

**Системы сбора и анализа статистики использования
электронных ресурсов: сравнительный обзор**
**Retrieval and Analysis Systems of Statistical Data
on the Use of Electronic Resources: Comparative Analysis**

**Системи збору й аналізу статистики використання
електронних ресурсів: порівняльний огляд**

Писляков В. В.

Библиотека Государственного университета — Высшей школы экономики, Москва, Россия

Vladimir V. Pislyakov

Library of the State University — Higher School of Economics, Moscow, Russia

Писляков В. В.

Бібліотека Державного університету — Вищої школи економіки, Москва, Росія

Исследуются наборы статистических параметров, предоставляемых различными вендорами — компаниями, продающими онлайн-информационные продукты, — своим подписчикам для анализа востребованности приобретенных ресурсов (на примере платформ Cambridge Journals Online, EBSCO, JSTOR, ProQuest и ScienceDirect). Проводится сравнительный анализ и выявляются показатели, доступные для каждой системы, встречающиеся только у отдельных вендоров и отсутствующие вообще. Затронуты вопросы о возможной интерпретации статистических сведений и об оперативности предоставления статистических данных библиотекам-подписчикам.

Taking the platforms of Cambridge Journals Online, EBSCO, JSTOR, ProQuest and ScienceDirect as an example, the author examines the sets of statistics provided by various vendors—companies selling online information products—to their subscribers for the purpose of studying the demand for the obtained resources. The author conducts a comparative analysis and shows indices either common to all systems, typical of individual vendors, or absent in general. The author touches upon the issues of possible interpretation of statistical data and the efficiency with which such data are provided to the subscribing libraries.

Досліджено набори статистичних параметрів, що надаються різними вендорами — компаніями, що продають інформаційні продукти он-лайн, — своїм передплатникам для аналізу попиту на придбані ресурси (на прикладі платформ Cambridge Journals Online, EBSCO, JSTOR, ProQuest і ScienceDirect). Проведено порівняльний аналіз і виявлені показники, доступні для кожної системи, які зустрічаються тільки в окремих вендорів або відсутні взагалі. порушено питання про можливу інтерпретацію статистичних відомостей і про оперативність надання статистичних даних бібліотекам-передплатникам.

1. Постановка проблемы

Аналитическое исследование рынка электронных ресурсов, проведение переговоров и оформление подписки на базы данных для библиотеки — только часть ответственности библиотекаря-специалиста по электронным информационным ресурсам. Не менее важной является «постподписная» фаза работы: анализ востребованности ресурса, его реклама среди пользователей и обучение. Эта стадия освоения информационной системы влияет на то, будет ли подписка на ресурс возобновлена по истечении первого подписного периода или же ресурс будет признан не соответствующим информационным потребностям читательской аудитории и подписка на него в дальнейшем будет прекращена.

Объективные выводы о востребованности базы данных и о влиянии различных просветительских мероприятий (тренингов, семинаров, email-рассылок и пр.) позволяет сделать *анализ статистики активности использования ресурса*. Такого рода статистика предоставляется каждым ресурсом на специальном административном сайте, доступ к которому имеет т. н. «локальный администратор» ресурса — сотрудник библиотеки, отвечающий и следящий за подпиской на эту базу данных. Ввиду того, что статистические сведения поставляет каждый вендор (компания-продавец информационного ресурса) по отдельности, наполнение статистических отчетов может быть разным для разных вендоров.

Целью настоящего доклада является сравнение предоставляемых вендорами параметров статистики использования электронных информационных систем и рассмотрение вопросов универсализации статистических отчетов.

2. Выбор систем, вошедших в анализ

Библиотека Государственного университета — Высшей школы экономики (ГУ-ВШЭ) создана в 1994 г. и является небольшой по меркам московских университетских библиотек: в настоящий момент печатный

фонд составляет ок.55 тыс. наименований и ок.300 тыс. экземпляров. Четыре года назад началось развитие электронной подписки на базы данных зарубежной периодики, в период с апреля по ноябрь 2001 г. была осуществлена подписка на ресурсы компаний JSTOR, ProQuest и EBSCO, которые до сих пор являются «ядром» электронного фонда библиотеки и используются очень активно. С 2004 г. также была оформлена подписка на журналы издательства Elsevier (платформа ScienceDirect) и издания Кембриджского университета (Cambridge Journals Online, CJO). Именно эти пять ресурсов взяты для анализа в настоящем обзоре. Каждый из них имеет свою систему предоставления статистики пользователю, особенности которой и войдут в наш доклад. При этом анализ будет касаться только содержательной стороны статистических отчетов (т. е. набора параметров, который может быть извлечен): технологические особенности и специфика интерфейсов не так важны и потому останутся за рамками исследования.

