perish"

Научометрические параметры, вычисляемые "Publish or Perish"

База данных, используемая системой "Publish or Perish", существенно отличается от БД ISI. Например, в отличие от ISI она содержит не только журналы, но монографии и даже материалы некоторых конференций (что не лишено смысла, т.к. процент цитирования не-журнальных источников в академической периодике колеблется от 6% до 65% в зависимости от отрасли знаний - см. [*Лисляков, 2005*] и дальнейшие ссылки там). Особо подчеркнем, что в БД "Publish or Perish" учтены практически все российские журналы, переводящиеся на английский язык, даже если они не входят в БД ISI (например, «Вестник МГУ. Серия 17: Почвоведение») а также некоторые из тех, которые выходят лишь на русском языке (например, «Агрохимия», «Лесоведение», «Криосфера Земли»). Очень важно и то, что "Publish or Perish" свободно доступна в Internet (www.harzing.com).

Для конкретного исследователя система "Publish or Perish" рассчитывает следующие ЛИЦ (те их обозначения, которые используются в интерфейсе "Publish or Perish", будут выделены ниже в тексте):

1) **Papers** – количество работ, опубликованных данным исследователем в изданиях, входящих в БД "Publish or Perish". Для примера на рис. 2А приведена эмпирическая плотность распределения количества публикаций для сотрудников факультета почвоведения МГУ им. М.В. Ломоносова. Из рисунка видно, что плотность этого распределения представляет собой очень резко убывающую кривую, иначе говоря, очень многие сотрудники имеют мало публикаций¹¹ и очень немногие – много.

¹¹ Правда, здесь необходимо помнить, что речь идет не о публикациях вообще, а о публикациях, учитываемых в системе "Publish or Perish". Впрочем, рис. 4 показывает, что эти последние довольно хорошо ($R^2 \approx 0.70$) скоррелированы с общим числом публикаций. Причем из уравнения регрессии видно, что из каждых 10 публикаций данного исследователя 3-4 учитываются в "Publish or Perish".

2) **Citations** – общее количество ссылок на работы данного исследователя (здесь и ниже под «общим количеством» имеются в виду, разумеется, только те данные, которые учтены в БД “Publish or Perish”). Для примера на рис. 2Б приведена эмпирическая плотность распределения количества цитирований публикаций сотрудников факультета почвоведения МГУ. Из рисунка видно, что плотность этого распределения, как и в предыдущем случае, представляет собой очень резко убывающую кривую. Вероятно, общее количество ссылок лучше характеризует работу исследователя, чем общее количество публикаций¹².

3) **Years** – количество лет, прошедшее с опубликования данным автором самой ранней работы, вошедшей в БД “Publish or Perish”.

4) **Cites/year** - среднее количество цитирований в год (СКЦ/Г), т.е. общее количество ссылок на работы данного автора, отнесенное к общему количеству лет, в течение которых были опубликованы эти его работы («Cites/year» = «Citations»/«Years»). Для примера на рис. 2В приведена эмпирическая плотность распределения интенсивности цитирований сотрудников факультета почвоведения МГУ. Из рисунка видно, что плотность этого распределения, как и в предыдущих случаях, представляет собой очень резко убывающую кривую. Очевидно, что интенсивность цитирования хорошо отражает, так сказать, научный потенциал конкретного сотрудника. Например, если к настоящему моменту времени на работы данного исследователя имеется небольшое количество ссылок, но интенсивность цитирования высока (т.е. это небольшое количество получено так же за небольшой промежуток времени), можно с большой вероятностью ожидать, что спустя некоторое время (можно даже примерно оценить – через какое время) указанный ученый получит весьма значительное количество ссылок. Однако, с другой стороны, написав работу, на которую интенсивно ссылаются в мировой литературе, исследователь уже сейчас (даже когда ссылок еще мало, а не спустя время, когда их будет много), фактически, доказал свою высокую квалификацию. Следовательно, интенсивность цитирования, все-таки, можно считать не только мерилom потенциального, но и актуального «класса» исследователя.

¹² В журнале Nature как-то раз было опубликовано письмо от читателя с замечательной фразой, приблизительно такой: "...правда нынешней науки состоит в том, что опубликовать можно все, что угодно, если только спуститься дотаточно низко по уровням журналов" [Галызин, 2006]. Да, опубликовать-то можно все, что угодно, но вот заставить независимых авторов ссылаться на это «все, что угодно» не представляется возможным.

