



# Кросс-ссылки в электронных информационных ресурсах

Писляков Владимир Владимирович

Начальник отдела информационных систем и электронных ресурсов библиотеки Государственного университета – Высшей школы экономики, г. Москва, Россия.

В работе описан процесс образования кросс-связей между различными электронными публикациями и введено понятие «библиогиперссылки», соединяющей в себе функции библиографического описания и гипертекстовой ссылки. Изложены возможные способы классификации кросс-ссылок, причем особое внимание уделено «внутренним» кросс-ссылкам, связывающим статьи, размещенные на одной платформе. Намечена связь между внутренними кросс-ссылками и «информационной замкнутостью», «самодостаточностью» электронного ресурса. Введена и подсчитана на тестовом примере количественная мера такой самодостаточности — «коэффициент интрацитирования».

## 1 Введение: библиографические ссылки в Интернет-среде.

Во второй половине прошлого века в мир научной коммуникации вошел принципиально новый объект — электронное издание и, в частности, электронный журнал. Новизна не ограничивалась простой трансформацией печатного материала и элементарным приспособлением его к электронной среде, но привела к возникновению совершенно новых свойств данного объекта. Задача о создании точного цифрового подобия, копирующего все возможности традиционного издания, достаточно быстро была решена, и электронный журнал стал наращивать функциональные способности, недоступные печатному аналогу и делающие электронный вариант более ценным.

Двумя первейшими преимуществами электронной среды перед печатной могут быть названы возможность полнотекстового поиска и предоставление

одновременного многопользовательского доступа к одному электронному изданию. В какой-то мере, ценой колоссальных трудовых и финансовых затрат, эти функции все-таки могут быть симитированы в печатном мире: первая — путем составления подробнейших приложений-индексов к изданию, подобных тем, что встречаются в классических академических трудах немецких ученых XIX века; вторая — путем наращивания экземплярности закупаемой библиотекой литературы. Пожалуй, следующим по важности преимуществом электронных изданий перед печатными является возможность мгновенного перехода по ссылкам между статьями или книгами, причем эта функция уже принципиально неповторима в печатном окружении<sup>1</sup>. Изучению именно этого потенциала электронной коммуникативной среды будет посвящена настоящая работа.

Речь идет о настройке и использовании гиперссылок из ци-

тирующей статьи на цитируемую статью. Фактически, такая ссылка (как логический элемент) уже существует в тексте печатной публикации, однако это библиографическая ссылка. Таким образом, наличие Интернет-ссылки на цитируемую статью уже как бы «заложено», «предполагается» самой структурой традиционной научной публикации и создание гипертекстовой связи само собой напрашивается, когда мы переходим из печатной среды в электронную. По этой причине научным и технологическим сообществом прилагаются немалые усилия для развития соответствующего сервиса: преобразование максимального числа ссылок из чисто библиографических в библиографически-гипертекстовые<sup>2</sup> с возможностью беспрепятственного (seamless) и мгновенного перехода пользователя к цитируемой публикации. Это значительный шаг вперед в эволюции технологий современной научной коммуникации.

<sup>1</sup> В дальнейшем мы для определенности будем говорить о периодических изданиях и, соответственно, о статьях в них, т. к. именно для электронных журналов описываемые функции кросс-связей развиты в настоящий момент наиболее полно и используются наиболее широко. Кроме того, «электронный журнал» мы по тем же причинам будем считать синонимом «онлайн-журнала» — издания, размещенного в сети Интернет и доступного по http-протоколу при помощи какого-либо Интернет-браузера.

## 2 Различные типы кросс-ссылок.

Для краткости будем называть описанные нами гипертекстовые ссылки из списка процитированной в электронной статье литературы на собственно цитируемые источники, размещенные в том или ином онлайн-ресурсе, кросс-ссылками<sup>3</sup>. Существуют различные типы кросс-ссылок и несколько способов их классификации.

