

**Государственный Университет
Высшая школа экономики**

В.В.Писляков

**АНАЛИЗ КОНТЕНТА ВЕДУЩИХ
ЭЛЕКТРОННЫХ РЕСУРСОВ
АКТУАЛЬНОЙ ЗАРУБЕЖНОЙ
ПЕРИОДИКИ**

Препринт WP2/2002/02

Серия WP2
Количественный анализ в экономике

Москва 2002

Писляков В.В. Анализ контента ведущих электронных ресурсов актуальной зарубежной периодики: Препринт WP2/2002/02. – М.: ГУ ВШЭ, 2002. – 32 с.

Работа посвящена всестороннему анализу контента электронных ресурсов иностранных периодических изданий с онлайн- и ондиск-доступом. Основной акцент сделан на базах данных изданий по общественным и экономическим наукам (экономика, бизнес, социология, политология, право). Осуществлен детальный анализ данных по количественным характеристикам контента ведущих систем. Предложена методика определения профессионального уровня и информативности электронных ресурсов с индексом цитируемости изданий в качестве основного критерия оценки. Приведены результаты применения методики для ряда ресурсов. Намечены перспективы использования разработанного метода для комплектования университетских библиотек, а также перспективы его дальнейшего совершенствования.

Pisliakov V.V. Analysis of the Content of Leading Electronic Databases of Foreign Actual Periodicals: Working paper WP2/2002/02. – Moscow: State University – Higher School of Economics, 2002. – 32 p. (in Russian).

The paper deals with a comprehensive analysis of the content of electronic databases of foreign periodicals with online and CD-ROM access. Main emphasis lies on social sciences databases (economics, business, sociology, political science, law). A detailed analysis of the quantitative characteristics of the content of leading databases is performed. A procedure for determining professional level and quality of the electronic database content with the impact factor as a principal criterion is proposed. Results of this method used for some of the content providers are presented. Perspectives of applying of the method for collection acquisition management in academic libraries as well as its further enhancing are outlined.

Писляков Владимир Владимирович
Государственный университет –
Высшая школа экономики
101990, Москва, Мясницкая, д. 20
E-mail: pislyakov@hse.ru

Pisliakov, Vladimir
State University – Higher School of Economics
20, Myasnitskaya, Moscow 101990, Russia
E-mail: pislyakov@hse.ru

Введение*

Общемировое развитие электронных систем, предназначенных для библиотек, и постепенное проникновение рынка информационных ресурсов такой направленности в Россию приводит к необходимости выбирать и сравнивать, проводить экспертные оценки различных продуктов. Справедливо сказать, что аналитика электронных ресурсов будет занимать все более значительное место в общем комплексе мероприятий по формированию фондов российских библиотек, в исследованиях по библиометрии и инфометрии, а также в программах обучения «библиотечной науке» (library science).

Именно такому анализу электронных библиотечных ресурсов посвящена настоящая работа, причем анализ не будет настолько обширен, чтобы охватить все или даже бóльшую часть существующих систем. На данный момент в мире существуют сотни провайдеров, «агрегаторов», накопителей и распространителей информационных баз данных с онлайн- или ондиск-доступом. Из них мы выделим системы, предоставляющие доступ к периодическим изданиям. Каждая компания на рынке таких систем ставит свои задачи, имеет свою «целевую аудиторию», по-своему строит отношения с издателями. Нас будут интересовать те продукты, которые создаются для библиотек, в первую очередь *академических библиотек* — библиотек вузов.

Ряд баз данных формируется в результате сбора, анализа и индексирования только «первичной» библиографической информации по периодике: оглавлений журналов, иногда — аннотаций статей¹. Другие системы организуют онлайн-доступ к полным текстам изданий, но исключительно к тем, которые *библиотека выписывает в печатном виде*. И вполне логично, что при этом доступ к электронным версиям может предоставляться даже бесплатно. Есть также системы, дающие пользователям возможность получить полные тексты публикаций на основе *pay-per-view*, за отдельную плату для каждой статьи. Все указанные ресурсы останутся за рамками нашего доклада. Мы будем исследовать системы, дающие доступ сразу к множеству (иногда нескольким тысячам) научных зарубежных журналов многих издательств, как правило, сразу по нескольким дисциплинам, причем выписывает библиотека издания в печатном виде или

* Статья является расширенной и переработанной редакцией одноименного доклада, прочитанного автором на конференции «Библиотека-InfoPro. Электронные ресурсы и онлайн-контент вузовских библиотек» (ГУ ВШЭ, 26–27 ноября 2001 г.). Автор выражает благодарность П. Г. Арефьеву, В. А. Бессонову, Н. Ю. Максимовой, А. В. Песоцкому за ценные замечания и плодотворные дискуссии в процессе подготовки настоящей работы.

¹ Среди таких систем, в частности, широко известные базы данных EconLit, Sociological Abstracts, Current Contents и др.

нет, это никак не влияет на ее доступ к электронному журналу. Фиксированная абонентская плата вносится за определенный срок (чаще всего за год), в течение которого университет неограниченно использует ресурс. При этом акцент в рассматриваемых в настоящей работе системах делается именно на свежих журналах, а не на архиве, — потому вне нашего внимания останется столь важный и масштабный проект, как JSTOR.

Описанных электронных библиотечных систем также несколько, и из них мы, вслед за авторами [1] и [2], выбрали 3 ведущие системы, 3 крупнейших «агрегатора», которые, к тому же, постепенно получают распространение в России. Это EBSCOhost (производитель — компания EBSCO Publishing, отделение EBSCO Information Services), ProQuest (производитель — ProQuest Company) и InfoTrac (производитель — Gale Group companies, подразделение The Thomson Corporation):

EBSCOhost	http://www.epnet.com
ProQuest	http://www.proquest.com
InfoTrac	http://www.galegroup.com

Близость «контентной политики», концептуального подхода к информационному наполнению данных ресурсов позволяет нам проводить их сравнение, и именно их контент будет рассматриваться при дальнейшем анализе.

Сразу сделаем оговорку, что, вообще говоря, ценность базы данных периодики — это далеко не только ее контент. Однако в рамках настоящей работы мы не будем говорить о столь немаловажных качествах системы, как ее интерфейс, поисковые возможности, профессиональный уровень аннотирования публикаций, наличие аккуратного и подробного тезауруса. Не будем мы говорить и о том, в каком формате представлены полные тексты статей². Мы также не станем обращать внимания на глубину архива каждого ресурса, на то, сколь *старые* издания можно найти в нем. Наконец, за рамками исследования останется и вопрос о ценах на ресурсы. Мы коснемся оценки качества контента баз данных, но не отношения «цена/качество».

Итак, единственное, о чем будет идти речь, — это *контент*, наполнение, «содержимое» баз данных периодики, причем «актуальный» контент, последние выпуски журналов, а не их архивы.

² Идеальным является представление полных текстов одновременно *всеми* типами файлов, используемыми на текущий момент в электронных библиотечных ресурсах, — html, html со вставленными рисунками и таблицами, pdf. Ибо, несмотря на все безусловные преимущества pdf-формата (электронная копия содержит изображение печатной страницы — в точности так, как выглядит бумажная версия издания, что делает возможным полноценное научное цитирование статей), некоторым пользователям неудобно или накладно каждый раз загружать статью в столь объемном формате. Форматы html значительно более компактны, что в случае низкоскоростного Интернет-соединения или дорогого трафика может оказаться предпочтительным.

Наше исследование включает количественный (раздел I) и качественный (раздел II) анализы контента ресурсов. Перед тем как приступить к содержательной части, отметим, что даже в западной литературе достаточно редко вплотную затрагивается данный вопрос. Причиной этому не только новизна темы, но, вероятно, и необходимость быть чрезвычайно осторожным при рассмотрении такого рода проблем. Тесная связь подобной аналитики с бизнесом, ее влияние на репутацию и положение на рынке тех или иных компаний, производящих электронные библиотечные ресурсы, приводит подчас к излишней тактичности авторов обзоров, стремлению скорее показать, что «все продукты одинаково хороши», чем провести объективный анализ. Быть может, поэтому в настоящее время «откровенно» тема обсуждается чаще на форумах и «листсерверах» (см., например, [3], [4]), чем в периодической печати и Web-журналах. Тем не менее, как недавно отметил сотрудник принстонской библиотеки, именно сейчас по данному вопросу все острее ощущается потребность в «значительно более точном и критическом подходе, чем нам обычно предлагают» [6].

I. Количественный анализ контента электронных баз данных зарубежной периодики

Мы начнем с количественного анализа: сколько изданий представлено в каждом ресурсе, каков объем «пересечения» контентов различных систем³. Хорошо известно, что два разных ресурса часто включают в себя некоторое количество одних и тех же изданий, и перед библиотекой нередко встает вопрос: заплатив деньги за расширение электронной подписки, за приобретение еще одного ресурса, не приобретет ли она те же самые журналы — или в большинстве своем те же самые?

