

УДК 681.327

ИСПОЛЬЗОВАНИЕ РЕСУРСОВ ИНТЕРНЕТ ДЛЯ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ ИССЛЕДОВАНИЙ

Р. Б. Сейфуль-Мулюков, А. В. Сорокин, О. В. Ломакина

Предлагается технология информационного мониторинга ядра зарубежной периодики, издаваемой по компьютерным и информационным наукам и помещаемой издателями на сайты сети Интернет. Обосновывается процедура выделения ядра 10 % журналов от всей их совокупности, представленной в Интернет по данной проблематике. Рассматривается организационная и экономическая эффективность проведения мониторинга специалистом на постоянной основе в целях информационного обеспечения исследований.

1. Информация в сети Интернет

Роль информации, представляемой в глобальной информационной сети Интернет (далее Интернет), стремительно растёт. Эта информация становится важным фактором социального, экономического и политического развития общества. В полной мере это относится к научно-технической информации. Ещё 20 лет тому назад основным источником информации служило печатное издание, доступ к которому со времён фараонов и римской империи осуществлялся через библиотеки. Библиотечные коллекции, вначале рукописные, а затем печатные, расширялись, пополнялись, и одновременно совершенствовались системы классификации хранящихся в них документов.

Компьютер и Интернет, соединивший в глобальную сеть миллионы компьютеров и электронных каталоги библиотек, открыли принципиально новые возможности использования огромных информационных массивов и доступа к ним. Как это вытекает из диалектики, одновременно и в сопоставимом масштабе, новые возможности породили новые проблемы и трудности, отличные от проблем поиска информации в библиотеке.

В библиотеке, независимо от способа представления её каталога в бумажном (каталожном) или электронном виде, основным средством поиска нужного издания служит алфавит. В библиотеках России он кириллический. Однако, как бы велика ни была коллекция, совершена классификация и алфавитный каталог, нужное издание, как результат поиска, «в живую» оказывалось в руках читателя.

В Интернете всё сложнее. При поиске нужно найти сайт, на котором может храниться нужная информация. Необходимо иметь элементарные навыки навигации на сайте. Поскольку большинство сайтов англоязычные, возникает требование вести поиск на английском языке.

Если говорить о научно-технической информации, то пользователь Интернет чаще «выходит» на аннотацию нужного издания и по её содержанию решает, нужен ли ему первоисточник целиком. В отличие от библиотеки, где издание представляется бесплатно, доступ к электронному первоисточнику, особенно тематическим выпускам, в большинстве случаев платный.

Отличия первоисточника, полученного в библиотеке, и информации в различных её формах, к которой пользователь получает доступ в сети Интернет, не ограничиваются только перечисленными особенностями. Помимо относительной простоты поиска нужного печатного издания в библиотеке, прочтение «живого» первоисточника имеет одно очень важное преимущество — рождение ассоциативных мыслей. Имея перед собой и читая издание, и как бы следуя за ходом мысли и рассуждений автора, читатель не только получает новую информацию, но часто находит решение своей проблемы, ассоциативной идее читаемого первоисточника, может быть, иной и далёкой от своей.

Ассоциативное мышление в психологии — это связь двух представлений, когда одно, ранее появившееся, вызывает в сознании другое (ассоциация идей). Идеи основателя ассоциативной психологии английского врача D. Hartley (1705–1757) популярны и ныне. На новой, компьютерной основе сегодня предпринимаются попытки формализовать процесс ассоциативного мышления, создать его математическую модель [1].

Для проблем, рассматриваемых в данной работе, важно, что ассоциативное мышление учёного можно рассматривать как один из важных побудительных мотивов чтения первоисточника. Поскольку ассоциативное мышление осуществляется в большей степени на подсознательном уровне, чтение короткой аннотации,

да ещё в электронном виде, по-видимому, значительно сокращает возможности ассоциативного мышления или вообще его не вызывает. Это объясняет большое значение первоисточника и необходимость чтения его «в живую».