Например, кросс-ссылки различаются по своему «источнику» — электронному журналу или базе данных, в которых находится цитирующая статья. Это может быть не только полнотекстовая версия издания и конкретной статьи, но и, например, «цитатная» база данных, не включающая в себя полные тексты публикаций, а специализирующаяся на сборе и предоставлении пользователю пристатейных списков цитирования наряду с библиографическими сведениями. В частности, такими являются базы данных Института научной информации (Institute for Scientific Information, ISI), входящие в Web of Science (Science Citation Index, Social Sciences Citation Index, Arts & Humanities Citation Index), а также продукт компании Elsevier, появившийся в 2004 г., Scopus. Аналогичным образом «источником» кросс-ссылки может быть «агрегатор», предоставляющий ряд изданий в полных текстах, а для некоторых журналов также собирающий только библиографию, аннотации статей и списки цитируемой литературы. Пример такого ресурса — базы данных EBSCO, например Academic Search Premier или Business Source Complete. Заметим, что и в том и в другом

случае пользователь ресурса может воспользоваться кросс-ссылками из статьи, даже не имея доступа к ее полному тексту. Этим такие ссылки отличаются от «обычных» — например, размещенных на сайте издательства на странице с полнотекстовой версией статьи.

Далее, кросс-ссылки могут отличаться друг от друга типами своих «целей», т.е. тех страниц, на которые они указывают. Во-первых, ссылка может вести прямо на полный текст в том или ином формате (pdf, html, реже другие). Во-вторых, кросс-ссылка может иметь менее глубокий «уровень разрешения» и вести, например, на оглавление выпуска журнала, в котором опубликована цитируемая статья, на страницу тома журнала, к которому относится данный выпуск, в конце концов просто на главную страницу данного журнала<sup>4</sup>. В этом случае пользователь самостоятельно должен найти цитируемую статью, пользуясь функциями навигации на сайте журнала. Это, конечно, менее удобно для читателя, однако по различным причинам не для всех ресурсов удается настроить ссылки непосредственно на размещенные в них статьи и иногда приходится довольствоваться «более общими» ссылками. Наконец, вообще необязательно «целью» должна служить полнотекстовая версия цитируемой статьи. Это может быть, например, промежуточная страница, где пользователю предлагается выбрать формат, в котором он хочет прочесть публикацию, или указать ресурс, из которого следует открыть статью (при наличии полнотекстовой версии в нескольких различных базах данных). В принципе это даже может быть просто аннотация данной публикации или пристатейный список цитируемой в ци-

тируемой статье литературы (таким образом, например, устроены кросс-ссылки внутри базы данных Scopus). Так или иначе, кросс-ссылка в самом общем случае должна вести на страницу или сайт, несущие некоторую информацию о цитируемой статье, дополнительную к той, что содержится в исходном пристатейном списке (т.е. к библиографической).

Еще одна важная характеристика, которой различаются кросс-ссылки между собой и которая будет интересовать нас в первую очередь, — это взаимное расположение «источника» ссылки и ее «цели». Размещены они в одном ресурсе — или в разных? Иными словами, находятся цитирующая и цитируемая статьи на одной и той же платформе или нет. В первом случае говорят о «внутренней» ссылке, во втором — о «внешней». Подробнее об этой типологии с примерами ее реализации в различных ресурсах см. в [1].

Когда возникла задача создания кросс-ссылок между статьями, размещенными в онлайн, то и логически и хронологически первой была поставлена цель настройки ссылок между источниками, находящимися на одной Интернет-платформе. Это проще как технически, так и методологически: и цитирующая и цитируемая электронные публикации принадлежат одной и той же компании, размещены на едином сервере, имеется полный доступ к их файлам, URL-адресам и пр. Таким образом, поддержка «внутренних» ссылок — стартовый этап в комплексном внедрении гипертекстовых связей между публикациями, и достаточно часто компания, предоставляющая какие-либо сервисы кросс-ссылок, этот этап уже прошла, имея полностью настроенный аппарат внутренних ссылок.