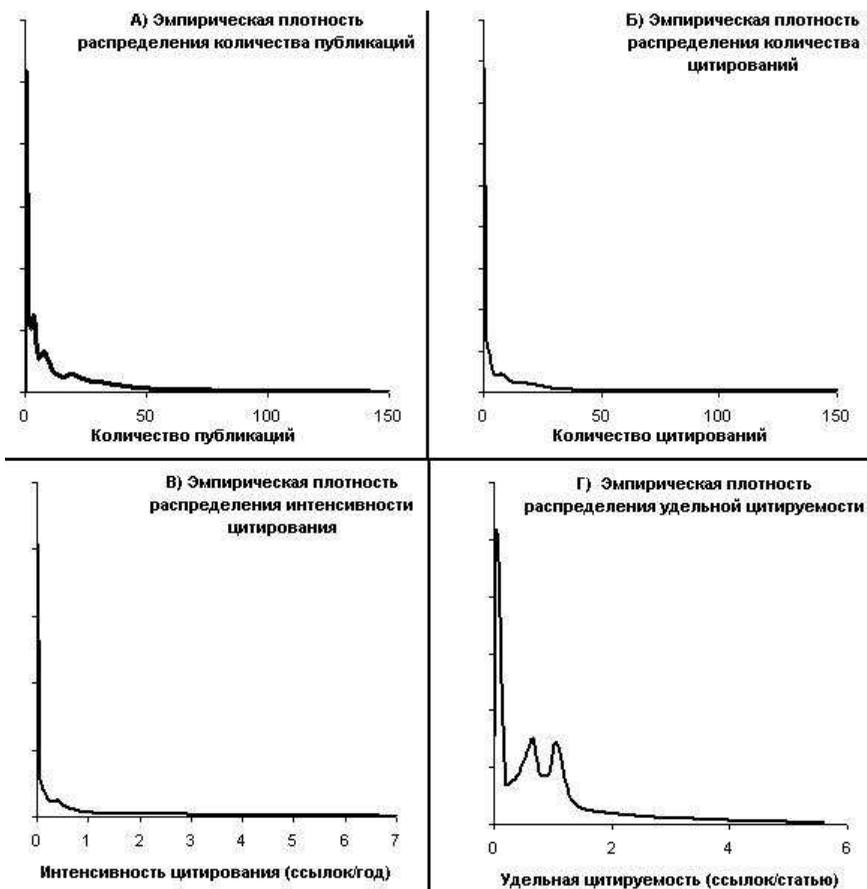


Рис. 2. Некоторые наукометрические показатели сотрудников факультета почвоведения МГУ (по состоянию на 07.09.2008), вычисленные по базе данных публикаций "Publish or Perish" (площадь под кривой равна количеству сотрудников факультета).

5) **Cites/paper** - среднее арифметическое количества цитирований публикации данного исследователя (СКЦ/П), т.е. общее количество ссылок на работы данного автора, отнесенное к общему количеству опубликованных им работ. Для примера на рис. 2Г приведена эмпирическая плотность распределения удельной цитируемости статей сотрудников факультета почвоведения МГУ. Из рисунка видно, что плотность этого распределения, несколько отличается от рассмотренных выше. Если три четко выделившихся пика не являются артефактами, связанными лишь с алгоритмом построения эмпирической плотности распределения, то, конечно, было бы интересно установить их физический

смысл. Очевидно, что удельная цитируемость отражает «среднее качество» статей данного исследователя. Можно предположить, что самый левый (самый высокий пик на рис. 2Г)¹³ соответствует группе «инженерно-технические работники». Представители этой группы лишь эпизодически занимаются написанием научных работ, и получающиеся в результате публикации оставляют желать много лучшего, а, следовательно, не могут набрать большого количества ссылок.

б) **Cites/author** – количество цитирований, приходящихся на одного автора, рассчитанное с учетом количества соавторов в каждой конкретной работе (например, если данный автор написал 8 статей, причем на статью, написанную четырьмя соавторами имеется 2 ссылки, на статью, написанную тремя соавторами – 1 ссылка и также по 1 ссылке приходится еще на две статьи, каждая из которых написана двумя соавторами, а на остальные 4 статьи ссылок вообще не было, то «Cites/author» = $2/4 + 1/3 + 1/2 + 1/2 \approx 1.83$).

7) **Papers/author** – количество статей, приходящихся на одного автора, рассчитанное с учетом количества соавторов в каждой конкретной работе (например, если данный автор написал 8 статей, причем 5 из этих статей было написано двумя авторами, 1 статья – тремя авторами и 2 – четырьмя, то «Papers/author» = $5/2 + 1/3 + 2/4 \approx 3.33$).

8) **Authors/paper** – средневзвешенное количество авторов, приходящееся на одну статью данного исследователя, например, если он написал 8 статей, причем 5 из этих статей было написано двумя авторами, 1 статья – тремя и 2 – четырьмя, то «Authors/paper» = $(5 \cdot 2 + 1 \cdot 3 + 2 \cdot 4)/8 \approx 2.63$.

9) Индекс Хирша (**h-index**), равный максимальному количеству h наиболее цитируемых публикаций данного автора, таких, что каждая из этих h публикаций процитирована не менее h раз (например, если автор А имеет всего три статьи, процитированные, соответственно, 0, 2 и 4 раза, то его h-индекс равен 2, поскольку есть 2 статьи, процитированные не менее 2 раз каждая, но нет 3 статей, процитированных не менее 3 раз каждая;

¹³ Два других пика интерпретировать сложнее. Пока мы склоняемся к следующему объяснению. Возможно, это связано с предпочтениями конкретных авторов – в каких журналах печататься. Больше всего сотрудники факультета почвоведения публикуются в двух журналах – «Почвоведение» и «Вестник МГУ, серия Почвоведение» (см. рис. 5). При этом цитируемость первого намного выше, чем второго. Следовательно, сотрудники из группы, больше уделяющей внимание публикациям в «Почвоведении» будут формировать на графике плотности распределения один пик, а те, кто, в основном, публикуется в «Вестнике МГУ» - другой. Заметим, что сказанное выше – это только наглядный пример. Мы проиллюстрировали только идею возможного объяснения. На самом деле мы не считаем, что правый пик действительно соответствует публикациям в «Почвоведении» (скорее, он соответствует публикациям в международных журналах). Точное разрешение поставленного вопроса (об интерпретации пиков) возможно, но требует огромных трудозатрат – необходимо проанализировать публикации каждого сотрудника попавшего в конкретный пик, после чего станет однозначно понятно, почему эти публикации получают именно столько ссылок (таким образом, можно объяснить положение пика), а также понять - что это за группы сотрудников (таким образом можно объяснить высоту пика и его площадь).