Данные для количественного анализа берутся нами из опубликованного летом 2001 г. обзора Ларри Круменэйкера [1], в котором приводились некоторые количественные характеристики контента систем EBSCOhost, ProQuest, InfoTracWeb. Важно отметить, что Круменэйкер брал для своего анализа *только полнотекстовые* журналы, представленные в ресурсах, причем им были исключены из рассмотрения разного рода информационные листки, сообщения информационных агентств и прочие издания, которые нельзя в строгом смысле назвать «журналами». Кроме того, упомянем здесь, что каждая из систем (ProQuest, EBSCO, InfoTrac) может предоставлять доступ ко многим (нескольким десяткам) базам данных периодики различной тематики или к политематическим базам данных. Такие базы данных, являющиеся частью всего информационного массива, к которому каждый агрегатор может

³ Здесь и далее, когда мы говорим «количество изданий», «количество журналов», «объем базы данных» речь всегда идет о числе *наименований* периодических изданий, каждое из которых, в свою очередь, представлено в ресурсах многими выпусками, номерами. Вопрос о *количестве номеров* того или иного журнала очевидно сводится к вопросу о глубине архива — которым, как отмечалось, мы заниматься не будем.

обеспечить доступ, мы называем далее «частными базами данных». Ряд из них — реферативные ресурсы, содержащие только аннотации статей, и потому они не добавляют никаких полнотекстовых документов. Другие являются подмножеством более крупных частных баз данных, доступных через того же провайдера⁴, а значит, также не влияют на проводимое здесь исследование. Так или иначе, основной вклад в полнотекстовую коллекцию каждой из трех «систем» вносят 2–5 частных баз данных (а в случае InfoTrac — даже одна база). Круменэйкер рассматривал *всю совокупность* частных баз, принимая во внимание все частные ресурсы, имеющиеся у каждого агрегатора. Таким образом, *вся* полнотекстовая периодика, к которой можно получить доступ через ProQuest, EBSCOhost, InfoTrac, учитывается в работе американского специалиста. Это необходимо иметь в виду при анализе выводов всей первой части настоящей статьи и при переходе ко второй части, где для исследования будет взят некий представительный «комплект» частных баз данных каждого провайдера.

Итак, из приведенных в работе Круменэйкера данных, проанализировав и дополнив их⁵, можно получить следующую таблицу количественных характеристик контента каждой системы и их «пересечений» (таблица 1).

Таблица 1. Полнотекстовые журналы в зарубежных электронных системах баз данных: количество «совпадений»

	Всего FT	EBSCOhost	ProQuest	InfoTrac	Везде
EBSCOhost	4039	2170	1466	1230	827
ProQuest	3602	1466	1742	1221	827
InfoTrac	2784	1230	1221	1160	827

В таблице показано количество полнотекстовых изданий, представленных *одновременно* в двух ресурсах. Серым цветом выделено количество «уникальных» журналов, присутствующих *только в одной* из трех систем. Также приведены суммарное количество полнотекстовых

⁴ Таковы, например, *Academic Search Elite* и *Business Source Elite*, являющиеся подмножеством соответственно *Academic Search Premier* и *Business Source Premier*, также доступных через EBSCOhost. Ср. [5].

⁵ Круменэйкер приводит только некоторые из цифр, а именно: суммарное количество полнотекстовых изданий в каждой из систем, количество журналов, представленных одновременно во всех ресурсах, и число «уникальных» изданий, присутствующих только в данной системе (для всех трех ресурсов) [1, р. 41–42]. Решением системы линейных уравнений легко восстанавливаются недостающие показатели. Автор статьи сообщает также, что количество журналов, представленных только в двух из трех ресурсов, варьируется от 391 до 632, не указывая, для каких именно пар систем достигаются минимум и максимум. По таблице 1 можно проверить эти «избыточные» данные и обнаружить, что они несколько отличаются от приведенных Круменэйкером ($1466-827=639$ для ProQuest/EBSCO и $1221-827=394$ для ProQuest/InfoTrac). Нам неизвестно, чем обусловлена неточность в обзоре американского специалиста, однако в процентном отношении она столь мала, что на ней не следует заострять внимание. Так или иначе, сомневаться (как это делает автор [7], [8]) в данных, представленных Круменэйкером, постоянным редактором авторитетного справочника *Net.Journal Directory*, на наш взгляд, нет серьезных оснований. Отличие численных показателей, приведенных в [2] или [5], обусловлено тем, что в них исследователи брали не полный «информационный массив» систем, а некоторый набор частных баз данных.

периодических изданий в каждом ресурсе («Всего FT») и число журналов, представленных во всех трех базах данных («Везде»). Например, EBSCOhost в общей сложности содержит 4039 полнотекстовых изданий, из которых 1466 есть также в ProQuest, 1230 — в InfoTrac, 827 — одновременно в ProQuest и InfoTrac, а 2170 журналов не содержится ни в одной другой из исследуемых систем. Для того чтобы оценить долю «уникального» и «неуникального» контента в системах, выразим все цифры в процентах от общего количества полнотекстовых периодических изданий, доступных в системе.

Таблица 2. Количество «совпадений» в процентном отношении к «Всего FT», %

	Всего FT	EBSCOhost	ProQuest	InfoTrac	Везде
EBSCOhost	100	54	36	30	20
ProQuest	100	41	48	34	23
InfoTrac	100	44	44	42	30

Принципиальный факт, который можно обнаружить в результате анализа таблиц 1 и 2, это достаточно небольшое пересечение контента различных ресурсов — меньше, чем можно было бы ожидать. Этот факт отмечает и сам Круменэйкер: «Едва ли установилась монополия», — пишет он, комментируя тот факт, что более 40% контента *каждой* системы является «уникальным» для нее и отсутствует в других службах [1, р. 42]. К аналогичному выводу приходит и Хилл, исследуя некоторые *частные* базы данных каждой из систем: «К нашему удивлению, нет значительного пересечения полных текстов, включенных в ресурсы». Далее он повторяет свое заключение для специфически научных (peer-reviewed) журналов [2, chapter «Title Comparison»].

Можно взглянуть на полученные данные с другой стороны, со стороны «журнальной массы» и ее распределения по системам баз данных. Из цифр Круменэйкера несложно получить, что всего изданий, представленных в полнотекстовом виде хотя бы в одном ресурсе, 7335. Из них только 827, или 11%, имеются во всех службах. В двух и только в двух системах представлены 1436 журналов, т. е. 20%. Наконец, основная масса, 5072 издания (69%), включена только в одну систему — или EBSCOhost, или ProQuest, или InfoTrac⁶. Представляя эти параметры на уровне конкретных ресурсов, получим таблицу 3, отражающую долю в общей массе тех журналов, что представлены в каждом ресурсе уникально, и тех, что имеются одновременно в двух или одновременно в трех системах (напомним, что здесь цифра, например, 17% пересечения EBSCOhost и InfoTrac включает в себя и те 11% всех журналов, которые включены сразу во все три ресурса).

⁶ Показательно, что очень близкие цифры (соответственно, 10, 20 и 70% для всех изданий и 8, 21 и 71% для peer-reviewed журналов) получает Хилл, берущий для анализа только по одной достаточно представительной частной базе данных из каждой из трех систем [2, loc. cit.].

Таблица 3. Количество «совпадений» в процентном отношении к общему числу журналов (7335), представленных в полнотекстовом формате хотя бы в одной системе, %

	Всего FT	<i>EBSCOhost</i>	ProQuest	InfoTrac	Везде
<i>EBSCOhost</i>	55	30	20	17	11
ProQuest	49	20	24	17	11
InfoTrac	38	17	17	16	11

Наконец, представим всю имеющуюся у нас информацию о пересечении контента различных систем в удобной графической форме, изобразив множества изданий, включенных в ту или иную систему баз данных, в виде трех пересекающихся окружностей (рис. 1).

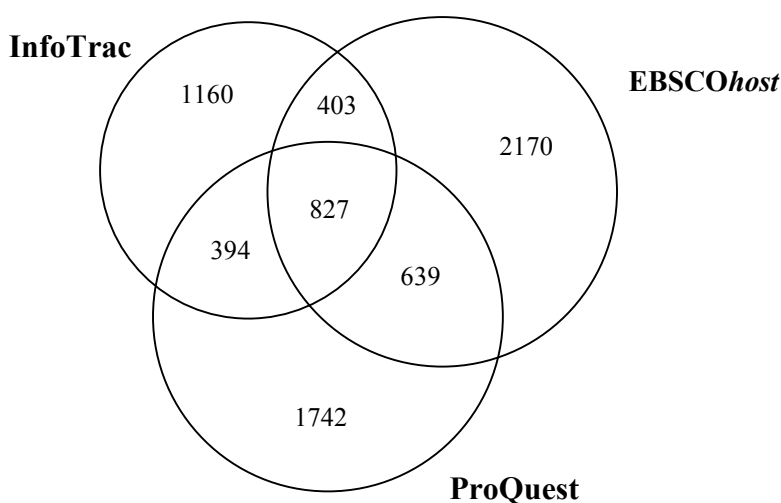


Рис. 1

В завершение первой части отобразим имеющиеся количественные характеристики контента еще одним образом: посчитаем число «несовпадений» или «дополнений» — изданий, отсутствующих в одном ресурсе, но представленных в другом (таблица 4).

Таблица 4. Полнотекстовые журналы в зарубежных электронных системах баз данных: количество «несовпадений»

	Всего нет FT	<i>EBSCOhost</i>	ProQuest	InfoTrac
<i>EBSCOhost</i>	3296		2136	1554
ProQuest	3733	2573		1563
InfoTrac	4551	2809	2381	

Каждое число в таблице 4 обозначает количество полнотекстовых журналов, не представленных в том ресурсе, в строке которого оно находится, но имеющих в системе, указанной в заголовке соответствующего столбца. Это число «дополнений» (новых периодических изданий), которые принесет ресурс-столбец в ресурс-строку при их

объединении. Включен также показатель «Всего нет FT», отражающий количество полнотекстовых журналов, которых нет в данном ресурсе из числа имеющихся во всем «информационном массиве» всех трех систем.