<sup>2</sup>Т.е. записи, которые содержат как библиографическую информацию о цитируемой публикации, так и гиперссылку на нее. Чаще всего при этом тэг Интернет-ссылки `<a href="http://...">` обрамляет либо некоторую часть библиографической записи цитируемой статьи (например, ее заглавие), либо отдельное указание на цитируемый электронный документ, выведенное за рамки библиографии (например, «Full Text», «PDF», «View at Publisher» и др.). Всю конструкцию целиком можно называть библиографической-гипертекстовой ссылкой или, если позволить себе неологизм, библиогиперссылкой.

<sup>3</sup>Терминологическое замечание: введенное таким образом понятие кросс-ссылки же, чем «библиогиперссылка»: речь идет только о цитировании одних источников другими и за рамками остаются, например, размещенный в сети список рекомендуемой литературы по теме, ссылка на электронный документ в электронном библиотечном каталоге и т.д. Все это «библиогиперссылки», которые, однако, не являются кросс-ссылками, если — повторим — оставаться в рамках использованных определений.

<sup>4</sup>Для такого рода ссылок в английской литературе используются обозначения «issue-level links», «volume-level links» и «journal-level links». Соответственно, ссылки непосредственно на статью называют «article-level links».

В дальнейшем развились более сложные отношения: сначала заключались двусторонние соглашения между различными компаниями, а затем решались и задачи стандартизации, ус-танавливавшие единые правила настройки и работы с кросс-ссылками для всех участников электронной научной коммуникации. Примерами таких стандартов могут быть OpenURL [см. работы [2], [3] и собственно стандарт [4]] или проект CrossRef [см. например, [5]], однако далее мы не будем затрагивать эти комплексные инициативы по организации внешних связей, сосредоточившись, наоборот, на внутренних кросс-ссылках, функционирующих в рамках одной платформы.

### 3 Внутренние кросс-ссылки и вопрос о «замкнутости» информационного ресурса.

По мнению автора настоящей статьи одна из причин, по которой внутренние кросс-ссылки ресурса заслуживают особенного внимания, — это их предполагаемая способность быть показателем «замкнутости» или «самодостаточности» электронной базы данных периодики. Чем больше внутренних кросс-ссылок имеется на той или иной платформе, тем в меньшей мере пользователь, читающий документ в данном ресурсе и обращающийся по библиографическим ссылкам к другим публикациям, будет испытывать потребность в иных информационных источниках, помимо данного ресурса. Другими словами, если ученый читает статью из некоторой базы данных периодических изданий, то вероятность того, что цитируемую в статье публикацию он сможет найти в полном тексте в том же ресурсе, находится в прямой зависимости от числа внутренних кросс-ссылок<sup>5</sup>. В частности,

можно представить ситуацию, когда специалист имеет доступ только к тому ресурсу, в котором читает статью. В таком случае вопрос о наличии внутренней кросс-ссылки у позиции в списке цитируемой литературы вообще эквивалентен вопросу о доступности цитируемого труда. Ресурсы, в которых внутренние кросс-ссылки встречаются чаще, обладают, таким образом, большей «информационной замкнутостью» или «самодостаточностью», позволяя пользователю в большей мере исследовать дерево цитирований, не покидая рамок данного ресурса.

Интересно поставить задачу о количественной оценке такого рода «самодостаточности» электронного информационного ресурса. Очевидно, что это должен быть не абсолютный, а относительный показатель, т.е. не суммарное количество внутренних кросс-ссылок, а средний процент библиографических записей в списках литературы, имеющих при себе гиперссылку, которая ведет на полный текст цитируемой публикации, размещенный на той же платформе. Назовем данную величину коэффициентом интрацитирования (передавая этим термином «внутреннее» цитирование статей, представленных в том же ресурсе).