подробности см. в [Hirsch, 2005]). Для примера на рис. 3А приведена эмпирическая плотность распределения сотрудников факультета почвоведения МГУ по индексу Хирша. В силу своей целочисленности индекс Хирша очень удобен для подразделения ученых на качественные классы. Учитывая статистику, приведенную в [Hirsch, 2005] (показывающую, что, например, физики, получившие Нобелевскую премию в течение последних примерно 20 лет, имели индекс Хирша, начиная приблизительно от 20), можно не без основания считать величайшими учеными современности (4%) тех сотрудников факультета, у которых индекс Хирша превышает 6 (максимальный на сегодняшний день индекс на факультете составляет 10!). К сожалению, только сотрудники двух кафедр факультета смогли достичь таких показателей.

10) Индекс Egghe (**g-index**), оказывающийся «менее грубым», чем h-индекс в том смысле, что он может различаться у двух исследователей, имеющих равные h-индексы (развивая предыдущий пример, допустим, что мы хотим сравнить автора А с автором Б, также имеющим всего три статьи, процитированные, соответственно, 0, 2 и 44 раза; очевидно, что автор Б – более «выдающийся ученый», однако, его h-индекс = 2 точно так же, как и у автора А, но вот g-индекс у автора А составляет 2, а у автора Б – 3);

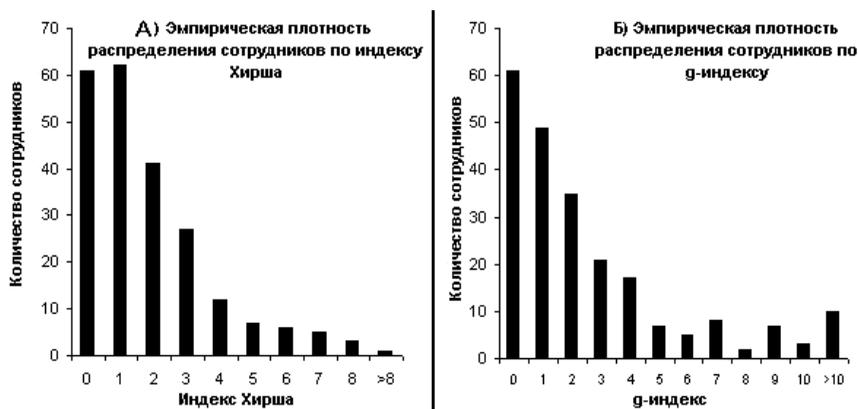


Рис. 3. Распределение сотрудников факультета Почвоведения МГУ по индексам цитирования Хирша и Egge.

11) для вычисления g-индекса статьи данного автора ранжируются на основании убывания количества ссылок следующим образом: на 1-ое место помещается статья, получившая максимальное количество ссылок, на второе – статья, получившая максимальное количество ссылок среди оставшихся статей и т.д.; g-индекс принимается равным максимальному номеру g статьи (считая от первой, имеющей максимальное количество ссылок), такому, что суммарное количество ссылок на первые g статей

составляет величину, не меньшую, чем g^2 (например, для автора А g -индекс = 2 потому, что расположив статьи в следующем порядке: 1-ая статья – это статья с 4 ссылками, 2-ая статья – с 2 ссылками, 3-я – с 0 ссылок, мы замечаем, что g могло бы быть равно 1, поскольку первая одна статья имеет 4 ссылки, что не меньше $1^2 = 1$, также g могло бы быть равно 2, поскольку первые две статьи в совокупности имеют $4 + 2 = 6$ ссылок, что не меньше чем $2^2 = 4$, но g не может быть равно 3, поскольку первые три статьи в совокупности имеют $4 + 2 + 0 = 6$ ссылок, что меньше чем $3^2 = 9$, теперь из возможных значений g выбираем максимальное и окончательно устанавливаем, таким образом, что $g = 2$; в случае же автора Б, рассуждая аналогично, мы приходим к выводу, что $g = 3$, поскольку для первых трех статей обнаружим, что в совокупности они имеют $44 + 2 + 0 = 46$ ссылок, что не меньше, чем $3^2 = 9$, и таким образом, возможными значениями g являются 1, 2 и 3, максимальное из которых – 3, оно-то и является окончательным значением g -индекса); дальнейшие подробности см. в [Egghe, 2006]). Для примера на рис. 3Б приведена эмпирическая плотность распределения сотрудников факультета почвоведения МГУ по g -индексу. В силу своей целочисленности g -индекс также удобен для подразделения ученых на качественные классы.

