НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Измерения науки

№ 03 (03), 2016

ОКНА АКАДЕМИЧЕСКОГО РОСТА

ПРИЛОЖЕНИЕ К ИНФОРМАЦИОННОМУ БЮЛЛЕТЕНЮ

InCites vs SciVal: сравниваем продвинутые наукометрические пакеты из подписки НИУ ВШЭ

Пару месяцев назад у Вышки появился доступ к продвинутому аналитическому наукометрическому пакету компании Thomson Reuters – InCites. Годом ранее Вышка приобрела SciVal Benchmarking – аналогичный модуль производства Elsevier. В этом материале мы рассказываем о назначении и функционале этих интересных инструментов, об их особенностях и ограничениях. Доступ к SciVal и InCites открыт всем сотрудникам изнутри сети НИУ ВШЭ, призываем попробовать их в действии.

В общем случае оба продукта представляют собой онлайн-сервисы по выводу многообразия наукометрических индикаторов для подбираемых пользователями наборов сущностей – авторов, организаций, стран и т.д. Они снабжены встроенными средствами визуализации данных, которые облегчают и ускоряют анализ, а также позволяют экспортировать таблицы с показателями. Ценность SciVal и InCites обусловлена коммерческим характером деятельности Thomson Reuters и Elsevier – по сути, это аналитические надстройки над Web of Science (WoS) и Scopus, которые вынесены в отдельные дорогие продукты. При этом в самих WoS и Scopus возможности для аналитики и экспорта специально существенно ограничены. Забегая вперед, заметим, что и в SciVal с InCites есть множество суровых лимитов на выгрузку данных, поэтому для крупномасштабных исследований все равно необходимо закупать выгрузки, подключать API или вручную выкачивать сырые данные кусочками разрешенного размера из WoS и Scopus. Напоминаем: любое использование роботов-автокликеров для скачивания информации из этих баз запрещено.

InCites и SciVal позиционируются как инструменты для аналитической поддержки управления университетом, оценки результативности, стратегического планирования. Прежде всего они предназначены для сравнения разных типов сущностей по разным наборам показателей, в том числе нормализованных, учитывающих различия в средней цитируемости по областям науки, престижность ссылающихся журналов и т.д., а также для изучения всех этих показателей в динамике. Полагаем, такие сравнения могут быть интересны не только руководству вуза, но и большинству его сотрудников.

Помимо аналитического функционала есть и чисто практический: авторы публикаций в Scopus могут редактировать свой профиль прямо в SciVal, добавляя непривязанные публикации, причем рассмотрение заявки на внесение коррективов происходит в два раза быстрее, чем при обращении из Scopus. Также можно проверить, попадают ли определенные публикации исследователя в узкую тематику, учитываемую тем или иным предметным рейтингом QS, что может быть важно для продвижения в программе «5-100». Каждый сотрудник в InCites по фамилии может найти свои публикации и привязать их к идентификатору Researcher ID.

Начнем с **InCites,** которая доступна сотрудникам по адресу https://incites.thomsonreuters.com.

Первое, что необходимо отметить: база построена на данных WoS, однако они могут не совпадать с теми, что доступны через интерфейс WoS. В InCites входят только Science Citation Index, Social Science Citation Index, Arts & Humanities Citation Index, все версии Conference Proceedings Citation Index и Book Citation Index, т.е. не входят Emerging Sources Citation Index и национальные индексы, включая Russian Science Citation Index. В момент написания этой заметки (16 сентября) в InCites загружены данные WoS на 27 мая. Помимо библиометрических данных в системе есть показатели по кадрам и финансам из Global Institutional Profiles –

масштабной базы вузов и научных организаций, создававшейся под рейтинг Times Higher Education World University Ranking. К сожалению, по крайней мере по российским организациям точность этих данных зачастую вызывает сомнения, поэтому в этом обзоре мы концентрируемся на библиометрии.