Приведённое выше, в некоторой степени, философское отступление об ассоциативном мышлении, имеет непосредственное отношение к теме настоящей статьи. Информационное обслуживание учёного должно строиться таким образом, чтобы, в конечном итоге, выбранное самим пользователем издание попало в его руки. Только так исследователь получит возможность работать с первоисточником и воспользоваться преимуществами ассоциативного мышления. Выбор издания среди всех возможностей, предоставляемых системой Интернет, должен обеспечить специально подготовленный специалист.

В последнее десятилетие большинство издателей научно-технической литературы помещают на свои сайты и сайты провайдеров полнотекстовые файлы статей или их расширенные аннотации. Вся совокупность журнальной периодики по дисциплине «информационные и компьютерные науки», представленной на сайте [6], превышает 250 наименований. Казалось бы, Интернет значительно расширяет возможности доступа к первоисточнику. Однако, это справедливо лишь в отношении пользователя, подготовленного к навигации в Интернете и способного читать литературу на английском языке, на котором публикуется около 85 % мирового потока научно-технической литературы.

Нужно иметь в виду, что, помимо журнальной периодики, важные данные содержатся в докладах, сборанных и публикуемых в трудах конференций. По проблемам информатики таких сборников трудов публикуется не менее 100 в год. Мониторинг названного объёма научно-технической литературы отдельно взятый учёный просто физически не может осуществить. Поэтому мониторинг Интернет — информатики и, соответственно, информационное обслуживание исследователей на её основе должны строиться с учётом реальных объёмов мирового потока информации по данной проблеме.

2. Тематический мониторинг Интернет-информации

В рассматриваемом контексте мониторингом будем считать спланированное, целенаправленное, постоянное слежение за выбранными наименованиями журнальной периодики и трудами

конференций определённого тематического спектра. Слежение означает фиксацию данного выпуска журнала на сайте, знакомство с аннотацией статей, предварительную оценку соответствия содержания тематическому спектру исследований института и помещение выбранных аннотаций или электронных версий статей на сервер внутренней корпоративной сети (Интранет). В таком виде эта «информация об информатике» будет доступна для любого пользователя данной научной организации.

В зарубежной литературе тематический мониторинг близок к понятию data mining, что на русском языке означает: «извлечь данные» или, дословно, «накопать данные». Data mining в зарубежной практике во многих случаях означает поиск данных и информации, необходимой для решения определённой проблемы, во всех возможных источниках, представленных на сайтах Интернет. Мониторинг в рассматриваемом контексте — это постоянное наблюдение или слежение за ограниченным, определённым заранее числом источников информации, покрывающих заранее определённую тематику.

При подобном подходе цель мониторинга — фиксировать наиболее значимые в научно-методическом плане и соответствующие тематике исследований института публикации. Подчёркнём, работа специалиста, проводящего мониторинг, это не реферирование зарубежных статей или перевод на русский язык аннотаций к ним. Это не составление обзоров типа экспресс — информации. Мониторинг в данном контексте — это в значительной степени *аналитическая обработка* зарубежной литературы заранее ограниченного тематического спектра. Перед такой работой стоит задача выделить наиболее значимые в научном и методическом плане публикации, содержание которых имеет непосредственное отношение к исследованиям института. В совокупности подобная информация призвана показать исследователям института, пользующимся результатами мониторинга, на каких направлениях и на каком уровне в настоящее время работает творческая мысль за рубежом в данном тематическом направлении.

Перечисленные особенности тематического мониторинга зарубежной периодики по проблеме «Информатика» определяют и требования к специалисту, привлекаемому к этой работе. Мониторинг должен осуществляться на постоянной основе. Его должен выполнять специалист в данной области знаний, хорошо представляющий основные направления исследований института

и владеющий английским языком, а также технологиями навигации на Интернет сайтах.