3. Результаты

В первой таблице сведены показатели, которые фиксируются сразу многими вендорами и являются, таким образом, основными:

Таблица 1

	полнотекст. статьи	аннотации/ биб. описания	поиски	сессии	IP	наимен-я
CJO	+	+				+
EBSCO	+	+	+	+	+	+
JSTOR	+	+	+		+	+
ProQuest	+	+	+	+		+
Science Direct	+	+	+	+	+	+

С одной стороны, анализ данной таблицы показывает, что все «основные» показатели предоставляются только двумя ресурсами — EBSCO и ScienceDirect, в то время как наименьший спектр данных фиксирует CJO (что, вероятно, объясняется тем, что онлайн-система Кембриджского издательства относительно молодая и еще не развита в совершенстве).

С другой стороны, взгляд на таблицу позволяет выделить «обязательные» показатели, которые подсчитываются всеми системами. В нашем случае это число открытых полнотекстовых статей, число просмотренных аннотаций (и/или библиографических описаний) и распределение статистики по наименованиям журналов. Указанные параметры, фиксируемые всеми вендорами, чрезвычайно важны. Во-первых, именно их следует использовать при *сравнении* использования различных баз данных в организации. Так, наилучшим показателем для компаративного анализа востребованности электронных ресурсов является число открытых полнотекстовых статей. Во-вторых, параметры, подсчитываемые всеми статистическими системами, являются аддитивными и могут служить для определения совокупной активности использования электронных ресурсов, когда вся онлайн-подписка библиотеки рассматривается как единый информационный массив.

Стоит также отметить, что те же самые показатели легли в основу рекомендаций для представления статистики по использованию электронных ресурсов — рекомендаций, выработанных специально собранными для этого экспертами «Project COUNTER»¹. Вот как выглядят основные стандартные отчеты, рекомендованные COUNTER:

- Journal Report 1: Number of Successful Full-Text Article Requests by Month and Journal
- Database Report 1: Total Searches and Sessions by Month and Database

Как видим, в эти отчеты входит количество загруженных полнотекстовых статей с распределением по наименованиям журналов, число поисков и сессий — все те параметры, которые фигурировали в таб.1.

Теперь перейдем к полезным показателям, которые, однако, встречаются в статистических отчетах лишь некоторых вендоров. Таковых три:

- среднюю продолжительность сессии считают EBSCO и ScienceDirect (этот параметр может считаться мерой «тщательности просмотра» ресурса пользователями);
- пути следования к полным текстам считают JSTOR и ScienceDirect (этот параметр позволяет оценить отношение частоты использования функции Search к функции Browse, тем самым давая представление о полноценности использования поисковых возможностей ресурса);
- частоту использования различных типов поиска фиксируют ProQuest и ScienceDirect (этот показатель также позволяет оценить уровень владения поиском у пользователей)

¹ Counter Code of Practice, Release 1: December 2002 // http://www.projectcounter.org/code_practice.html.

Наконец, можно отметить несколько немаловажных параметров, каждый из которых представлен только в одной системе из пяти.

Для JSTOR это:

- число просмотренных полнотекстовых страниц (параметр даже более точный, чем количество открытых статей, однако не всегда могущий быть посчитанным ввиду технологических особенностей платформ);
- объем распечатанных материалов (также весьма говорящий показатель, но тоже зависящий от способа представления полнотекстовых статей);
- дисциплинарное распределение востребованных материалов (позволяет выявить картину тематических интересов пользователей);
- посуточная активность (в частности, позволяет исследовать влияние тренингов, семинаров, иных мероприятий на использование ресурса);
- сравнение статистики использования с другими организациями-подписчиками.

Для ScienceDirect:

- число пользовательских компьютеров, с которых совершались заходы в систему (параметр не менее информативен, чем считаемое многими системами количество сессий);
- активность использования «персональных профилей» (параметр, позволяющий проследить поведение пользователей, зарегистрировавшихся и создавших «персональную зону» на ресурсе и потому вероятнее всего составляющих его «читательское ядро»).

Для ProQuest:

- распределение по выбранному способу доставки (позволяет получить соотношение между числом загрузок статей в браузер и числом публикаций, отправленных по электронной почте или факсом).