12) Современный индекс Хирша (**hc-index**), равный максимальному количеству h^c наиболее цитируемых публикаций данного автора, таких, что для каждой из этих h^c публикаций выполняется соотношение $S^c(i) \geq h^c$, а любая из оставшихся публикаций имеет $S^c(i) < h^c$; величина $S^c(i)$ для i -ой статьи определяется следующим образом: $S^c(i) = \gamma \cdot C_i / (Y_{now} - Y_i + 1)^\delta$, где C_i – количество ссылок на i -ую статью, опубликованную в Y_i год, Y_{now} – текущий год; $\gamma = 4$; $\delta = 1$ [Sidiropoulos et al., 2006]. Эмпирически было установлено, что в “Publish or Perish” дробная величина $S^c(i)$ предварительно округляется до ближайшего целого и только потом сравнивается с h^c .

13) Индивидуальный индекс Хирша (**hI-index**), равный отношению индекса Хирша h к среднему числу авторов тех наиболее цитируемых h статей, которые и определили такой индекс Хирша [Batista et al., 2006].

14) Нормализованный индивидуальный индекс Хирша (**hI,norm**), равный максимальному количеству $h_{I,norm}$ наиболее цитируемых публикаций данного автора, таких, что для каждой из этих $h_{I,norm}$ публикаций выполняется соотношение $Q(i) \geq h_{I,norm}$, а любая из оставшихся публикаций имеет $Q(i) < h_{I,norm}$; здесь величина $Q(i)$ для i -ой статьи определяется следующим образом: $Q(i) = C_i / A_i$, где A_i – количество авторов i -ой статьи.

15) **AWCR** - взвешенная скорость цитирования (Age-Weighted Citation Rate), равная сумме (по всем статьям¹⁴) отношений количества ссылок к возрасту статьи.

¹⁴ Первоначально взвешенная скорость цитирования была введена в заметке Jin B. 2007. The AR-index: complementing the h-index // *ISSI Newsletter*, 3(1), p. 6.

- 16) **AW-index** – квадратный корень из **AWCR**; AW-index имеет тот же смысл, что и индекс Хирша, но лишен многих недостатков последнего.
- 17) **AWCRpA** - взвешенная скорость цитирования в расчете на автора (per-Author Age-Weighted Citation Rate), равная сумме (по всем статьям) отношений количества ссылок к возрасту статьи и к количеству соавторов.
- 18) а-индекс (**Hirsch a**) – отношение суммарного количества ссылок, полученных данным автором, к квадрату индекса Хирша; а-индекс характеризует «разброс» качества статей данного автора. Большое значение а-индекса подразумевает, что на некоторые публикации данного исследователя было сделано существенно больше ссылок, чем на остальные его работы [*Sidiropoulos et al.*, 2006].
- 19) «Скорость зарабатывания» индекса Хирша данным автором (**Hirsch m**) – отношение индекса Хирша к количеству лет, прошедших с опубликования данным автором самой ранней работы, вошедшей в БД “Publish or Perish”.
- 20) **Cites/paper median** - медиана количества цитирований публикаций данного исследователя.
- 21) **Cites/paper mode** - мода количества цитирований публикаций данного исследователя.
- 22) **Authors/paper median** – медиана количества авторов, приходящихся на одну статью данного исследователя.
- 23) **Authors/paper mode** - мода количества авторов, приходящихся на одну статью данного исследователя.

Недостатки системы “Publish or Perish”

К сожалению, система “Publish or Perish” не лишена недостатков, свойственных вообще всем системам подобного рода. Во-первых, для исследователей, имеющих распространенные фамилии, в автоматическом режиме невозможно подсчитать индивидуальные ЛИЦ, поскольку в “Publish or Perish” запрос определяется, главным образом, инициалами и фамилией. Но, например, запросу «AV Ivanov» может соответствовать как Александр Васильевич Иванов, так и Антон Валерьевич Иванов - мы специально приводим в качестве примера двух реально существующих ученых, одновременно работающих не просто в одной и той же области науки (!) – в почвоведении, но вместе работающих на одном факультете (!!), более того – на одной кафедре (!!!). Радикальное решение вопроса состоит в том, чтобы иметь список трудов конкретного исследователя и в ручном режиме вычеркивать найденные системой “Publish or Perish” названия публикаций, не соответствующие указанному списку (к счастью, в системе предусмотрена возможность такой ручной

где предлагалось осуществлять суммирование только по наиболее цитируемым статьям, определяющим индекс Хирша, однако в “Publish or Perish” суммирование производится по всем статьям.

корректировки). Для менее распространенных фамилий возможно использование автоматических фильтров по соавторам. Например, в БД "Publish or Perish" обнаруживается несколько исследователей по запросу «NA Shnyrev», однако сотрудник факультета Почвоведения МГУ Н.А. Шнырев никогда не писал статей в соавторстве с I.R. Bariliak, с которым есть совместные работы у Н.А. Шнырева-биолога. По фильтру "Bariliak" оказывается возможным выделить работы Н.А. Шнырева-«почвоведа».