Именно такое представление удобно и полезно при планировании и исследовании различных «сценариев подписки» («subscription scenarios» — удачное выражение, заимствованное нами из [5]). Исходя из этих данных можно строить «количественную стратегию» *при расширении электронной подписки*. Таблица 4 наглядно показывает, чего лишена в настоящий момент организация, оформившая подписку только на один ресурс, и что она приобретет при получении дополнительного доступа к еще одной системе — или к двум сразу. Выразим то же самое в процентном отношении к количеству журналов в каждом ресурсе и получим таблицу 5.

Таблица 5. Количество «несовпадений» в процентном отношении к «Всего FT», %

	Всего нет FT	EBSCOhost	ProQuest	InfoTrac
EBSCOhost	82		53	38
ProQuest	104	71		43
InfoTrac	163	101	86	

Таким образом, библиотека, уже подписанная, например, на EBSCO, оформив подписку на ProQuest, приобретет доступ к 2136 *новым* изданиям, а при приобретении InfoTrac — к 1554. При этом объем электронной подписки увеличится соответственно на 53% и 38%. Оформление подписки сразу на два ресурса (с сохранением, разумеется, доступа к EBSCO), даст 3296 новых полнотекстовых изданий, расширив электронный фонд периодики на 82%.

II. Качественный анализ контента электронных баз данных зарубежной периодики

Теперь мы займемся анализом *качества* контента онлайн-баз данных. Иными словами, следует оценить, насколько признанные, профессиональные и авторитетные издания включены в тот или иной ресурс. Насколько «информативны» и «содержательны» различные системы баз данных. Ведь может оказаться, что хорошие количественные характеристики получены за счет малоизвестных и узкоспециальных журналов — или, наоборот, масс-медийных, неисследовательских, нередатируемых публикаций. Поэтому прежде всего нам надо выбрать критерий оценки, критерий сравнения журналов. Нам необходимо знать, какой из журналов *лучше* и *насколько* он лучше. Ясно, что это крайне сложная и ответственная задача, потому особое внимание будет уделено методике наших оценок.

1. Индексы цитируемости ISI

Мы будем отталкиваться от того, что ценность академического журнала — это в первую очередь его ценность для исследователя, читаемость и объем использования его материалов в научном сообществе. Такую величину можно оценить при помощи индекса цитируемости журнала. Как правило, чем чаще журнал используют в научно-исследовательском процессе, тем чаще ссылаются на него, чаще цитируют и, следовательно, тем выше его индекс цитируемости.

Одной из наиболее авторитетных организаций, проводящей комплексный подсчет индексов цитируемости, является базирующийся в Филадельфии Институт научной информации — Institute for Scientific Information, ISI (<http://www.isinet.com>). Остановимся на методике вычисления индекса цитируемости, изложенной основателем ISI Юджином Гарфильдом в [9], [10]⁷. ISI определил некое множество журналов, в которое входит большинство авторитетных научных изданий на разных языках. Список постоянно обновляется и пополняется, на данный момент по социальным наукам включены 1764 журнала, список их может быть найден в [11]. Индекс цитируемости (импакт-фактор) отражает количество цитат, ссылок на «среднюю статью» в некотором периодическом издании и считается на определенный год. Если это год N, то импакт-фактор равен числу всех цитирований статей данного журнала, вышедших в годах N–1 и N–2, в статьях всех журналов из множества ISI, вышедших в году N, деленному на общее количество статей, опубликованных в данном журнале за годы N и N–1. Например, чтобы вычислить индекс цитируемости журнала J на 2000 г., необходимо проанализировать все статьи, опубликованные в 2000 г. во всех периодических изданиях «множества ISI» (включая сам журнал J), сосчитать количество ссылок в них на статьи, опубликованные в J в 1998 и 1999 г., и разделить это число на общее количество работ, вышедших в J в 1998–1999 гг. Кроме того, ISI считает импакт-факторы за более длительные промежутки времени, за 4 года и даже за 19 лет (см., например, [12]), однако основным показателем полагает индекс цитируемости именно с двухгодичной «глубиной».

Импакт-факторы для социальных и экономических наук считаются ISI на основе своей базы данных «цитат-индексов» *Social Sciences Citation Index (SSCI)* и помещаются в другой продукт ISI — *Journal Citation Reports (JCR)*⁸. На данный момент, насколько нам известно, в России нет ни одной установки весьма дорогостоящего ресурса SSCI, остался недоступен нам в полном объеме и JCR, поставляемый, как правило, в качестве приложения к SSCI. Тем не менее, в «открытом Интернете», на сайте, поддерживаемом ISI, регулярно помещаются списки десяти наиболее цитируемых журналов по различным дисциплинам с указанием их импакт-факторов.

⁷ Здесь, равно как и во всех изданиях ISI, индекс цитируемости называется импакт-фактором (impact factor), причем английское «impact» в данном случае обозначает именно «воздействие», влияние публикации или издания на академическое сообщество.

⁸ См.: описания SSCI, <http://www.isinet.com/isi/products/citation/ssci/index.html>, и JCR, <http://www.isinet.com/isi/products/citation/jcr/index.html>.

Именно на эти данные опираются наши дальнейшие выкладки. Очередная версия JCR появляется ежегодно летом и содержит импакт-факторы за предыдущий год. К моменту написания настоящей статьи в общедоступных источниках были опубликованы индексы цитируемости 2000 г. десятков изданий по экономике, бизнесу, социологии, политологии⁹. Индексы журналов по праву, используемые в нашем исследовании, были взяты по состоянию на 1998 г.¹⁰

Наконец, необходимо сделать оговорку о том, что сравнение импакт-факторов различных журналов имеет смысл только внутри какой-либо одной дисциплины. Издания разных направлений имеют различные средние и максимальные индексы цитируемости, что отражает отличающиеся друг от друга степени заимствования, темпы развития и темпы восприятия новых наработок, уровни разработки отдельных проблемных областей и вообще разные культуры цитирования, сложившиеся в конкретных дисциплинах. Поэтому ISI приводит сравнительные показатели изданий в той или иной научной области, и наш анализ также будет проводиться внутри конкретных дисциплин, без междисциплинарного сравнения «абсолютных» импакт-факторов.

2. Использование индексов цитируемости. Индекс цитируемости и «престиж» журнала

Введенный таким образом индекс цитируемости (импакт-фактор) можно считать показателем, достаточно корректно отражающим уровень использования того или иного издания в научном процессе, в процессе творческого изучения наработанного материала и, как результат, создания нового. Авторы методики рекомендуют, помимо прочего, использовать их показатель библиотекарям — именно для планирования состава подписки на периодические издания. «В первую очередь, — пишет Гарфильд, — мы даем библиотекарям и исследователям средство управления журнальными фондами библиотек» [9]. Использование объективной характеристики, индекса цитируемости, более продуктивно, чем попытки опираться на «престиж журнала», это «абстрактное психологическое понятие», по выражению авторов [13].

Сравнению «престижа», оцениваемого на основании опроса квалифицированных респондентов, с результатами анализа корректных библиометрических показателей, каковым является индекс цитируемости, было посвящено немало работ, причем многие из них касались положения дел в периодике по общественным наукам: [14], [15], [16], [17], [18]. Исследователи отмечали, что «отдельные журналы занимают относительно высокое положение в сознании людей, однако в действительности их используют весьма слабо» [16]. Встречаются соответственно и обратные ситуации, когда «многие новые и менее престижные издания пользуются меньшим доверием, чем должен бы обеспечивать их индекс цитируемости» [15]. В

⁹ См.: [12], [21], [22], [26].

¹⁰ Они также публиковались на сайте ISI, однако в связи с его произошедшей реконструкцией в настоящий момент уже недоступны. Мы нашли их среди сохраненных при прежних просмотрах страниц в кэше одной из поисковых машин Интернета (Google).

результате делается вывод: такого рода исследования являются «доказательствами того, что измерения “репутации” журналов скорее отражают устойчивые стереотипы, чем дают свод информации о действительном влиянии изданий» [15].

Уилл Х. Мур проводит различие «прямых» и «косвенных» показателей качества журнала. Измерения престижа при помощи опросов он причисляет к косвенным характеристикам, «так как, не определяя напрямую само значение, которое ученые придают изданию при проведении собственных исследований, опрос выявляет их *мнения* о придаваемом ими значении различным журналам». Таким образом, «имеются два типа показателей, полезных для оценки качества журналов: связанные с опросом репутации (косвенные) и подсчетом цитат (прямые)». Далее автор приходит к выводу, что «прямые цитатные измерения лучше оценивают качество издания, чем косвенные исследования репутации» [17, р. 3–5, курсив наш]. Помимо введения понятия прямых и косвенных характеристик, Мур подробно останавливается на «разрешающей способности» различных методов. В случае замеров «престижа» изданий¹¹ оказывается, например, что разница между показателями журналов, занимающих двадцатые места в итоговом рейтинге, и журналов, попавших на пятидесятые места, составляет всего 1 балл (при использовании десятибалльной шкалы). Далее исследователь оценивает погрешность измерений «престижа», получает значение около 0,7 баллов и констатирует невозможность достоверно разделить на шкале «репутаций» издание, занимающее 20-е место, и журнал, находящийся на 55-й строчке рейтинга престижа. Так мы приходим к выводу, что исследование репутации изданий делает возможным «некоторое различие между журналами, достаточно далеко отстоящими друг от друга, однако в общем и целом опрос престижа журналов показывает, что нет особенной разницы, где печатать свою работу». С другой стороны, «можно заметить, что разброс результатов цитатных измерений значительно выше». Это вновь приводит исследователя к заключению о превосходстве анализа цитат над обработкой опросов репутации журналов [17, р. 9–11]¹².