Введенный нами коэффициент можно считать для различных множеств статей. Сначала в качестве иллюстрации рассмотрим самый простой случай — коэффициент интрацитирования для отдельной статьи, который просто равен отношению числа цитирований в ней тех источников, которые размещены на той же платформе, к общему числу цитирований. Например, электронная версия статьи, размещенная на сайте издательства, содержит 20 пунктов в пристатейном списке процитированной литературы, из них 5 снабжены внутренними кросс-ссылками (следовательно,

полные тексты цитируемых работ также представлены на сервере издательства). В этом случае коэффициент интрацитирования для данной статьи на платформе данного издательства составляет 25%. Разумеется, для той же самой работы, размещенной в другом ресурсе (например, в базе данных агрегатора), коэффициент может быть уже иным. Более того, с течением времени данный показатель может меняться, отражая изменение контента соответствующих ресурсов: скажем, если издательство оцифровало дополнительные архивные выпуски своих журналов, в том числе какие-то публикации, процитированные в рассматриваемой статье, то коэффициент увеличится; или напротив, у агрегатора может закончиться договор на предоставление на своей платформе каких-то источников, процитированных в этой статье, что приведет к уменьшению коэффициента интрацитирования, посчитанного для данной статьи и данного агрегатора. Еще один очевидный механизм изменения введенного нами коэффициента со временем — это завершение «эмбарго» (требуемой некоторыми издательствами искусственной задержки между выпуском печатной версии журнала и размещением его полного текста в базе данных) и выход новых источников в агрегаторе в полный текст.

Коэффициент интрацитирования для одной публикации, конечно, мало характеризует весь ресурс в целом. Поэтому логично расширять множество статей и считать показатель для целого номера журнала, для нескольких номеров, для нескольких периодических изданий или даже для целой предметной области (взяв достаточное количество выпусков различных журналов, относящихся к данной тематике)<sup>6</sup>.

<sup>5</sup>Т. к. выше было введено расширительное толкование термина «кросс-ссылка», следует оговориться, что в данном случае мы имеем в виду кросс-ссылки на полнотекстовые версии цитируемых статей — только при доступности полного текста можно говорить о «замкнутости».

<sup>6</sup>Отметим, что, считая коэффициент интрацитирования для более чем одной статьи, было бы ошибочно найти такой коэффициент для каждой статьи, а потом взять их среднее. Это было бы подсчетом «среднего отношений», в то время как необходимо определить «отношение средних»: среднего числа внутренних кросс-ссылок во взятом множестве статей к среднему числу позиций в пристатейных списках литературы. Для этого надо найти суммарное число внутренних кросс-ссылок во всех статьях, принятых к рассмотрению, затем — суммарное число пунктов в их списках цитируемой литературы и, наконец, посчитать отношение первого ко второму.

Можно ли посчитать коэффициент интрацитирования для всего электронного ресурса в целом? Конечно, это слишком трудоемко и едва ли выполнимо даже для небольших баз данных периодически: пришлось бы просмотреть все статьи, представленные на платформе. Такого рода задачи принципиально решаемы только для самого владельца ресурса, который мог бы собирать эти данные (процент пунктов в пристатейных списках, которые снабжены внутренними кросс-ссылками) в автоматическом режиме. С другой стороны, даже если технически мы могли бы считать такой коэффициент, едва ли стоило бы охватывать весь ресурс: скорее всего более информативным в таком случае будет срез по одному-двум годам — например определение коэффициента интрацитирования на множестве всех журналов, вышедших за два последних года.

Таким образом, при практическом подсчете введенный нами показатель может серьезно зависеть от:

- выбора предметной области, по которой он считается (скорее всего база данных будет иметь высокий коэффициент в своих «профильных» дисциплинах, по которым имеется много изданий)

- выбора конкретных наименований журналов в заданной дисциплине

- выбора конкретных выпусков отобранных наименований (например, старые номера будут, вероятно, реже цитировать статьи, имеющиеся в электронном виде, т. к. архив оцифрованной литературы в ресурсе нередко ограничен)

В любом случае коэффициент интрацитирования меняется в пределах от 0 до 100% и чем он больше, тем о большей «полноте», «самодостаточности», «замкнутости» базы данных можно говорить.