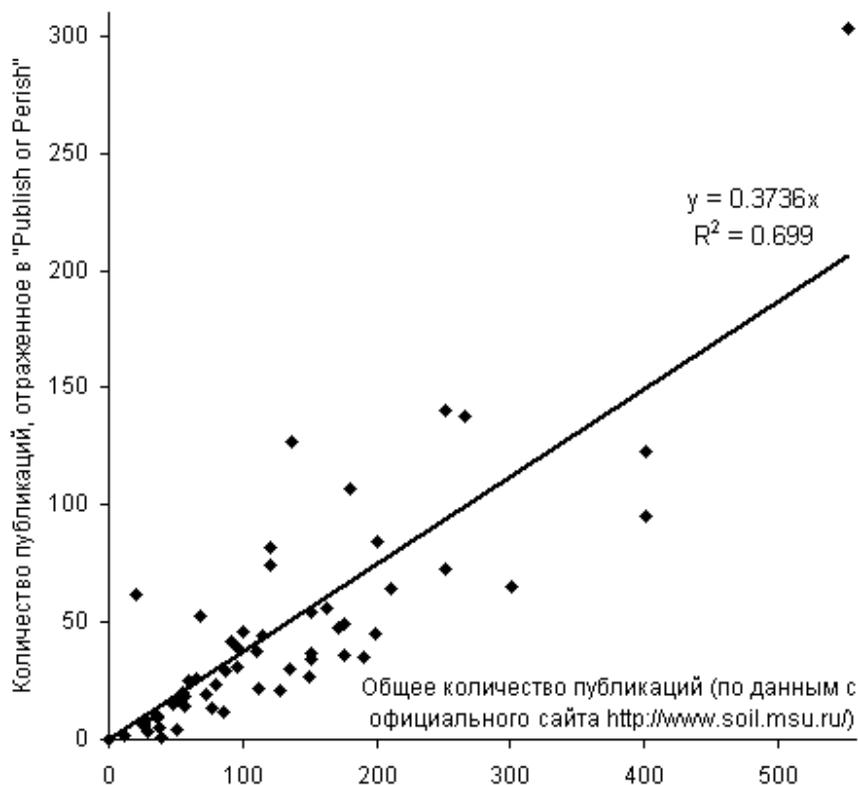


Рис. 4. Количество публикаций сотрудников факультета почвоведения МГУ на 07.09.2008 (каждая точка на графике представляет конкретного сотрудника).

Во-вторых, различное написание авторами названий некоторых журналов при цитировании приводит к искусственному увеличению количества публикаций у цитируемого автора и, следовательно, к занижению такого «показателя качества», как СКЦ/П. Например, журнал

“Почвоведение” публикуется на английском языке под названием «Eurasian Soil Science». Однако иногда при цитировании этого журнала используется написание «Pochvovedenie». Представим, что некий автор опубликовал за свою жизнь лишь одну статью, на которую было 2 ссылки (таким образом, СКЦ/П = 2 ссылки / 1 статью = 2 ссылки/статью).

Однако, если один раз на его статью сослались, как на опубликованную в журнале «Eurasian Soil Science», а другой раз указали журнал «Pochvovedenie», то он «становится» автором двух статей; но поскольку количество ссылок остается прежним, то качество работы данного автора «уменьшается» в 2 раза (2 ссылки / 2 статьи = 1 ссылка/статью). Есть и другие источники генерирования фиктивных публикаций. Например, иногда при цитировании названия русскоязычных публикаций переводятся на английский язык, а иногда русские слова просто записываются латиницей.

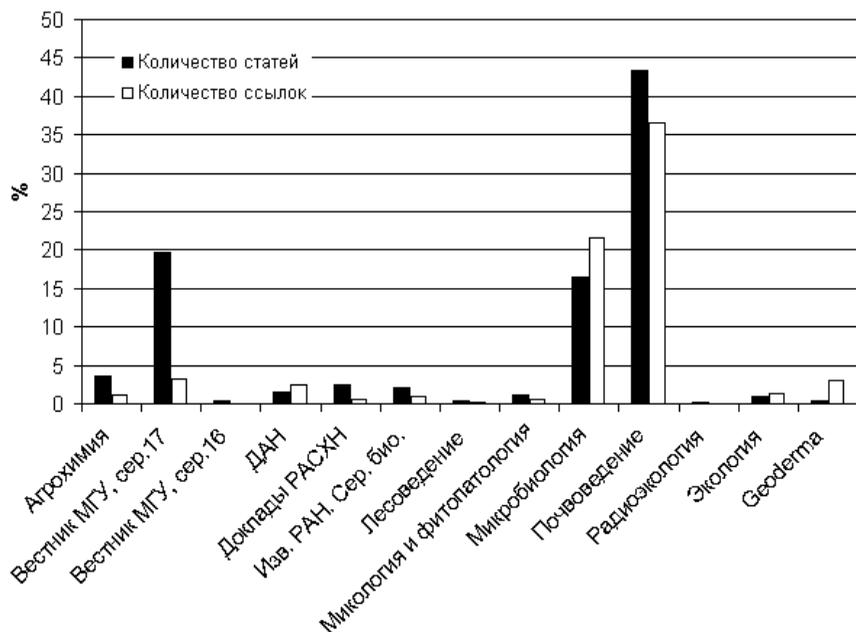


Рис. 5. Распределение основных публикаций сотрудников факультета почвоведения МГУ (на 07.09.2008) по журналам.