Мы подробно разбирали отличия WoS от Scopus ранее (https://okna.hse.ru/news/160270181.html), поэтому сейчас лишь напомним, что в WoS отбор гораздо строже, т.е. качество индексируемых журналов выше, а их число ниже, чем в Scopus. Это большое достоинство и одновременно большой недостаток, который всегда надо иметь в виду при выборе одного из инструментов.

Объектами изучения (сущностями) в InCites являются исследователи, организации, регионы, источники, области исследования, каналы финансирования (грантовые фонды). Все эти сущности уже внесены в систему, но точность привязки из-за проблемы унификации названий и имен сильно разнится, особенно это касается уровня отдельных авторов. Увы, многие из них распались на 5-6 записей, объединить которые в единые профили не получится – для этого надо покупать отдельный продукт, который все равно позволит работать только с профилями сотрудников НИУ ВШЭ. Это связано с тем, что в самом WoS, в отличие от Scopus, авторских профилей нет.

Для каждого объекта можно вывести целый ряд показателей, часть из которых доступна для визуализации. Здесь представлен широкий спектр различных типов графического отображения данных: временные ряды, сети соавторства (увы, только для авторов), столбчатые и радиальные диаграммы, treemaps, географические карты. Также есть возможность добавлять или убирать нужное количество объектов на график с помощью кнопки «Pin to top». Все показатели можно экспортировать в csv в динамике, но установлен лимит – 50 тысяч объектов.

Имеется развитая система фильтров для учитываемых в расчете индикаторов публикаций: по годам, типам (в любых комбинациях), странам, организациям, тематикам. Сами получающиеся списки-рейтинги объектов также можно фильтровать по пороговым значениям.

В InCites доступны 14 классификаторов научных тематик, из которых 11 схем представляют собой переходники на внешние классификаторы (например, Organization for Economic Cooperation and Development Fields of Science (OECD FoS)), а три являются разработками Thomson Reuters (WoS Subject Categories – узкая категоризация, состоит из 252 тематических категорий в естественных, социальных и гуманитарных науках, журнал может иметь одну и более тематическую категорию, Essential Science Indicators – широкая категоризация, состоит из 22 тематических категорий только в естественных и социальных науках, журнал может принадлежать только одной предметной области, GIPP – очень широкая категоризация, состоит из 6 тематик, охватывая все области исследования, представляет собой агрегирование тематических категорий схемы WoS, использовавшееся в рейтинге университетов Times Higher Education World University Rankings).

Наиболее корректными являются сравнения различных достаточно крупных наборов публикаций: стран и организаций. Профили организаций в InCites соответствуют профилям Organization-Enhanced в WoS, для ведущих российских вузов они достаточно выверены, для НИИ, особенно отраслевых, могут быть лакуны. РАН существует и как совокупность НИИ, и как отдельные организации. Всего российских организаций в системе сейчас 160. Организации разделены на эклектичные категории: academic, government, corporate, health, museum и т.д.

Возможность вносить в InCites собственные наборы публикаций для анализа существует, но в довольно специфической форме: можно внести набор публикаций (до 50 тысяч, вышедших не ранее 1980 года) и считать показатели для всех изначально присутствующих в системе объектов внутри этого набора. Т.е. сравнить два набора с ходу не получится.

Ключевые показатели

Основное достоинство InCites по сравнению с WoS заключается в наличии продвинутых показателей цитируемости:

Category Normalized Citation Impact (Нормализованный по тематическим категориям индекс цитирования, CNCI) является отношением текущего количества цитирований публикации к ожидаемой норме цитирований публикаций того же типа, года выхода и предметной области, т.е. число цитирований публикации делится на среднюю цитируемость публикаций того же типа, года и тематики. Если тематик у публикации несколько, берется среднее. CNCI набора публикаций представляет собой среднее CNCI компонентов набора. При этом окно цитирования составляет весь промежуток с момента выхода статьи до последнего обновления InCites Dataset.