## 4 Пробный подсчет коэффициента интрацитирования и анализ результатов.

В работе [6] сделана первая попытка просчитать коэффициент интрацитирования для двух электронных ресурсов на ограниченном множестве экономических журналов. В качестве объектов исследования взяты две платформы, которые являются одними из наиболее прогрессивных в плане развития сервисов внутренних (и не только внутренних) кросс-ссылок. Это EBSCO (взяты конгломерат баз данных Academic Search Premier, Business Source Complete и Masterfile Premier) и журналы издательства Эльзевир на ScienceDirect (рассматривается множество всех журналов на данной платформе). Предметной областью выбрана экономика и решено сделать выборку из нескольких журналов, попавших на верхние позиции рейтингов ISI по рубрике Economics в 2003 г. (рейтинги составляются на основе импакт-факторов, см. [7]).

Из шести журналов, стоящих на первых местах в рейтинге, было взято пять (занимающий 4-е место NBER Macroeconomics Annual не имеет росписи пристатейных списков ни в одном из двух ресурсов). Т. к. особенный интерес представляют издания, одновременно доступные в обоих ресурсах, был добавлен еще один такой журнал, занимающий 12-е место в рейтинге. В полученной шестерке три издания представлены на ScienceDirect (в полном тексте) и все шесть — на платформе EBSCO (в виде аннотаций или в полном тексте, но с обязательной росписью пристатейных списков цитирования). Окончательный лист отобранных журналов: Journal of Economic Literature, Quarterly Journal of Economics, Journal of Accounting and Economics, Journal of Financial Economics, Journal of Economic Perspectives, Journal of Health Economics.

Что касается временного охвата, то для каждого журнала

было взято 4 самых свежих номера, в которых расписаны пристатейные списки.

Интересны результаты, полученные в данном пилотном исследовании: если считать коэффициент интрацитирования только по трем журналам, представленным в обоих ресурсах, то для ScienceDirect его значение будет равно 20%, а для EBSCO — 32%. Из 3411 библиографических описаний, найденных в пристатейных списках статей из 4 номеров каждого журнала, 669 снабжены внутренними кросс-ссылками на платформе ScienceDirect (т. е. цитируется работа, также представленная в полном тексте на ScienceDirect) и 1103 — на платформе EBSCO. Если же для EBSCO посчитать коэффициент по всем шести отобранным журналам, то он окажется несколько меньше — 28% (из 9029 пунктов библиографии 2502 ссылаются на статьи, размещенные в полнотекстовом формате на том же ресурсе).

Полученные результаты весьма любопытны и даже отчасти парадоксальны. Логичнее было бы ожидать, что на сайте издательства Эльзевир большее число ссылок будет вести на журналы того же издательства, следовательно также размещенные на ScienceDirect и потому снабженные внутренней кросс-ссылкой. В частности, цитирование журналом самого себя («самоцитирование») дает вклад в коэффициент интрацитирования ScienceDirect и не дает — в EBSCO (т. к. полные тексты эльзевировских журналов практически не размещаются в иных ресурсах помимо ScienceDirect). Однако оказывается, что агрегатор, широко представляющий журналы различных издательств, содержит больше полнотекстовых статей, на которые ссылаются в трех эльзевировских изданиях по экономике, чем сама платформа Эльзевира. Более того, аналогичный показатель EBSCO для трех журналов других издательств, размещающих свои

<sup>7</sup> В [6] использовался термин «самоцитирование», от которого мы решили отказаться в пользу «интрацитирования». Сделано это во избежание ассоциаций с известным в библиометрии коэффициентом самоцитирования авторов научных публикаций и журналов, описывающим процесс использования ссылок на собственные работы или на работы, опубликованные в том же издании. Там термин «самоцитирование», в противоположность нашему случаю, имеет в какой-то степени негативный подтекст.

полные тексты в этом агрегаторе<sup>8</sup>, оказывается хуже: добавление их в расчет понижает коэффициент интрацитирования с 32% до 28%.