В-третьих, необходима осторожность при транскрипции имен отдельных ученых, которые в переводе с кириллицы на латиницу, используемую в БД, записываются по-разному [Касимова, 2004]. В качестве примера приведем некоторые написания фамилии известного

почвоведом проф. Л.О. Карпачевского, причем подчеркнем, что статьи с такими написаниями фамилии действительно имеются в БД “Publish or Perish”: Карпачевski (не менее 1 работы), Карпачевskii (не менее 79 работ), Карпачевskiy (не менее 15 работ), Карпачевsky (не менее 13 работ).

В-четвертых, поскольку источники (по крайней мере, российские) покрываются БД “Publish or Perish” неравномерно по времени (наиболее полно учитываются лишь публикации последних лет), то показатель Cites/year может оказаться занижен у корифеев отечественной науки, поскольку их главные работы иногда оказываются не внесенными в БД. К счастью, БД “Publish or Perish” непрерывно развивается и сравнение результатов, полученных в 2007 и 2008 гг. по одним и тем же запросам, показывает, что за истекший год БД пополнилась многими старыми классическими работами. Не может не вызывать радости (и удивления!) тот факт, что в настоящее время по запросу “VV Dokuchaev” система выдает, среди прочих, ссылки, например, на «Abridged historical account and critical examination of the principal soil classifications existing» (“Transactions of the Petersburg Society of Naturalists”, 1879), «О podzole» (“Tr Vol'nogo ekonomicheskogo obshchestva”, 1880), «Russian Chernozem» (St. Petersburg, 1883) и ряд других старых классических работ.

К счастью, “Publish or Perish” позволяет сохранить в различных удобочитаемых файлах все найденные в БД ссылки на статьи по данному запросу. В дальнейшем возможно обработать эти файлы вручную или при помощи несложных программ, учитывающих все вышеперечисленные недостатки. Для 125 исследователей факультета Почвоведения МГУ, имеющих СКЦ/П от 0.2 до 7.2 ссылок/публикацию, мы обработали исходные данные при помощи исправленных программ и обнаружили, что СКЦ/П возросло в среднем на $26 \pm 20\%$ по сравнению с тем, что выдавала “Publish or Perish” (для конкретных исследователей величина прироста СКЦ/П может сильно различаться). Правда, и здесь следует отдать должное создателям системы: сравнительный анализ, проведенный нами в 2008 г. для ряда цитирований сотрудников, по которым в 2007 г. “Publish or Perish” допускала ошибки, показал, что теперь многие ошибки исправлены.

К сожалению, есть и неустраняемые погрешности системы. Повидимому, независимо от реального числа авторов той или иной статьи, в БД “Publish or Perish” попадает не более пяти авторов. Таким образом, при наукометрическом обчете исследователя, который писал свои работы преимущественно с большим числом соавторов, все параметры,

принимающие в расчет количество авторов статьи, будут вычислены неверно. Исправить этот недостаток системы возможно только путем очень больших трудозатрат – по краткой ссылке, указанной в БД “Publish or Perish”, следует найти полную библиографическую ссылку и по ней установить истинное количество авторов данной публикации.

Как повысить личный индекс цитирования?

На рис. 6 приведена зависимость реальной цитируемости статей от ИФ журнала для некоторых сотрудников факультета почвоведения МГУ. К сожалению, ISI рассчитывает классический ИФ лишь для небольшого числа российских журналов, причем почвоведы и экологи публикуют свои работы лишь в 5-6 из них. Для десятков других отечественных журналов, в которых обычно печатаются статьи по экологии и почвоведению, ИФ ISI не известен. Однако для очень многих из них известен ИФ РИНЦ¹⁵, рассчитанный ООО "НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА" (г. Москва) в связи с выполнением государственного контракта по созданию Российского Индекса Научного Цитирования (РИНЦ). На рис. 6 мы приводили средний (за период 1994-2007 гг.) ИФ ISI для тех журналов, для которых он существует, а для остальных журналов использовали ИФ РИНЦ на 19.09.2007.

Из рис. 6 видно, что цитируемость статьи в среднем почти прямо пропорциональна ИФ журнала (Цитируемость ~ ИФ^{0,96}), однако разброс цитируемости очень велик¹⁶. Причем в подавляющем большинстве

¹⁵ РИНЦ – это многофункциональная информационная система, в которой обрабатывается библиографическая информация, аннотации и пристатейные списки цитирования из российских научных журналов. Поисковые и информационные сервисы базы данных эффективно реализуют различные виды поиска информации, анализируют и рассчитывают индексы цитирования отдельных авторов, научных коллективов и организаций, тематических направлений, импакт-факторы журналов. Авторам предоставлена возможность самостоятельно вводить и корректировать информацию о том, что, где и когда они опубликовали, используя для этой цели интерфейс Единого реестра научных публикаций. Используя систему управления ссылками, возможно без труда выходить не только на полные тексты статей, которые обрабатываются в самом РИНЦ, но и на статьи, которые цитировались в этих публикациях

¹⁶ Этот разброс имеет, как минимум, две составляющих. Безусловно, «качество» разных статей в одном и том же журнале различно, и это – один источник разброса. Но даже если бы «качество» статей было абсолютно одинаковым, разброс все равно существовал бы. Ведь, вычисляя «реальную цитируемость», мы для разных журналов берем различные промежутки времени: от момента опубликования первой статьи по настоящее время (статья должна удовлетворять двум условиям: во-первых, она найдена нами в списках литературы сотрудников, до сих пор работающих на факультете Почвоведения МГУ и, во-вторых, она имеется в БД “Publish or Perish”). Понятно, что на протяжении этого промежутка времени ИФ мог существенно меняться (причем по-разному – для различных журналов), а мы используем

случаев реальная цитируемость *ниже* той, которую можно было бы ожидать, исходя из ИФ (что вполне естественно, поскольку ИФ представляет собой линейную модель, в то время как на длительных промежутках времени накопление ссылок происходит в соответствии с логистической кривой). В любом случае, фактом остается то, что количество ссылок на конкретную статью может на порядок (!!!) отличаться от ИФ. Это явление хорошо известно, однако, подчас из него делают совершенно неверные выводы.