Так или иначе, по мнению многих ученых, «проведенные обширные исследования доказывают приемлемость количества цитирований как характеристики относительного уровня публикации» [19] и даже «число цитирований работы ученого часто может быть рекомендовано в качестве *наилучшего* из показателей признания в научном сообществе» [20, курсив наш]. Следовательно, мы имеем все основания взять импакт-факторы ISI как критерий при анализе качества контента электронных баз данных периодики.

¹¹ Мур использует исследования «репутации» политологических журналов, проведенные в 1988 г. и изложенные в [28]. Опрос проводился для 78 журналов при участии 215 респондентов.

¹² Сам Мур, правда, отстаивает большую ценность 10-летнего «кумулятивного» индекса цитируемости, для нахождения которого считается количество цитирований всех выпусков журнала за некоторый год в течение последующих десяти лет. Ученый из Флориды полагает, что такой индекс имеет приоритет по крайней мере в исследуемой им «менее динамичной» дисциплине, политологии. См. также [31], [32].

3. Виды представления журналов. Коэффициенты представленности

Для комплексной оценки контента ресурса необходимо не только учитывать сам факт наличия журнала в базе данных, но принимать во внимание и то, *в каком виде* представлено издание. Даже самый авторитетный журнал с высоким индексом цитируемости, будучи представленным только аннотациями статей или же полным текстом, поступающим в базу с запаздыванием в несколько лет, мало ценен для исследователя, следовательно, мало влияет на качественный уровень электронного ресурса. С другой стороны, учитывать *исключительно* полнотекстовые журналы, поступающие в систему без задержки, на наш взгляд, также некорректно¹³.

Для подсчета интегрального показателя представленности той или иной дисциплины в электронных ресурсах укажем все возможные виды представлений изданий по данной дисциплине и оценим их «веса», чтобы таким образом ввести понятие «коэффициента представленности» того или иного журнала. Так, если журнал имеется в базе данных в полнотекстовом формате и без задержки, то его «коэффициент представленности» логично принять равным единице. Далее мы вправе создать некоторую математическую модель, оценивающую «неполноту» представления того или иного журнала, и присвоить значения остальным коэффициентам представленности.

Таблица 6. Виды представления журналов в онлайн-базах данных и соответствующие коэффициенты представленности

Вид представления	Обозначение	Коэффициент представленности
Полный текст (full text) без задержки	FT	1
Полный текст с задержкой («эмбарго») 6 месяцев	6	0,75 (при эмбарго от 6 до 11 месяцев)
Полный текст с эмбарго 12 месяцев	12	0,5 (при эмбарго от 12 до 23 месяцев)
Полный текст с эмбарго 24 месяца	24	0,3 (при эмбарго от 24 до 35 месяцев)
Полный текст с эмбарго 36 месяцев	36	0,2 (при эмбарго от 36 месяцев и более)
Журнал был недавно исключен из ресурса, последний полный текст — 12-месячной давности	12 x	0,5 (приравнивается коэффициенту при соответствующем эмбарго)
Только аннотации статей (abstracts), полного текста нет	Ab	0,1
Журнал никак не представлен или представлен только оглавлениями номеров		0

¹³ Количественные исследования [1], [5] касаются только полнотекстовых изданий, причем задержка в поступлении электронной версии по сравнению с выходом печатной никак не учитывается (в [1] делается лишь оговорка о суммарном количестве «запаздывающих» журналов). В [2] анализируется также общее число «индексируемых» изданий (т. е. представленных как минимум аннотациями статей) в каждом ресурсе, однако такой анализ проводится отдельно от исследований полнотекстового контента.

В таблице 6 приведены возможные виды представления периодических изданий в электронных базах данных. Ряд издателей позволяют производителям электронных библиотечных ресурсов помещать полный текст своих журналов только по истечении некоторого времени после выхода печатной версии — это так называемое «эмбарго». Чаще всего встречаются 6-, 12-, 24- и 36-месячные эмбарго. Например, 24-месячное эмбарго означает, что в феврале 2002 г. в ресурсе можно найти журнал за февраль 2000 г. (или за I квартал 2000 г. для ежеквартальников) или более старые выпуски. Чем больше эмбарго журнала в той или иной базе данных, тем меньшую ценность представляет эта электронная версия издания для специалиста и, следовательно, тем меньшим должен быть введенный нами коэффициент представленности указанного журнала в данном ресурсе. Например, мы оценили, что если полный текст журнала имеется в ресурсе, но поступает в него с задержкой от одного до двух лет, то соответствующий коэффициент равняется 0,5. Грубо говоря, в нашей модели электронная версия журнала, полный текст которого представлен с 12-месячным эмбарго, в два раза менее ценна для исследователя, чем тот же журнал при условии наличия в базе самых свежих его выпусков.

Встречается также ситуация, когда некоторое время назад полный текст издания прекратил поступать в ресурс, однако имеющиеся прежние полнотекстовые выпуски сохранены. В этом случае мы приравниваем коэффициент представленности коэффициенту для журнала с соответствующим эмбарго. Несмотря на то, что таким образом мы весьма полноценно учитываем «исключенное» издание, необходимо помнить, что с течением времени его коэффициент представленности будет постоянно уменьшаться (ввиду «устаревания» последнего доступного полнотекстового выпуска).

Большое количество журналов (как правило, 30–50% от всего объема базы), индексируемых в ресурсе, вообще не имеют в нем полнотекстового представления. Для таких изданий имеются только аннотации статей, которые фактически всегда поступают без существенной задержки. Коэффициент представленности подобных журналов примем равным 0,1. В исследуемых нами системах практически не встречаются издания, представленные только оглавлениями, даже без аннотаций статей. Если таковые все же есть, мы все равно никак не будем их учитывать. Оглавления несут незначительную информацию, если сравнивать ее с полнотекстовым представлением (для которого коэффициент равен 1) и, кроме того, содержание почти каждого журнала можно найти в «открытом Интернете», на сайтах соответствующих издательств.

Построенная нами система коэффициентов представленности (таблица 6) является, конечно, нашим субъективным видением «степени» (или, наоборот, «неполноты») представленности журналов, находящихся в электронных базах данных в том или ином виде. Наши принципы можно оспаривать и по-иному оценивать различные виды представления. Ввиду отсутствия, насколько нам известно, аналогичных работ с комплексным подходом к анализу

разнородного контента библиотечных ресурсов, мы вынуждены строить математическую модель сами, не опираясь на прежний опыт и исследовательскую традицию. Тем не менее предложенные оценки, как нам кажется, вполне разумны, логичны и внутренне непротиворечивы.

4. Базы данных, выбранные для анализа

Теперь сделаем оговорку о том, какие «частные» базы данных мы возьмем для исследования. Как уже говорилось, каждая система (EBSCOhost, ProQuest, InfoTrac) может дать доступ к нескольким, порой около 20–30, базам различной тематики — при условии оформления подписки на них. В связи с тем, что подписываться на *все* базы данных нецелесообразно, да и едва ли какая-либо академическая библиотека может себе это позволить, нам необходимо определить, какие «частные» базы данных выбрать для качественного анализа их контента. В дальнейшем мы будем оценивать контент всей системы, учитывая только издания, представленные *хотя бы в одной* из выбранных нами частных баз данных. Исходя из того, что наш анализ коснется социальных и экономических дисциплин, можно составить набор частных баз, приведенный в таблице 7.

Таблица 7. «Частные» базы данных, выбранные для анализа

EBSCOhost	Academic Search Premier Business Source Premier MasterFILE Premier
ProQuest	ABI/INFORM Global Academic Research Library ProQuest Education Complete Social Sciences PlusText
InfoTrac	InfoTrac OneFile

Помимо того, что перечисленные базы данных составляют основной общественно-научный контент исследуемых ресурсов, именно эти продукты получают распространение в российских университетских и публичных библиотеках. Базы данных серии «Premier» заявляются компанией EBSCO Publishing как наиболее представительные собрания периодических изданий¹⁴, и именно к перечисленным базам дается доступ в рамках широко известной программы Electronic Information for Libraries (EIFL *Direct*), число российских библиотек — участников которой составляет несколько сот. Базы данных, выбранные нами для исследования контента ProQuest, входят в подписку консорциума четырех московских и одного новосибирского университетов¹⁵. Наконец, собранные воедино практически все ресурсы InfoTrac

¹⁴ См., например, высказанное вице-президентом EBSCO Publishing пожелание использовать при исследованиях по сравнению различных баз данных именно Premier-продукты EBSCO [23].

¹⁵ Участники консорциума в 2001/2002 г.: ГУ ВШЭ, МВШСЭН, РГГУ, РАГС, НГУ. Консорциум открыт для новых членов и планирует в 2002 г. начать активное сотрудничество с «Научной электронной библиотекой» (elibrary.ru).