Специалист, основной обязанностью которого является тематический мониторинг, может рассматриваться как своего рода первопроходец. В результате его работы пользователи в институте получают своеобразное «научное сырьё», с которым они дальше более углублённо могут работать. Мониторинг — это извлечение и первичное обогащение научной информации.

Для института же мониторинг представляется достаточно эффективной затратой средств на информационное обеспечение научно-исследовательского процесса. Взамен многих специалистов из каждого научного отдела, самостоятельно занимающихся поиском нужной Интернет — информации, причём нередко в одних и тех же источниках, один или два специалиста, отвечающие выше перечисленным требованиям, будут готовить информацию из зарубежных источников для всех научных подразделений, делая это на профессиональном уровне и постоянной основе.

Мониторингом можно не охватывать отечественную научную техническую литературу, поскольку доступ к ней не ограничен. Литература в тематическом диапазоне института имеется в его библиотеке. Добавим также, что специалист любого отдела, имеющий опыт работы с зарубежной литературой и многие годы самостоятельно производящий поиск информации на сайтах зарубежных издателей, как правило, не нуждается в помощи. Однако, на примере ИПИ РАН можно констатировать, что на зарубежные источники ссылаются не более 10% от общего числа авторов статей и отчётов [5]. Это свидетельствует о том, что использование исследователями зарубежной научно-технической литературы, сравнительно невелико. Поэтому, проведение тематического мониторинга зарубежной периодики, в качестве эффективного средства для анализа достижений коллег за рубежом, вполне оправдано, как с организационной точки зрения, так и экономически.

Практика показала, что специалист, отвечающий приведённым выше требованиям, может обеспечить тематический мониторинг 20–25 наименований периодических журналов, что соответствует количеству, примерно, 150–180 статей в месяц. При этом в результате его деятельности создаётся информация, включающая:

— название журнала, заглавие статьи и её автор/авторы;

— тематика журнала и статей, привязанных к выбранной для Итранета системе классификации — рубрикации;

— авторская аннотация из первоисточника;

— аннотация в кратком переводе на русский язык для наиболее важных работ, имеющих прямое отношение к тематике научных подразделений института;

— полный текст статьи в оригинале, предоставляемый по запросу пользователя;

— набор ключевых слов, представленных авторами статей для поиска других, аналогичных источников в сети Интернет, подготовленных этими авторами, а также другими, работающими в данной области.

Имея доступ к такому набору информации, пользователь решает, какой из названных видов информации ему интересен и с каким из этих видов информации он будет дальше работать. При этом надо иметь в виду, что информационное обслуживание в наибольшей степени должно соответствовать особенностям восприятия пользователем каждого вида информации, к которой он имеет доступ, и в наибольшей степени способствовать удвоению его информационных запросов, включая механизм ассоциативного мышления.

Эффект такого информационного обслуживания может продублировать интерес к работам других исследователей в данной области, чтобы знать, кто ещё работает в интересующей или близкой по тематике проблеме. Интерес к этому является достаточно мощным побудительным мотивом для знакомства с зарубежной литературой. Это может быть интерес к определённому кругу конкретных исследователей, знакомых пользователю и к результатам, достигнутым ими на определённом этапе, что диктуется часто проблемами научной конкуренции и приоритета.

Стимулом может явиться также и необходимость знать определённое положение, факт или конкретные данные, параметры, лежащие в поле интересов пользователя. Интерес к этому определяется возможностью использовать конкретику, подтвердить или опровергнуть научные выводы пользователя на основании данных или фактов, полученных другими исследователями.

Наконец, это может быть необходимость использования полноразмерного первоисточника как самого полноценного источника информации, с точки зрения его возможностей для рождения ассоциативных мыслей.

Если для проведения тематического мониторинга может быть выделен один специалист, то возникает необходимость отобрать из всей совокупности 250 журналов, публикуемых по дисциплине «Информатика», не более 20–25 первоисточников, с которыми этот специалист будет в состоянии эффективно работать. Это должно быть «ядро» наименований журналов, в наибольшей степени покрывающее тематику института и соответствующее максимальному количеству статей данной тематике.