В ходе предварительного исследования [6] было просмотрено более 12 тыс. библиографических описаний на двух платформах и зафиксировано более 3 тыс. внутренних кросс-ссылок. Тем не менее исследование должно быть продолжено: интересно расширить как охват журналов по экономике, так и посчитать аналогичные показатели в других дисциплинарных областях. Наконец, можно увеличить набор исследуемых ресурсов — например, JSTOR завершил тестирование сервиса «Reference Linking», предполагающего расстановку внутренних кросс-ссылок [8], и с 1 марта 2006 г. активировал данную функцию для пользователей; а в недавней e-mail рассылке компания ProQuest, ближайший конкурент EBSCO, также допустила возможность внедрения аналогичного сервиса на своей платформе.

## 5 Заключение.

Возможности веб-среды позволяют по-новому взглянуть на библиографическую ссылку — увидеть в ней новое измерение, возможность превращения ее в гипертекстовую связь и формирования т. н. библиогиперссылки, не только включающей в себя сведения о цитируемом документе, но и содержащей возможность немедленного перехода к нему.

Помимо практических преимуществ, которые предоставляет такая трансформация, она также позволяет исследователю по-новому взглянуть на все пространство библиографического цитирования, задавать новые вопросы и проводить новые эксперименты. В частности, можно сконцентрироваться на внутренних кросс-ссылках, поставить задачу оценки «замкнутости» информационного ресурса и пытаться ввести меру его «самодостаточности», т. е. определить, в какой степени публикации, размещенные в базе данных, цитируют источники, которые также представлены на той же платформе.

Естественным показателем в таком случае является «коэффициент интрацитирования», равный проценту пунктов в пристатейных списках литературы, которые снабжены внутренней кросс-ссылкой. Первые попытки оценить данный коэффициент привели к нетривиальным результатам: не всегда по характеру контента ресурса можно сказать, будет ли интрацитирование активным, а ресурс — «замкнутым». Автор предполагает, что подобные оценки необходимо продолжать и совершенствовать — коэффициент интрацитирования может стать новой содержательной характеристикой информационных ресурсов наряду с такими показателями, как процент ISI-расписываемых изданий, число журналов из десятков наиболее цитируемых в той или иной дисциплине, средний импакт-фактор и т. п.

Автор благодарит М. С. Аксентьеву, П. Г. Арефьеву и Г. В. Калягина за плодотворные дискуссии во время написания настоящей работы.

## ЛИТЕРАТУРА

1. Grogg J. E. Thinking About Reference Linking // *Searcher*. 2002. Vol. 10, Iss. 4. P. 56–61 // <http://www.infotoday.com/searcher/apr02/grogg.htm>.
2. Van de Sompel H., Beit-Arie O. Open Linking in the Scholarly Information Environment Using the OpenURL Framework // *D-lib Magazine*. 2001. Vol. 7, No. 3 // <http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>
3. Ferguson C. L., Grogg J. E. OpenURL Link Resolvers // *Computers in Libraries*. 2004. Vol. 24, Iss. 9. P. 17–24.
4. ANSI/NISO Z39.88-2004. The Open URL Framework for Context-Sensitive Services. Bethesda: NISO Press, 2005 // [http://www.niso.org/standards/standard\\_detail.cfm?std\\_id=783](http://www.niso.org/standards/standard_detail.cfm?std_id=783)
5. Brand A. Publishers Joining Forces through CrossRef // *Serials Review*. 2004. Vol. 30, Iss. 1. P. 3–9.
6. Писляков В. В. Ах, зачем нам кто-то еще? — Индекс самоцитирования или самодостаточности электронных баз данных. Доклад на VI международной конференции «SCIENCE ONLINE: электронные информационные ресурсы для науки и образования». 2005 // <http://elibrary.ru/projects/conference/egypt2005/presentations/Self-citing.ppt>.
7. Garfield E. The Impact Factor // *Current Contents*. 1994. No. 25. P. 3–7 // <http://scientific.thomson.com/tree/essays/journalcitationreports/impactfactor/>
8. Reference Linking. 2006 // [http://www.jstor.org/about/reference\\_linking/reflinking.html](http://www.jstor.org/about/reference_linking/reflinking.html)

<sup>8</sup>А именно *American Economic Association*, выпускающая *Journal of Economic Literature* и *Journal of Economic Perspectives*, и *MIT Press*, публикующий *Quarterly Journal of Economics*.