

Комментарий к обновлению данных в 2016 году

1. Какая информация доступна сейчас в Web of Science (WoS).

База **WoS** в настоящее время активно расширяет свой функционал, подключая все большее число приложений. С 2015 г. одним из таких приложений стал Russian Science Citation Index (RSCI). [Вот как это происходило](#).

Сведения о числе статей и цитировании каждого автора, извлекаемые из **WoS**, требуют указания не только опции, используемой при поиске и обработке данных, но и поискового массива, и с увеличением числа приложений зависимость результата от выбора массива резко усиливается. При поиске по опции **Basic Search** в массиве **All Databases** для каждой статьи **WoS** выводит справа сверху страницы общее число цитирований [N TimesCited](#), а также (чуть ниже на той же странице) указывает вклады разных приложений:

- [All Databases](#)
- [Web of Science Core Collection](#)
- **BIOSIS Citation Index**
- **Chinese Science Citation Database**
- **Data Citation Index**
- **Russian Science Citation Index**
- **SciELO Citation Index**

Известны примеры отличия цитирования в **All Databases** и **Core Collection** на порядок величины. Их причины в каждом конкретном случае требуют отдельного анализа. Обновление списков <http://expertcorps.ru/science/whoiswho> в 2015 г. впервые, в порядке эксперимента, проводилось по массиву **All Databases**, в попытке более полно отразить низкоцитируемые научные области и направления. Имеющиеся наблюдения (см. ниже) вынуждают нас пока отказаться от использования этого массива.

2. Опасности дублирования:

- Система идентификаторов **WoS** в принципе позволяет избежать дублирования статей, реферируемых как в **Core Collection**, так и в приложениях (или одновременно в разных приложениях). Это должно было бы обеспечить корректное определение выводимого в списке <http://expertcorps.ru/science/whoiswho> показателя **M_ref** (число статей, реферируемых в **WoS**). Однако, как выяснилось, эта система не спасает от дублирования при существовании версий (издаваемых на разных языках) одной и той же статьи.

- С появлением **RSCI** проблема такого дублирования числа статей усугубилась, поскольку возник большой массив русско- и англоязычных версий одних и тех же статей, не всегда идентифицируемых по **DOI**. Это может приводить к значительному завышению **M_ref** для авторов, активно публикующихся в российских журналах. Еще сложнее обстоит дело с показателем **M_all**, включающим скрытые ссылки - поскольку у соответствующих строк, выводимых по опции **Cited Reference Search**, вообще нет идентификаторов. Это увеличивает вероятность дублирования.

- Но наиболее острой является сопряженная проблема дублей цитирования, которые влияют на показатели **CI_tot**, **CI_7** и **h**. Чтобы выявлять дубли цитирования, нужно сопоставлять большие массивы цитирующих работ. Эту задачу своими средствами пытается решать **WoS**, но выборочная проверка показывает, что до хорошего решения еще далеко. В **RSCI** регулярным образом "размножаются" ссылки из одной и той же цитирующей статьи, путем замены одной ссылки на любую из версий цитируемой статьи двумя ссылками - на русско- и англоязычную версии. И **WoS** пока что не всегда это обнаруживает.

- Безусловно, аналогичная проблема существует и в **Core Collection**. Встречаются, хотя и редко, уважаемые международные журналы, издаваемые в двух версиях, и требующие от авторов указания обеих версий в ссылках. Например, [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1521-3773/homepage/2002_guideline.html#id2221813](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1521-3773/homepage/2002_guideline.html#id2221813)
When citing publications from *Angewandte Chemie*, please quote both the German and the International editions of this journal, starting with the International edition. Examples:
[1] a) H. J. Ache, *Angew. Chem. Int. Ed. Engl.* **1989**, *28*, 1; *Angew. Chem.* **1989**, *101*, 1; b) H. Frey, *Angew. Chem. Int. Ed.* **1998**, *37*, 2193; *Angew. Chem.* **1998**, *110*, 2313.

Выборочная проверка показывает, что цитирования в таком формате **WoS** учитывает дважды. Аналогичный формат встречается и в других журналах, и в конференционных сборниках (иногда по инициативе авторов). Например, в одной ссылке может содержаться ссылка на журнальный вариант (часто - "в печати") и на препринт той же статьи. При обработке в **WoS** такие случаи иногда трактуются как одна ссылка, а иногда как две. Так что и цифры из **Core Collection** не следует рассматривать как абсолютные. Но в случае с российскими журналами размножение ссылок производится самим **РИНЦ** даже в том случае, когда формат цитирования к дублированию не приводит.

3. Оргвыводы:

- Во избежание путаницы и в ожидании дальнейшего продвижения **WoS** в трудном деле оптимизации приложения **RSCI** в 2016 г. мы будем выводить в списке <http://expertcorps.ru/science/whoiswho> данные только по **Web of Science Core Collection** (включает, среди прочего, [Science Citation Index Expanded](#) и [Conference Proceedings Citation Index](#)). По-прежнему будет использоваться опция **Cited Reference Search**, которая позволяет отразить цитирование не реферируемых в **Core Collection** статей, если их цитируют в реферируемых там статьях. Просим отнестись с пониманием к тому, что у некоторых авторов цифры 2016 года могут оказаться ниже, чем 2015. В дальнейшем таких осложнений возникать не будет.

- Мы будем по-прежнему загружать из **WoS** весь массив **All Databases**, но соответствующие цифры не будут выводиться в списке на сайте до тех пор, пока не станет ясно, что проблема дублирования в цитировании русскоязычных публикаций в **WoS** будет сведена к минимуму.

4. Особенности организации обновления данных в 2016 г.

За время, прошедшее после переноса списков с scientific.ru на expertcorps.ru, число людей в списках увеличилось в 4.4 раза, а работа по обновлению данных производилась до 2015 года включительно все теми же силами, уже на пределе человеческих возможностей. Подробнее - здесь: <http://trv-science.ru/2016/01/26/project-expertcorps-2016/>

В 2016 г. решено внести следующие изменения в традиционный порядок обновления данных:

1. До появления каких-либо дополнительных ресурсов, позволяющих обеспечить работу со списками в прежнем режиме, обновление данных в общем случае проводить не ежегодно, а раз в два года.
2. Сохранить ежегодное обновление данных для всех, кто имеет **MyResearcherID** и регулярно актуализирует разметку статей в этом ресурсе. Соответствующий номер **MyResearcherID** нужно сообщить администратору проекта по адресу corr.lists@gmail.com или через индивидуальные окна в списках (открываются нажатием на фамилию).
3. Предоставить институтам, желающим ежегодно (или чаще) обновлять данные своих сотрудников в списках собственными силами, возможность использования имеющихся инструментов. В этом случае после подробного инструктажа и работы в тестовом режиме может быть создан селективный доступ к карточкам сотрудников института.