Average Percentile (Средний процентиль). Процентиль цитируемости для каждой публикации соответствует ее положению в рейтинге публикаций того же типа за тот же предметный год в той же тематической категории, причем 100 соответствуют нулевой цитируемости, а наиболее близкая к нулю – наивысшей. Если публикация относится к нескольким тематикам, присваивается наилучший (наименьший) процентиль. Средний процентиль равен среднему гармоническому процентилей публикаций в наборе.

% Documents in top 1% и % Documents in top 10% (процент документов, входящих в 1% или 10% наиболее цитируемых) показывает долю публикаций в наборе, входящих в топ-1% или топ-10% самых цитируемых публикаций в соответствующей тематической категории с учетом года выхода и типа публикации. Публикации, входящие в 1% самых цитируемых хотя бы в одной из укрупненных

тематических категорий Essential Science Indicators, называются Highly Cited Papers. Важно, что эти метрики в силу своего определения хорошо работают только на достаточно больших массивах публикаций, сравнивать по ним отдельных ученых необходимо крайне осторожно.

Journal-Normalized Citation Impact (Нормализованный по журналу индекс цитируемости, JNCI) показывает отношение числа цитирований, полученных данной публикацией, к среднему числу цитирований публикаций в выпустившем ее журнале за тот же период, т.е. лучше или хуже, чем в среднем по журналу, цитируется данная статья. Для набора публикаций опять же берется среднее арифметическое отдельных значений JNCI.

Принципиально важно, что все вышеупомянутые метрики не исключают самоцитирования, это один из их существенных недостатков в текущих условиях наукометрической горячки в управлении наукой.

Множество прочих интересных показателей InCites описано в соответствующем руководстве http://researchanalytics.thomsonreuters.com/m/pdfs/indicators-handbook.pdf. В их числе, например, процент работ в международном соавторстве.

Аналитическая система SciVal использует данные базы научного цитирования Scopus.

Объектами изучения являются организации и их группы, исследователи и группы исследователей, области исследования, страны и множество внесенных пользователями публикаций. В отличие от InCites, в системе SciVal не предусмотрена работа с источниками (журналы) как с объектом изучения. Полный комплект системы предполагает наличие четырех модулей: Overview, Benchmarking, Collaborations, Trends. Подписка ВШЭ распространяется только на использование ключевого модуля Benchmarking (цена на остальные, на наш взгляд, завышена, а функционал зачастую дублируется). При первом обращении к системе необходимо зарегистрироваться, если есть пароль\логин от Scopus, можно использовать их.

SciVal основывается на данных Scopus, обновляется более оперативно, чем InCites, – раз в две недели, – но включает не все типы публикаций (нам не встречались correction и некоторые другие второстепенные вещи). Более печально, что нельзя отдельно фильтровать документы типа book и book chapter, хотя в базе они учитываются в общем числе публикаций.

В отличие от InCites, в SciVal многие сущности хоть и внесены в базу, но в неявном виде, то есть с ходу строить рейтинги в ней не получится, а в случае авторов вообще потребуется внести все объекты для анализа вручную либо списком с указанием персональных идентификаторов. Зато SciVal позволяет объединять в группы для анализа различные организации и, главное, различных авторов. Можно внести в него в виде таких групп факультеты и кампусы, отдельные категории сотрудников типа ординарных профессоров, получателей внутренних грантов и т.д. Группы могут вкладываться друг в друга.

Вносить исследователя можно не только по идентификатору (Scopus Author ID, ORCID), но и по фамилии. Здесь же можно редактировать профиль исследователя, добавляя непривязанные публикации. Исследователи и группы исследователей по умолчанию доступны только пользователю, который их создал, также есть возможность предоставлять доступ другому пользователю по e-mail к созданным объектам. Группы организаций доступны после подтверждения всем пользователям мира, расшаривать их не требуется. Мы так внесли, например, группу вузов «5-100».