образовали базу данных «InfoTrac OneFile», которая, по словам вице-президента Gale Group Джона Барнса, является «идеальным продуктом для больших библиотек и консорциумов». Доступность, наличие приемлемых и уже опробованных способов подписки на исследуемые нами базы данных переводит наш анализ в практическую плоскость и делает его реальным руководством по планированию электронной коллекции¹⁶.

Итак, выбор баз данных сделан таким образом, что можно утверждать: если журнала по экономической или социальной тематике нет в указанных продуктах, то с очень высокой вероятностью его нет и во *всех* «частных» базах данных, предоставляемых данной системой. То есть если издание по экономике отсутствует, например, в «Academic Search Premier», «Business Source Premier» и «MasterFILE Premier», то его нет ни в одной базе данных, к которым можно получить доступ через EBSCOhost.

Для всех перечисленных в таблице 7 баз данных имеются составленные производителем «списки журналов» (title lists), в которых отражены временные промежутки представления изданий (дата первого включенного номера, дата последнего присутствующего в базе номера), виды их представления (FT или Ab), информация о существующем эмбарго. Списки доступны в сети Интернет:

EBSCOhost: <http://www.epnet.com/maglists/maglist.htm>

ProQuest: <http://www.proquest.com/division/tl-menu.shtml>

InfoTrac: <http://www.galegroup.com/tlist/sb5101.html>

В списки журналов вносятся изменения в зависимости от новых соглашений, заключаемых производителем ресурса с издателями, или прекращения действия прежних соглашений. Используемые нами данные соответствуют состоянию контента баз данных на ноябрь 2001 г. К базам данных EBSCO и ProQuest имеет доступ ГУ ВШЭ, потому приводимая в дальнейшем информация о наличии в этих системах тех или иных журналов и о виде их представления проверялась нами непосредственно (вплоть до открытия полнотекстовых версий статей в самых свежих из имеющихся в базах номерах издания). Для InfoTrac OneFile использовался только «список журналов» (взятый по приведенному выше Интернет-адресу), без дальнейшей проверки содержащихся в нем сведений¹⁷.

¹⁶ Говоря об «опробованных» и «приемлемых» вариантах, мы имеем в виду в первую очередь консорциумную подписку. Информационный продукт может оказаться недоступным *ни для одной российской библиотеки*, если в библиотечном сообществе нет консорциума по доступу к этой базе данных, внутри которого можно получить значительные скидки и приобрести продукт в полном объеме по меньшей цене. Без организации такой структуры путь ресурса на российский рынок может оказаться полностью закрытым, а интерес к его контенту становится попросту гипотетическим.

¹⁷ Вообще говоря, такая проверка оправдана: нам встречались случаи, когда издание, полный текст которого, согласно списку журналов, должен был присутствовать в ресурсе, на практике оказывалось недоступным или имело существенное эмбарго. Кроме того, однажды был найден обратный пример (!). Чем обусловлено такое несовпадение, как быстро оно обнаруживается и исправляется работниками компаний, нам неизвестно.

5. Качественный анализ электронных ресурсов по дисциплинам: экономика

Теперь мы можем провести процедуру оценки уровня представленности, «информационной наполненности» различных ресурсов по предлагаемой нами методике. Как уже говорилось, сравнение корректно проводить по конкретным дисциплинам. Мы начнем с экономики, а в дальнейшем аналогичные действия повторим для четырех других дисциплин.

Итак, возьмем из [12] десятку ведущих экономических журналов с их импакт-факторами (индексами цитируемости). Изучим представление этих изданий в каждой из трех систем (EBSCOhost, ProQuest, InfoTrac), проверив их наличие в базах данных, перечисленных в таблице 7, и составим таблицу 8, используя обозначения из таблицы 6.

Таблица 8. Десятка ведущих журналов по экономике. Представление в электронных базах данных

Место	Индекс цитирования 2000	EBSCO	ProQuest	InfoTrac
1	Journal of Economic Literature (6,68)	36	36	Ab
2	Quarterly Journal of Economics (4,02)	12	FT	9 x
3	Brookings Papers on Economic Activity (3,93)	FT	FT	FT
4	Journal of Economic Perspectives (2,34)	36	36	Ab
5	Journal of Political Economy (2,32)	FT	FT	FT
6	Economic Policy (2,21)	12	Ab	
7	NBER Macroeconomics Annual (2,00)	FT		
8	Health Economics (1,97)			
9	Journal of Health economics (1,92)	Ab	Ab	Ab
10	Economic Geography (1,91)	Ab	FT	FT

Кратко анализируя таблицу 8, отметим, что у двух журналов из десяти полных текстов нет ни в одной системе. Один журнал вообще никак не отражен в ресурсах¹⁸. У пяти изданий полные

¹⁸ Это «Health Economics». Его присутствие в десятке, наряду с «Journal of Health Economics», может озадачить исследователя и заставить предположить, что эти издания получили высокий импакт-фактор цитатами из медицинских журналов и не имеют соответствующего уровня собственно в экономике. Наличие подобного влияния естественных наук на индексы цитируемости в общественно-научной дисциплине предполагает И. В. Маршакова-Шайкевич, отмечая, что в случае психиатрии высокие показатели цитирования «объясняются тем, что психиатрия тесно связана с клинической медициной, отнесенной к естествознанию, имеющей высокие показатели цитирования» [27, с. 142]. Ввиду отсутствия у нас доступа к полной базе JCR, мы не можем решительно сказать, что высокий рейтинг данных журналов обусловлен повышенным интересом именно к экономике здравоохранения, а не к самой медицине.

тексты без эмбарго имеются хотя бы в одной из исследуемых нами систем. Только в паре ProQuest — InfoTrac первый по всем изданиям не хуже второго. В остальных парах ряд журналов представлен лучше в одном из ресурсов, но при этом как минимум одно издание, наоборот, в более полном виде присутствует в другой системе.

Чтобы провести комплексную оценку представленности авторитетных журналов по экономике в электронных библиотечных ресурсах, необходимо одновременно учесть индекс цитируемости каждого включенного в систему издания и вид его представления, «степень» его представленности. Для этого подставим в таблицу 8 введенные нами «коэффициенты представленности» из таблицы 6. Далее умножим индекс цитируемости журнала на его коэффициент представленности в той или иной системе. Сложив такие произведения по всей десятке, получим величину, характеризующую степень представленности ведущих информационных источников по экономике в каждом ресурсе. Далее разделим этот показатель на аналогичный индикатор, посчитанный для гипотетического «идеального» ресурса, в котором *каждый* журнал из десятки представлен в полнотекстовом виде без эмбарго (эта величина, очевидно, будет равна простой сумме всех импакт-факторов в десятке). Назовем это отношение, выраженное в процентах, *коэффициентом информационной наполненности (КИН)* по экономике. Результаты отражены в таблице 9.

Таблица 9. Десятка ведущих журналов по экономике. КИН

Место	Индекс цитирования 2000	EBSCO	ProQuest	InfoTrac
1	Journal of Economic Literature (6,68)	0,2	0,2	0,1
2	Quarterly Journal of Economics (4,02)	0,5	1	0,75
3	Brookings Papers on Economic Activity (3,93)	1	1	1
4	Journal of Economic Perspectives (2,34)	0,2	0,2	0,1
5	Journal of Political Economy (2,32)	1	1	1
6	Economic Policy (2,21)	0,5	0,1	
7	NBER Macroeconomics Annual (2,00)	1		
8	Health Economics (1,97)			
9	Journal of Health economics (1,92)	0,1	0,1	0,1
10	Economic Geography (1,91)	0,1	1	1
	Итоговый показатель:	13,6	14,4	12,3
	В процентном отношении (коэффициент информационной наполненности, КИН):	46	49	42

Мы видим, что интегральный показатель, каковым является КИН, варьируется от одного ресурса к другому несильно, составляет около 45% (что говорит в целом о не очень хорошей представленности ведущих экономических журналов в электронных библиотечных ресурсах) и максимальное значение (49%) имеет для ProQuest.

6. Качественный анализ электронных ресурсов по дисциплинам: бизнес

Теперь мы можем аналогичным образом считать КИН электронных ресурсов по различным дисциплинам и анализировать результаты. Для анализа нами выбраны бизнес, социология, политология и право — дисциплины, являющиеся профильными для ГУ ВШЭ.

Данные по представлению ведущих журналов по бизнесу и КИН ресурсов приведены в таблице 10.

Таблица 10. Десятка ведущих журналов по бизнесу. КИН

Место	Индекс цитирования 2000	EBSCO	ProQuest	InfoTrac
1	Academy of Management Review (3,91)	FT	FT	9 x
2	Administrative Science Quarterly (3,33)	FT	FT	FT
3	California Management Review (2,88)	FT	FT	Ab
4	Harvard Business Review (2,56)	FT	Ab	Ab
5	Strategic Management Journal (2,53)	Ab	Ab	Ab
6	Journal of Consumer Research (2,46)	FT	FT	FT
7	Academy of Management Journal (2,38)	FT	FT	9 x
8	Journal of Marketing (2,04)	FT	FT	Ab
9	Sloan Management Review (1,79)	FT	FT	FT
10	Journal of Marketing Research (1,68)	FT	FT	Ab
	Итоговый показатель:	23,3	21,0	13,5
	КИН (%):	91	82	53

Только один журнал не представлен полными текстами. Остальные издания в полнотекстовом виде отражены в EBSCO, причем эмбарго на них нет. В случае ведущих изданий по бизнесу мы имеем довольно редкую возможность выстроить электронные ресурсы в строгом порядке по «представленности» *всех единичных изданий*: нет ни одного журнала, который был бы представлен в InfoTrac лучше, чем в ProQuest; а в EBSCO, в свою очередь, каждое издание

представлено так же, как в ProQuest, или в более полном виде (в случае Harvard Business Review). Посчитанные КИН показывают, что наполнение EBSCO и ProQuest по бизнесу очень качественное. InfoTrac значительно отстает, но имеет приемлемый коэффициент.