4. Обсуждение результатов

Как выяснилось, большинство вендоров предоставляют ряд ключевых параметров статистики использования своих систем в единообразном формате, так что они пригодны для сравнения востребованности различных электронных ресурсов и определения суммарного спроса на онлайн-подписку организации. На рис.1 показано, как это сделано на сайте библиотеки Государственного университета — Высшей школы экономики (данные за весь 2004 г.):

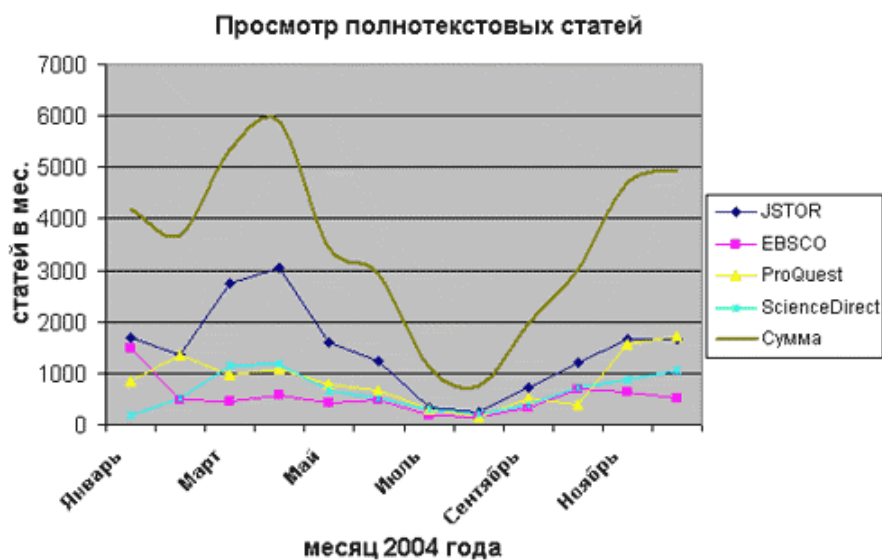


Рис.1

Однако некоторые «показатели второго плана» считаются только частью компаний и даже имеются параметры, предоставляемые только каким-либо одним вендором. Следует также указать на ряд показателей, которые вообще не считаются ни в одной из систем, несмотря на то что их анализ был бы весьма полезен для библиотек-подписчиков. Например, это:

- статистика использования *уникальных* статей, за вычетом дублирования обращений (в настоящий момент все статистические системы считают две загрузки одной и той же статьи как два разных обращения, каждое из которых учитывается в статистике);
- хронологическое распределение запрашиваемых материалов (число обращений к статьям или журналам в зависимости от давности их выхода в свет);
- самые популярные статьи (статистика обращений к конкретным публикациям);
- статистика использования поисковой строки (напр., наиболее часто встречающиеся при поиске ключевые слова).

Наконец, нельзя обойти вниманием *оперативность* предоставления статистики различными вендорами. Чем скорее статистические сведения попадают подписчику (библиотеке), тем быстрее можно проводить их анализ и эффективнее принимать решения на их основании. На рис.2 приведены данные, сопоставляющие запаздывание предоставляемых статистических показателей. Как видим, большая часть компаний выдает статистику уже на следующие сутки: можно обрабатывать сведения за вчерашний день. Кембриджская платформа предоставляет отчеты понедельно, а ScienceDirect готовит данные дольше всех, 3—4 недели. Столь большая задержка уже является критической, особенно для тех подписчиков, которые проводят анализ использования баз данных на регулярной основе. Так, библиотека Высшей школы экономики по этой причине долгое время не включала ScienceDirect в свои ежемесячные обзоры активности пользования электронными ресурсами¹: отчеты готовились к 15-му числу каждого месяца, в то время как данные по ScienceDirect приходили позже. Впрочем, анализ сказанного в 3-м пункте настоящего доклада позволяет заметить, что список параметров, предоставляемых ScienceDirect, один из наиболее полных и в известном смысле объяснимо, что эта статистика готовится столь долго.

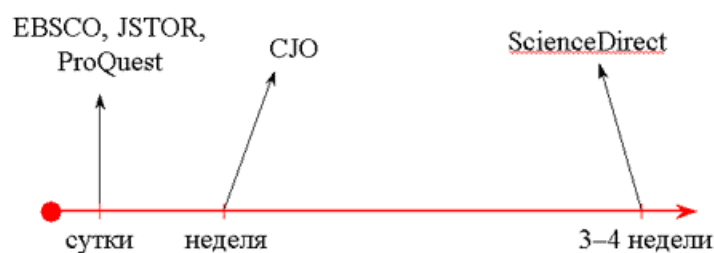


Рис.2

¹ См. <http://library.hse.ru/stat/stat.htm>