Предусмотрена возможность внесения своих наборов публикаций, порциями по 20 тысяч штук, которые можно объединять в наборы по 100 тысяч. Это позволяет в случае необходимости добиваться большей корректности: как и в InCites, готовые профили организаций в SciVal могут быть неточными, а исправить их изнутри системы нельзя.

Средства для визуализации и экспорта данных в принципе аналогичны InCites, но лимит на выгрузку значительно жестче – всего 20 тысяч позиций.

Показатели SciVal

Ключевые показатели SciVal напоминают индикаторы InCites, но есть ряд существенных отличий и дополнений:

Field-Weighted Citation Impact (Индекс цитирования, взвешенный по дисциплине, FWCI), аналог CNCI, представляет собой отношение числа цитирований публикаций объекта к среднему числу цитирований, полученных всеми остальными схожими публикациями в мире. Для расчета индикатора учитываются цитирования, полученные в год опубликования работы, а также в последующие три года, среднемировое значение равно единице.

Outputs in Top Percentiles (Публикации в верхних процентилях) показывает количество публикаций (или долю публикаций), которое входит в 1, 5, 10 или 25% наиболее цитируемых публикаций. Метрика рассчитывается для текущего года, не ранее 1 июля, аналогично InCites. Для этого индикатора предлагается также возможность его нормализации по предметной области, ее надо обязательно использовать, выбирая галочку «field-weighted».

Как и в InCites, исключить самоцитируемость в этих «продвинутых» индикаторах невозможно.

Индикатор **Publications In Top Journal Percentiles (Публикации в журналах из верхних процентилей)** считает число публикаций в первых 1, 5, 10 или 25% наиболее цитируемых журналов. Ранжирование журналов происходит по журнальным метрикам, учитывающим различия в цитировании по областям – SNIP или SJR. Этот индикатор крайне важен в практике управления наукой, так как позволяет оценить средний уровень работ и ожидаемую цитируемость в первые годы после публикации, когда считать реальную цитируемость еще невозможно.

H-indices (индекс Хирша и аналоги). Помимо стандартного индекса Хирша предлагается g-index, он лучше

учитывает вклад высокоцитируемых единичных выбросов. Если публикации отранжировать по цитируемости, то g – это такое максимальное число публикаций, что их цитируемость в сумме равна g2 или более. Есть и m-индекс – введенный самим Хиршем показатель, равный h-индексу, деленному на число лет с момента первой публикации автора. Все три индекса – h, g, m – можно посмотреть и без вклада самоцитируемости.

Доля публикаций в соавторстве (Collaboration) в случае SciVal – заметно более продвинутый индикатор: можно считать число и долю публикаций в иностранном соавторстве, национальном, институциональном, число и долю статей с одним автором. Причем в первых двух вариантах можно смотреть долю статей в соавторстве, нормализованную на среднее по области.