Так, проф. Сеглен указывает, что 15% наиболее цитируемых статей дают 50% цитирований; 50% наиболее цитируемых статей дают 90% цитирований. Т.е. наиболее цитируемая половина статей цитируется, в среднем в 10 раз (на порядок!) чаще, чем наименее цитируемая половина. Отсюда проф. Сеглен в своей статье с характерным названием «Почему импакт-фактор журналов не должен использоваться для оценки исследования» делает следующий вывод: так как вклад небольшого числа высоко цитируемых статей определяет значение импакт-фактора журнала, то приписывание такого же коэффициента (ИФ журнала) всем статьям скрывает огромное различие в реальной цитируемости (даже нецитируемым статьям придается полный вес!) [Vibe and Jakobsson, 2005], что, казалось бы, является точной противоположностью того, что предполагали достичь, например, вводя в институтах РАН надбавку за публикации, пропорциональную именно ИФ журналов, в которых публиковался конкретный сотрудник в течение последних двух лет.

Конечно, если речь идет об одной единственной статье у каждого исследователя, то, если судить лишь по ИФ журнала, в котором они опубликовали эту свою единственную статью, какие-то авторы будут переоценены, а какие-то – недооценены. Но в реальности у большинства авторов имеется множество публикаций, какие-то работы получили больше ссылок, чем можно было ожидать из ИФ журнала, какие-то – меньше, но в среднем, как видно из рис. 6, реальную цитируемость действительно можно принимать пропорциональной ИФ журнала.

в лучшем случае усредненное за 9 лет значение. Кроме того, следует помнить об описанном в [Писляков, 2005] явлении: в ряде дисциплин «устаревание» знания очень велико и публикации теряют свою актуальность быстро, в то время как в других отраслях исследовательские работы могут цитироваться из года в год. И те и другие издания, конечно, могут иметь близкие величины ИФ, но реальная цитируемость у них будет сильно различаться если она рассчитывается по периоду времени, существенно превышающему цитатное окно ИФ (т.е. 2 года).

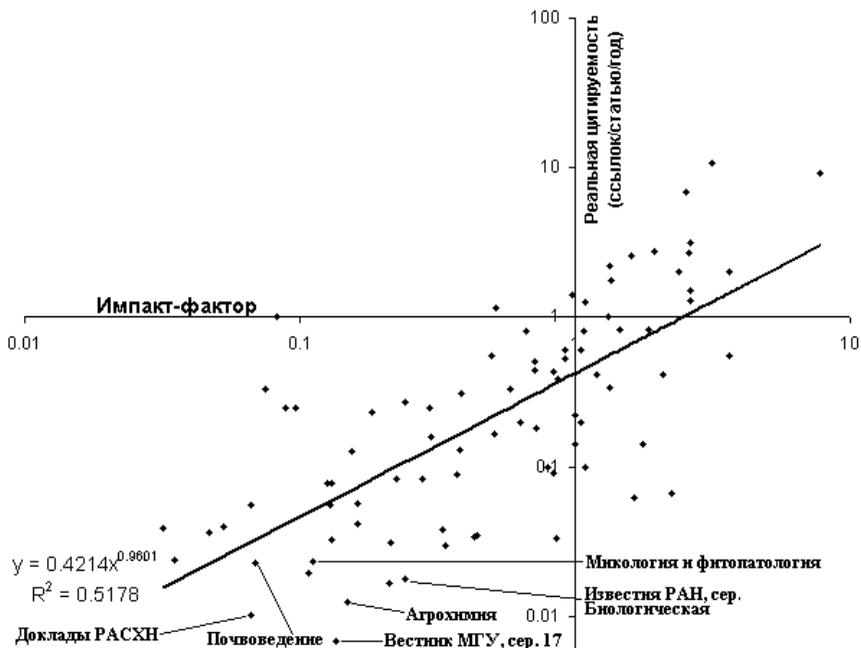


Рис. 6. Взаимосвязь реальной цитируемости статьи с импакт-фактором журнала, в котором она опубликована.

С огромным сожалением, следует отметить, что большинство основных журналов, в которых публикуются сотрудники факультета Почвоведения МГУ («Агрохимия», «Вестник МГУ, сер. Почвоведение», «Доклады РАСХН», «Известия РАН, серия биологическая», «Микология и фитопатология», «Почвоведение») имеют самую низкую реальную цитируемость¹⁷ – см. левый нижний угол рис. 6.