7. Качественный анализ электронных ресурсов по дисциплинам: социология

Представление ведущих изданий по социологии в электронных библиотечных ресурсах отражено в таблице 11.

Таблица 11. Десятка ведущих журналов по социологии. КИН

Место	Индекс цитирования 2000	EBSCO	ProQuest	InfoTrac
1	American Sociological Review (3,26)	Ab	FT	Ab
2	American Journal of Sociology (2,83)	FT	FT	FT
3	Annual Review of Sociology (2,81)	Ab	12	FT
4	Journal of Marriage & the Family (2,14)	FT	FT	Ab
5	Discourse & Society (1,80)	24		Ab
6	Law & Society Review (1,78)	12	12	Ab
7	Sociology of Health & Illness (1,48)	12	Ab	Ab
8	Sociology (1,41)	Ab	Ab	FT
9	Politics & Society (1,38)	12	12	36
10	Journal of Leisure Research (1,30)	FT	FT	FT
	Итоговый показатель:	11,2	12,8	9,7
	КИН (%):	55	63	48

Полнотекстовая версия каждого журнала присутствует хотя бы в одном ресурсе, причем шесть из десяти изданий поступают как минимум в одну систему без запаздывания. Хуже всего представлен занимающий пятое место по индексу цитируемости «Discourse & Society»: полный текст есть только в EBSCO, причем с двухгодичным эмбарго. Ни один ресурс не превосходит какой-либо другой по всем журналам. В InfoTrac отсутствуют пять полных текстов из изданий десятки, в EBSCO и ProQuest — по три. При этом в ProQuest отсутствуют полнотекстовые версии менее цитируемых изданий, потому его КИН ощутимо выше, чем у EBSCO. В целом КИН ресурсов не очень высок, что говорит о недостаточно полной представленности ведущих социологических журналов.

8. Качественный анализ электронных ресурсов по дисциплинам: политология

В таблице 12 приведены импакт-факторы наиболее цитируемых изданий по политологии и их представление в онлайн-базах данных.

Таблица 12. Десятка ведущих журналов по политологии. КИН

Место	Индекс цитирования 2000	EBSCO	ProQuest	InfoTrac
1	American Political Science Review (2,21)	Ab	FT	9 x
2	American Journal of Political Science (1,79)	12	36	Ab
3	New Left Review (1,72)	Ab		Ab
4	Review of International Political Economy (1,61)	12		
5	Political Science (1,38)			
6	Political Geography (1,23)		Ab	Ab
7	East European Politics & Societies (1,11)	FT		FT
8	Comparative Political Studies (1,07)	12	12	Ab
9	Comparative Politics (0,90)	Ab	Ab	Ab
10	Journal of Democracy (0,88)	Ab	Ab	Ab
	Итоговый показатель:	3,9	3,4	1,9
	КИН (%):	28	24	13

Налицо крайне низкая степень представленности политологических журналов в электронном виде. Пяти полных текстов нет ни в одной базе данных. Одновременный доступ к пяти остальным полнотекстовым версиям также не дает ни одна система. Одно издание («Political Science») вообще никак не проиндексировано в ресурсах. Ни одна из баз не превосходит какую-либо из двух других по всем журналам. КИН для политологических изданий низок у всех ресурсов и не превышает 28%.

9. Качественный анализ электронных ресурсов по дисциплинам: право

Как мы уже отмечали, данные по импакт-факторам юридических изданий за 2000 г. остались нам недоступны. Поэтому мы пользуемся здесь индексами цитируемости журналов по состоянию на 1998 г. Представление ведущих изданий по праву в электронных библиотечных базах данных показано в таблице 13.

Таблица 13. Десятка ведущих журналов по праву. КИН

Место	Индекс цитирования 1998	EBSCO	ProQuest	InfoTrac
1	Harvard Law Review (6,46)	Ab	Ab	Ab
2	Yale Law Journal (6,01)	Ab	FT	FT
3	Columbia Law Review (4,61)	Ab	Ab	Ab
4	Michigan Law Review (4,14)	FT	FT	FT
5	Stanford Law Review (4,11)	Ab	Ab	FT
6	California Law Review (4,02)	Ab	Ab	FT
7	University of Pennsylvania Law Review (3,83)	FT	Ab	Ab
8	American Journal of International Law (3,83)	Ab	6	Ab
9	Journal of Legal Studies (3,76)	Ab		Ab
10	Cornell Law Review (3,73)			FT
	Итоговый показатель:	11,3	15,3	24,3
	КИН (%):	25	34	55

Как видно, представленность ведущих изданий по праву не очень высока в InfoTrac, а в EBSCO и ProQuest еще более низкая. В полнотекстовом виде ни в одной системе нет трех журналов, в том числе занимающих первое и третье места в рейтинге цитируемости. Максимальное число полных текстов, доступных из одной базы данных, составляет 5 — для InfoTrac. Преимущества по всем изданиям ни у одной системы нет.

10. Сводный качественный анализ электронных ресурсов

Наконец, мы можем объединить результаты, полученные нами в подразделах 5–9, и оценить качество контента электронных библиотечных ресурсов, рассматриваемых нами, для всей совокупности исследуемых экономических и социальных дисциплин. Результат представлен в сводной таблице 14.

Таблица 14. Сводная таблица КИН электронных библиотечных ресурсов, %

Дисциплина	EBSCO	ProQuest	InfoTrac
Экономика	46	49	42
Бизнес	91	82	53
Социология	55	63	48
Политология	28	24	13
Право	25	34	55
Средний КИН	49	50	42

Отметим прежде всего, что сравнение самих по себе индексов цитируемости в разных дисциплинах, как уже отмечалось, бессмысленно. Бессмысленно также сравнение абсолютных показателей информационной наполненности (графа «Итоговый показатель» таблиц 9–13), взятых для различных дисциплин. Однако выраженные в процентах коэффициенты информационной наполненности (КИН) уже являются относительными величинами, и их междисциплинарное сравнение и вычисление среднего возможны.

Анализ таблицы 14 показывает, с одной стороны, что каждый ресурс имеет как минимум одну дисциплину, в которой качество его контента, посчитанное по десятке наиболее цитируемых журналов, будет выше, чем у конкурентов. С другой стороны, по четырем дисциплинам (за исключением права), один из ресурсов, InfoTrac, уступает одновременно обоим другим системам. Что касается EBSCO и ProQuest, то их конкуренция в представлении экономических и общественно-научных изданий дает, согласно таблице 14, более сложную картину. Несмотря на то, что EBSCOhost достигает высочайшей для всех трех ресурсов величины КИН по одной из дисциплин (бизнес: 91%), по трем из других отраслей знания он уступает ProQuest, так что средний коэффициент информационной наполненности оказывается для двух ресурсов практически одинаковым.

Быть может, один из основных результатов, который можно почерпнуть из сводной таблицы и ее сравнения с данными раздела I нашего исследования, — это выявленное расхождение количественных и качественных характеристик электронных библиотечных ресурсов. Если при анализе численных показателей безусловным лидером был EBSCO (см., например, таблицу 1), то сравнение КИН не продемонстрировало его явного превосходства, и даже наоборот, по большому числу дисциплин первенствует ProQuest, имеющий также чуть более высокий средний КИН. Это доказывает правоту нашего основного предположения о недостаточности сопоставления только количественных характеристик ресурсов и необходимости тщательного анализа *качества* контента.

Наконец, обратим внимание на то, что средний КИН по общественным наукам не высок у всех ресурсов и не поднимается за отметку 50%. Отсюда можно заключить, что, формируя качественную подписку на иностранную научную периодику, никак нельзя приобретать только

одну базу данных и, кроме того, на данный момент не следует, вероятно, опираться исключительно на электронные ресурсы, полностью уходя от печатного комплектования¹⁹.

Заключение

Анализ информативности баз данных периодических изданий должен включать в себя как количественную, так и качественную составляющие. В то время как в области количественных исследований имеются определенные наработки и проблеме уделяют все больше внимания, изучение и сопоставление *качества* контента являются практически совсем не разработанной областью. Обширные исследования профессионального уровня печатных периодических изданий и методов определения их импакт-фактора, обильные публикации на эту тему проходят отдельно и вне связи с попытками оценить различные электронные системы баз данных, их ценность и актуальность для библиотек и пользователей. Например, даже в [2], где непосредственно идет речь о данных ISI, используется только факт *существования* импакт-фактора издания в «Science Citation Index» для отнесения электронного журнала к «научным» или «ненаучным». Абсолютное значение индекса цитируемости ISI не учитывается. При этом проблема, безусловно, существует и «ощущается»: например, как сообщает Курт Холлеман, «после того как консорциум OhioLink с помпой приобрел для своих участников полный доступ ко всем электронным журналам издательства Elsevier, ряд библиотекарей высказали критическое отношение к этой сделке по причине ненужности многих изданий для большинства библиотек и подняли <...> вопрос о качестве» [29, p. 704].