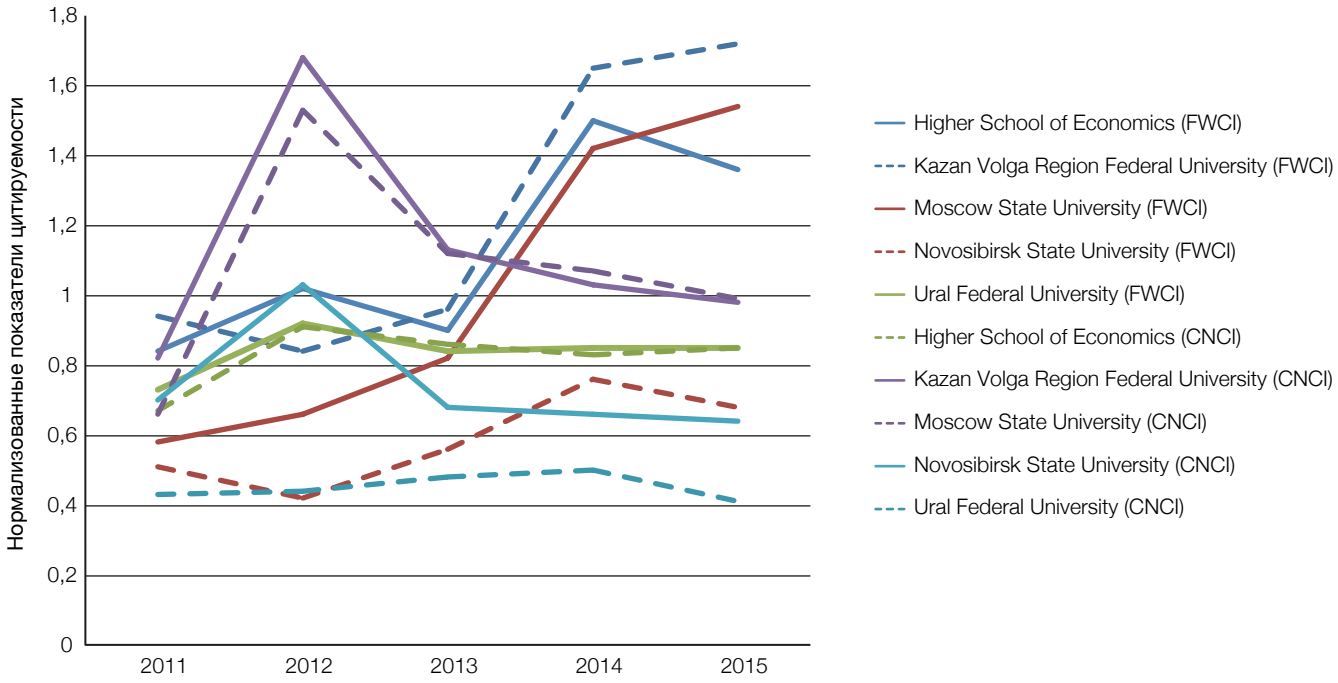
Какую же из двух систем выбрать? Единого ответа не существует, все зависит от целей и объектов изучения.

InCites содержит готовые списки организаций, авторов, журналов и стран, позволяющие быстро выявлять лидеров. Лимиты выгрузки показателей в нем более дружелюбные, выгружать в динамике можно сразу несколько показателей, а система визуализации сравнения объектов содержит больше вариантов графиков и позволяет одновременно анализировать не три, а значительно большее число метрик, включая уникальные типа JNCI. Фильтрация по типам публикаций и по тематикам реализована более гибко и удачно позволяет формировать комбинированные фильтры. При этом корректировка авторских профилей в нем затруднена, а внесение собственных наборов публикаций реализовано так, что сравнить их между собой очень непросто.

SciVal обладает в целом более удобным интерфейсом и четкой визуализацией. Впрочем, это дело вкуса, а вот несомненные его достоинства – это возможность корректировать авторские профили в интерфейсе, сравнивать собственные наборы публикаций и, главное, создавать любые группы авторов, а также метрики числа\доли публикаций в топовых журналах. Система расшаривания внесенных сущностей коллегам в SciVal также более развитая. SciVal с недавних пор включает также альтметрики – число просмотров из Scopus, однако эти данные пока столь же проблематично использовать в реальной оценке, как и кадровую и финансовую статистику из InCites.

В завершение мы публикуем прямое сравнение коронных показателей двух систем – CNCI InCites и FWCI SciVal – на примере пяти российских вузов. Объяснение расхождений для нескольких вузов кроется в том, что у них существенно выше средней доля публикаций в журналах Scopus, отсутствующих в WoS, при этом в одном случае именно в таких журналах сотрудники вуза особенно активно заняты самоцитированием. Так что в определенном смысле индикаторы InCites являются более устойчивыми.

Сравнение нормализованных показателей цитируемости



Выпускается НИУ ВШЭ. Редактор Светлана Фомина.
Материалы подготовлены при участии Ивана Стерлигова.
Контакты: okna@hse.ru, (495) 621-4514. Дата выпуска: 29.09.2016.
Подписаться на рассылку можно здесь: http://okna.hse.ru/subscribe