В любом случае, очевидно, что необходимо стремиться к публикации работ в журналах с высокими ИФ. Действительно, посмотрим на рис. 5: например, в журнале «Вестник МГУ, сер. Почвоведение»

¹⁷ Точнее говоря, самую низкую *положительную* цитируемость. На публикации в некоторых журналах за многие годы так и не было получено ни одной ссылки. Впрочем, нулевая цитируемость объясняется, скорее всего, не низким качеством журналов, а чисто вероятностными причинами: в этих журналах сотрудники факультета опубликовали единичные работы (в некоторых журналах за многие годы, даже за десятилетия, опубликована лишь одна статья, принадлежащая перу сотрудников факультета почвоведения), следовательно при цитируемости значительно меньше 1 (а именно такова цитируемость в большинстве случаев, как это хорошо видно из рис. 6), вероятнее ожидать, что наугад выбранная статья (или даже 2-3 статьи) окажутся не имеющими ссылок. Сколь-нибудь надежно о реальной цитируемости можно судить лишь на большом статистическом материале, как это имеет место лишь для нескольких журналов, таких как «Почвоведение», «Вестник МГУ, сер. Почвоведение», «Микробиология», «Агрохимия» и некоторых других.

(ИФ РИНЦ = 0.135) сотрудники факультета Почвоведения публикуют 20% своих работ, но это приносит им лишь 3% ссылок, в то время как те же 3% ссылок приносит журнал «Geoderma» (ИФ ISI≈1.34), в котором опубликовано лишь... 0.3% статей сотрудников факультета!

Конечно, требования к статьям, предъявляемые со стороны редакции и рецензентов, тем выше, чем выше ИФ журнала. Вполне очевидной является следующая рекомендация молодому исследователю: **начать публиковаться в журналах, имеющих небольшие ИФ, но каждую следующую статью стараться опубликовать в журналах с бо́льшим ИФ, чем предыдущую.**

Хорошо известно, что ЛИЦ работ, выполненных в соавторстве с иностранцами, выше, чем мононациональных. Сама возможность опубликовать работу в престижном зарубежном журнале в соавторстве с коллегой-иностранцем (в особенности, англоязычным) легче, чем в чисто российском варианте [Касимова, 2004]. **Для российских авторов очень часто основной проблемой при попытке опубликовать свою работу в международном журнале, имеющем высокий ИФ, является недостаточное знание иностранного (чаще всего - английского) языка.** Понятно, что участие в интернациональном коллективе соавторов очень часто решает эту проблему, т.к. в таком коллективе обычно оказываются один или несколько носителей языка.

Список литературы

Арутюнов В.С. 2007 // *Ж. Рос. хим. об-ва им. Д.И. Менделеева*, **LI(3)**, 5-15.

Васьковский В.Е. 1993 // *Химия и Жизнь*, № 9, 42 - 45.

Касимова Р.Г. 2004. Библиометрические базы данных как инструмент научного менеджмента. (Статья доступна по адресу: <http://www.ecsocman.edu.ru/images/pubs/2004/04/22/0000155817/bd.pdf>).

Кохановский В.П., Лешкевич Т.Г., Матяш Т.П., Фатхи Т.Б. 2007. Основы философии науки. – Ростов н/Д: Феникс. – 608 с.

Писляков В.В. 2005. Наукометрические методы и практики, рекомендуемые к применению в работе с Российским Индексом Научного Цитирования // Приложение к отчету за 2005 г. по проекту «Разработка системы статистического анализа российской науки на основе данных Российского Индекса Цитирования» (государственный контракт от 31 мая 2005 г. № 02.447.11.7001). – М.: ООО НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА.

Сонькин В.Д. 2005. Реформа науки: взгляд изнутри. (Статья доступна по адресу: <http://www.lenta.ru/articles/2005/10/17/science1/>).

Талызин А.В. 2006 // Интервью сайту ScienceRF -Национальный информационный центр по науке и инновациям. (Статья доступна на: <http://www.researcher-at.ru/index.php?option=content&task=view&id=344>).

Batista P.D., Campiteli M.G., Kinouchi O., Martinez A.S. 2006 // *Scientometrics*, **68**(1), 179-189.

Egghe L. 2006 // *Scientometrics*, **69**(1), 131-152.

Hirsch J.E. 2005 // *Proc. Nat. Acad. Sci.*, **102**, 16569.

Sidiropoulos A., Katsaros D., Manolopoulos Y. 2006. Generalized h-index for Disclosing Latent Facts in Citation Networks. (Препринт доступен на <http://arxiv.org/abs/cs/0607066v1>).

Vibe A.-M., Jakobsson A. 2005 // World Library and Information Congress: 71th IFLA General Conference and Council “Libraries – A voyage of discovery”, August 14th – 18th 2005, Oslo, Norway. (Перевод на русский язык доступен по адресу

http://www.ifla.org/IV/ifla71/papers/008r_trans-Vibe_Jakobsson.pdf).

**ELEMENTS OF SCIENTOMETRICS
IN THE SOIL SCIENCE AND ECOLOGY
(WITH EXAMPLES FROM MOSCOW STATE UNIVERSITY)**

Glagolev M.V., Yanin M.V.

We present some basic concepts of scientometrics and free Internet-system “Publish or Perish” which allow to calculate many “citation indexes” for each of scientists. Main errors and difficulties of “Publish or Perish” are discussed. All questions are considered at examples from scientific publication activity of Moscow State University Soil Science Faculty.