Потому основной целью настоящей работы было разработать или хотя бы приступить к разработке *методики* оценки качества контента электронных ресурсов периодических изданий. Важно было объединить вопросы о контенте онлайн-баз данных, о видах представления включенных в них статей и об индексе цитируемости печатных изданий, чтобы создать таким образом искомый метод информационного анализа.

Помимо этого, получены предварительные результаты применения данной методики для социальных и экономических наук. Таблицы 9–13 введенных нами коэффициентов информационной наполненности (КИН) по дисциплинам, а также сводная таблица 14 могут использоваться не только библиотекарями для решения о подписке на тот или другой ресурс и определения оптимальных «сценариев подписки», но также и самими производителями систем, агрегаторами, — для активизации переговоров с издателями на тех направлениях, по которым КИН ресурса невелик.

¹⁹ Конечно, ряд изданий могут быть представлены в других системах, не затронутых нашим исследованием, но отсутствие журнала или существенное эмбарго на его полные тексты часто бывает продиктовано намеренной политикой издательства, считающего нецелесообразным предоставлять доступ к электронной версии журнала тем, кто не является подписчиком бумажной копии.

Принципиальным итогом анализа качества контента стало и заметное отличие его результатов от изложенного в первом разделе количественного обзора ресурсов. Это не только дает качественному анализу право на существование, но подчеркивает его превосходство над чисто количественными исследованиями. Дело в том, что важность данного вопроса не ограничивается только финансовой стороной и опасностью потратить библиотечные средства на бессодержательный ресурс. Библиотека не просто теряет деньги — она приобретает *низкокачественную литературу*. Учитывая многие преимущества именно электронного доступа и тот факт, что самое доступное, как правило, становится самым используемым и читаемым, библиотека, собрав электронные фонды периодики сомнительного качества, рискует сформировать второсортный круг чтения — особенно у студентов и начинающих исследователей. Лучшие издания, авторы и публикации могут остаться за скобками, а все внимание будет поглощено освоением обильного, удобного в использовании, но низкопрофессионального массива «агрегированной» информации. Таким образом, при рассмотрении принципов комплектования библиотеки анализ качества контента электронных ресурсов даже *более важен*, чем аналогичная политика в отношении печатных фондов. Этот вывод должен служить мотивацией для дальнейшего совершенствования предложенных в нашей работе методик.

В качестве будущего развития исследования в первую очередь можно предложить количественный анализ представленности в различных ресурсах *ведущих* изданий. Проведя анализ, аналогичный выполненному в разделе I, но не для всей массы журналов, а только для тех, что занимают первые места в рейтингах ISI, можно сравнить его с результатами, полученными при совокупном количественном анализе. При этом мы отказываемся от учета абсолютного значения импакт-фактора (важно только, чтобы издание входило в десятку) и вида представленности журнала (издания с эмбарго учитываются как полнотекстовые, издания с аннотациями вообще не учитываются), зато методика становится абсолютно идентичной процедуре количественного анализа всех изданий в разделе I. Построив таблицы, аналогичные таблицам 1–5, можно было бы провести их сопоставление как с результатами совокупного количественного анализа, так и с итогами качественного исследования раздела II.

Также интересно было бы провести качественный анализ представленности ведущих журналов в базах данных «попарно» по ресурсам: насколько хорошо охватывают десятку наиболее цитируемых изданий по различным дисциплинам ProQuest и EBSCO вместе, EBSCO и InfoTrac вместе, InfoTrac и ProQuest вместе (разумеется, КИН в таблицах 9–14 не являются аддитивными по различным ресурсам). Эти результаты, «парные коэффициенты информационной наполненности», будут важны при выборе библиотеками стратегии расширения своей подписки на электронные базы данных периодики — в том случае, если уже имеется один из ресурсов. Наряду с количественными сведениями о «пересечении» и, наоборот,

«несовпадении» контентов, приведенными в разделе I настоящей работы, такие «парные КИН» станут ориентиром при составлении и выборе «сценария подписки».

Весьма важным представляется дальнейшее развитие исследования корректности использования и авторитетности индекса цитируемости, вычисляемого ISI. Например, необходимо изучить колебания импакт-фактора из года в год — и, как следствие, изменение КИН баз данных. В какой мере КИН являются стабильными — следовательно, в какой мере можно опираться на них в долгосрочной перспективе? Отражают ли они реальный информационный вес баз данных и тщательность политики их производителей (которые должны стремиться заключать соглашения с наиболее авторитетными изданиями) — или же это всего лишь результат случайной конъюнктуры, сложившейся на данный год, выведшей очередные журналы из забвения в «хит-парад» и *top ten*, а вместе с ними поднявшей КИН ресурса, которому в данный момент просто «повезло»? Все это заслуживает тщательного рассмотрения.

Отметим в этой связи, что осуществление исследования, аналогичного нашему, но с *классификацией журналов по издательствам*, может дать картину реального уровня активности сотрудничества между производителями электронных ресурсов и теми или иными издательскими домами, а изменение наполненности баз данных авторитетными изданиями какого-либо издателя будет сигнализировать о произошедших изменениях в деловых отношениях. Логично предположить, что чем выше реальный авторитет журнала, следовательно, чем выше его индекс цитируемости, тем бóльшую цену вправе запрашивать издательство за размещение его полных текстов в электронной базе данных, тем сложнее агрегатору договориться о включении издания в свой ресурс, тем более весомой, наконец, можно считать состоявшуюся сделку. На значительность соглашения будет влиять не только величина импакт-фактора, но, конечно, и вид представления журнала, установленное издательством эмбарго, — фактор, также учитывающийся при подсчете КИН. Не исключено поэтому, что КИН конкретного ресурса, посчитанный не для дисциплин, а для издательств, будет являться вполне корректным индексом деловой активности между данным агрегатором и академическими издательствами, а ряд аналогичных КИН для всех производителей электронных библиотечных систем даст бизнес-карту рынка информационных ресурсов периодики.

Совершенствуя методику оценки качества контента электронных ресурсов, надо помнить и о другой стороне ограниченной применимости публикуемого в *Journal Citation Reports* импакт-фактора. Показатель *средней* цитируемости статьи не является идеальным при оценке *интегральной* ценности издания. Например, журнал, выходящий с пятью статьями два раза в год, и журнал, выходящий ежемесячно с 20 статьями того же уровня (столь же часто цитируемыми в других публикациях), будут иметь одинаковый импакт-фактор, поскольку среднее число цитирования статей одинаково. Тем не менее, вряд ли два таких издания представляют равную

ценность для библиотеки²⁰. Таким образом, вместо количества цитирований, деленного на количество опубликованных статей, следует, быть может, использовать просто количество цитирований (то есть фактически величину импакт-фактора ISI, умноженную на число вышедших за два предыдущих года публикаций в данном издании). Умножив этот показатель на коэффициент представленности, просуммировав по десятке журналов и нормировав результат аналогично тому, как это делалось при вычислении КИН, мы получим выраженный в процентах «собранный» по статьям лидирующих журналов, «агрегированный» коэффициент информационной наполненности, АКИН. Отметим, что если задать линейную зависимость коэффициента представленности от эмбарго $k = 1 - E/24$, где E — эмбарго, выраженное в месяцах, то АКИН будет оценкой вероятности того, что, найдя в свежем периодическом издании ссылку на статью, опубликованную в журнале из ведущей десятки (для данной дисциплины) за последние 2 года, нам удастся найти цитируемую публикацию в электронной базе данных, для которой посчитан АКИН.

В известном смысле пределом такого рода вычислений и совершенствования методики был бы выход за рамки десятки рейтинга и подсчет количества цитирований в текущем (или прошедшем) году всех включенных в данный ресурс статей. Таким образом мы, с одной стороны, анализируем все журналы, а не только по десять из каждой дисциплины. С другой стороны, в отличие от простого количественного анализа, мы считаем не наименования журналов, не количество их выпусков, даже не количество статей, но число актуальных (свежих) цитат на включенные в базу данных публикации. Этот метод является сочетанием количественного и качественного анализов, а вычисленный коэффициент будет равен вероятности найти в ресурсе статью, ссылку на которую исследователь встретит в каком-либо журнале за предыдущий год. Дисциплинарные границы, очевидно, необходимо оставить, однако оценка «коэффициентов представленности» для разных типов представления изданий (подобная таблице 6) будет упразднена. Методика практически безупречна, однако непросто реализуема — по крайней мере никакого готового программного продукта, осуществлявшего бы накопление и предоставление необходимых показателей, на данный момент нет. Необходимо вручную просматривать несколько десятков журналов по каждой дисциплине и комбинировать данные из *Journal Citation Reports* со списками включенных в базы данных изданий.

Наконец, естественно предположить, что индекс цитируемости печатных изданий и коэффициент информационной наполненности электронных ресурсов *взаимно* влияют друг на друга. Если в базах данных хорошо представлена какая-либо дисциплина, это не только ведет к

²⁰ Именно в этом смысле Стивен Хартер и Томас Низонгер инициировали дискуссию о некорректном использовании Институтом научной информации термина «journal impact factor» (досл. «коэффициент влияния журнала»). По их мнению, подсчеты ISI определяют «article impact factor» («коэффициент влияния статьи»), а для оценки авторитетности всего издания необходимо брать само количество цитирований, а не производить его нормировку на число вышедших в журнале публикаций [30]. Ответ Гарфильда см. в [31].

увеличению соответствующего КИН ресурса, но и способствует бóльшей доступности авторитетных журналов по данной теме для академического сообщества. Если учесть, что электронные коллекции периодики предоставляют широкие и удобные возможности для поиска информации, столь важного в научном процессе, то можно предположить, что к соответствующим электронным журналам специалисты будут чаще обращаться и чаще их цитировать, что, в свою очередь, поднимет индекс цитируемости отраженных в ресурсе изданий. Анализ такой «обратной связи» также может стать прекрасной темой для дальнейшего развития нашего исследования.

Литература

1. *Krumenaker L.* A Tempest in a Librarian's Teapot. EBSCO, ProQuest, Gale: Exclusive and Unique Titles // *Searcher: The Magazine for Database Professionals*. 2001. Vol. 9. Iss. 7 (July/August). P. 40–45.
2. *Hill J. B.* Aggregated Science: An Examination of Three Multidisciplinary Databases // *Issues in Science and Technology Librarianship*. Spring 2001 // <http://www.library.ucsb.edu/istl/01-spring/article4.html>.
3. *Curtis D.* EBSCO and ProQuest database content // <http://www.library.yale.edu/~llicense/ListArchive/0105/msg00145.html> (онлайн-сообщение, отправленное 24.05.2001).
4. *Krumenaker L.* Tempestuous Arguments // <http://www.library.yale.edu/~llicense/ListArchives/0107/msg00075.html> (онлайн-сообщение, отправленное 24.06.2001).
5. *Conger J.* The GALILEO Comparison of EBSCOhost and ProQuest Product Suites. Paper, presented at COMO/GLA. Jekyll Island Convention Center, 2001 // http://www.arches.uga.edu/~jconger/compare_GLApaper.doc.
6. *Goodman D.* Re: EBSCO and ProQuest database content. <http://www.library.yale.edu/~llicense/ListArchives/0105/msg00150.html> (онлайн-сообщение, отправленное 25.05.2001).
7. *Sampson W.* RE: Two articles on aggregator exclusivity deals and embargoes // <http://www.library.yale.edu/~llicense/ListArchives/0107/msg00071.html> (онлайн-сообщение, отправленное 23.07.2001).
8. *Sampson W.* RE: Tempestuous Arguments // <http://www.library.yale.edu/~llicense/ListArchives/0107/msg00079.html> (онлайн-сообщение, отправленное 25.07.2001).
9. *Garfield E.* The Impact Factor // *Current Contents*. 1994. No. 25. P. 3–7 // <http://www.isinet.com/isi/hot/essays/journalcitationreports/7.html>.
10. *Garfield E.* Using the impact factor. *Current Contents*. 1994. No. 29. P. 3–5 // <http://www.isinet.com/isi/hot/essays/journalcitationreports/8.html>.
11. Social Sciences Citation Index Journal List // <http://www.isinet.com/cgi-bin/jrnlst/jlresults.cgi?PC=J&Alpha=A>.

12. Journals Ranked by Impact: Economics // http://in-cites.com/research/2001/november_12_2001-2.html.
13. *Vocino T., Elliott R. H.* Journal prestige in public administration // *Administration and Society*. 1982. Vol. 14. P. 5–14.
14. *Garfield E.* Prestige versus impact: Established images of journals, like institutions, are resistant to change // *Essays of an Information Scientist*. Philadelphia, 1989. Vol. 10. P. 263–264.
15. *Christenson J. A., Sigelman L.* Accrediting knowledge: Journal stature and citation impact in social science // *Social Science Quarterly*. 1985. Vol. 66. No. 4. P. 964–975.
16. *Colson H.* Citation Rankings of Public Administration Journals // *Administration and Society*. 1990. Vol. 21. Iss. 4. P. 452–471.
17. *Moore W. H.* Citation versus Reputation: Assessing Political Science Journals. Working paper. 2000. August // http://www.fsu.edu/~polisci/wksp_papers/workingpapers/p_20004wmoore.pdf.
18. *Roche T., Smith D. L.* Frequency of citations as criterion for the ranking of departments, journals, and individuals // *Sociological Inquiry*. 1978. Vol. 48. P. 49–57.
19. *Cole S.* The Hierarchy of the Sciences // *American Journal of Sociology*. 1983. Vol. 89. July. P. 111–139.
20. *Hargens L. L., Felmler D.* Structural Determinants of Stratification in Science // *American Sociological Review*. 1984. Vol. 49. October. P. 685–697.
21. Journals Ranked by Impact: Business // http://www.in-cites.com/research/2001/december_10_2001-2.html.
22. Journals Ranked by Impact: Sociology // http://www.in-cites.com/research/2001/oct_29_2001-1.html.
23. *Brooks S.* Re: EBSCO and ProQuest database content. // <http://www.library.yale.edu/~license/ListArchives/0105/msg00158.html> (онлайн-сообщение, отправленное 27.05.2001).
24. *Quint B.* Gale Group's InfoTrac OneFile Creates Web-Based Periodical Collection for Libraries // *Information Today*. 2000. November. Vol. 17. Iss. 10. P. 12.
25. Scope Notes. Social Sciences Citation Index // <http://sunweb.isinet.com/isi/journals/scope/ssci/index.html>.
26. Journals Ranked by Impact: Political Science // http://in-cites.com/research/2002/january_7_2002-2.html.
27. *Маршакова-Шайкевич И. В.* Анализ вклада России в развитие социальных и гуманитарных наук // *Вопросы философии*. 2000. № 8. С. 139–149.
28. *Giles M. W., Mizell F., Patterson D.* Political Scientists' Journal Evaluations Revisited // *PS: Political Science and Politics*. 1989. Vol. 22. Iss. 3. P. 613–617.
29. *Holleman C.* Electronic resources: Are Basic Criteria for the Selection of Materials Changing? // *Library Trends*. 2000. September. Vol. 48. No. 4. P. 694–710.

30. *Harter S. P., Nisonger Th. E.* ISI's Impact Factor as Misnomer: A Proposed New Measure to Assess Journal Impact // *Journal of the American Society for Information Science*. 1997. December. Vol. 48. No. 12. P. 1146–1148.
31. *Garfield E.* The Multiple Meanings of Impact Factors // *Journal of the American Society for Information Science*. 1998. Vol. 49. No. 8. P. 768.
32. *Garfield E.* Long-term vs. Short-Term Journal Impact: Does It Matter // *The Scientist*. 1998. Vol. 12 (3). P. 11–12.

Оглавление

Введение	3
I. Количественный анализ контента электронных баз данных зарубежной периодики	5
II. Качественный анализ контента электронных баз данных зарубежной периодики	9
1. Индексы цитируемости ISI	10
2. Использование индексов цитируемости. Индекс цитируемости и «престиж» журнала	11
3. Виды представления журналов. Коэффициенты представленности	13
4. Базы данных, выбранные для анализа	15
5. Качественный анализ электронных ресурсов по дисциплинам: экономика	17
6. Качественный анализ электронных ресурсов по дисциплинам: бизнес	19
7. Качественный анализ электронных ресурсов по дисциплинам: социология	20
8. Качественный анализ электронных ресурсов по дисциплинам: политология	21
9. Качественный анализ электронных ресурсов по дисциплинам: право	21
10. Сводный качественный анализ электронных ресурсов	22
Заключение	24
Литература	28

Препринт WP2/2002/02
Серия WP2
Количественный анализ в экономике

Писляков Владимир Владимирович

Анализ контента ведущих электронных ресурсов актуальной зарубежной периодики

Публикуется в авторской редакции
Оформление серии *А.М.Павлов*
Корректор *Е.Е.Андреева*

ЛР № 020832 от 15 октября 1993 г.

Отпечатано в типографии ГУ ВШЭ с представленного оригинал-макета.

Формат 60×84 $\frac{1}{16}$. Бумага офсетная. Тираж 150 экз. Уч.-изд.л. 2,35. Усл.печ.л. 1,98.

Заказ №57. Изд. №201.

ГУ ВШЭ. 117312, Москва, ул.Вавилова, 7а

Типография ГУ ВШЭ. 125319, Москва, Кочновский проезд, 3

Препринты ГУ ВШЭ
Серия WP2 "Количественный анализ в экономике"

<http://stat.hse.ru/>

2002

Писляков В.В. Анализ контента ведущих электронных ресурсов актуальной зарубежной периодики: Препринт WP2/2002/02 – М.: ГУ ВШЭ, 2002. – 33 с.

Губанов В.А. Непараметрическое выделение динамических сезонных циклов: Препринт WP2/2002/01 – М.: ГУ ВШЭ, 2002. – 33 с.

2001

Поспелов И.Г. Экономические агенты и системы балансов: Препринт WP2/2001/03 – М.: ГУ ВШЭ, 2001. – 68 с.

Бессонов В.А. Об измерении динамики российского промышленного производства переходного периода: Препринт WP2/2001/02 - М.: ГУ ВШЭ, 2001. - 34 с.

Алексеевкова М.В. Факторы отраслевого анализа для российской переходной экономики: Препринт WP2/2001/01 - М.: ГУ ВШЭ, 2001. - 34 с.