Выбор ядра связан с законом рассеивания информации, впервые сформулированным Брэдфордом. Особенностью этого закона рассмотрены в работе [4]. Конкретное количество ядра наиболее информативных журналов, отбираемых для тематического мониторинга по проблеме «Информатика», может быть проведено с учётом теории группового выбора [2, 3]. При этом, исходными данными служат: тематический спектр исследователей института, отдельные направления и разделы которого могут быть выражены в индексах полной классификации информатики. Важным является также стоимость доступа к электронным версиям первоисточников на сайтах Интернет, определяющая количество журналов или отдельных статей, которые институт может оплатить, в случае их необходимости.

Поэтапная процедура выбора «ядра» наименований журналов подробно рассмотрена в работе [5]. Предлагаемая методика позволяет ранжировать журналы по степени соответствия содержания публикуемых статей тематике института (от максимального до минимального значения), а также по стоимости доступа к их электронным версиям на сайтах Интернет.

Таким образом, в зависимости от информативности источника, финансовых возможностей института и возможностей специалиста — эксперта обработать при мониторинге текущие публикации журналов, можно выбрать их оптимальное количество. Причём, журналы, подлежащие обработке, можно расположить в убывающей по информативности или важности последовательности. Ниже приведён список журналов по дисциплине «Информатика», отобранный по описанной выше методике для ИПИ РАН.

— Издательство Elsevier/North Holland/Pergamon Journals:

1. Systems and Control Letters.
2. Data & Knowledge Engineering.
3. Information & Software Technology.

4. Information Systems.
 5. Theoretical Computer Science (TCS).
- Издательство IEEE (Institute of Electronic and Electric Engineering):
6. Automatic Control.
 7. Information Theory.
 8. IEEE Signal Processing.
 9. Parallel and Distributed Systems (TPDS).
 10. Knowledge and Data Engineering (TKDE).
 11. Computers.
- Издательство SIAM (Society for Industrial and Applied Mathematics):
12. SIAM journal on applied mathematics.
 13. SIAM journal on control and optimization.
 14. SIAM journal on optimization.
 15. SIAM review.
 16. SIAM journal on mathematical analysis.
- Издательство Springer:
17. Probability Theory and Related Fields.
- Другие издательства:
18. IBM Systems Journal.
 19. Science.
 20. Nature Magazine.

В качестве эталона для сравнения тематики института и тематики журналов предлагается использовать классификацию отработанную в рубрикаторе Computer and Information Sciences, изданном Международной ассоциацией ACM (Association of Computer Machinery), которой пользуются многие зарубежные издатели. В России дисциплина «Информатика» примерно соответствует тому, что в западной литературе в совокупности понимается как «компьютерная и информационная наука» или Computer and Information Sciences.

Полный английский оригинал рубрикатора ACM и соответствующий ему русский перевод приведён в (5). Это самая полная современная классификация информатики в виде 4-х уровней рубрикатора. Достоинством этого рубрикатора является то, что многие зарубежные издатели научно-технических журналов по информатике просят авторов статей при направлении их в издательство индексировать статьи индексами рубрикатора ACM.

В предлагаемом варианте мониторинга они могут использоваться не только для классификации или/и индексирования тематики института, но и для индексирования статей, отображенных в результате мониторинга, и их последующего хранения в базе данных. Это позволяет при поиске интересующей статьи, который пользователь осуществляет в базе данных внутренней, корпоративной сети Интернет, пользоваться индексами АСМ, как одним из информационно-поисковых языков, поскольку он сопряжен с тематикой института, и с тематикой журнала.

Эффективность системы мониторинга Интернет-информации зависит от организации доступа пользователей к результатам этой работы. Любой исследователь института, как пользователь, должен получить текущий информационный бюллетень, как некую экспресс-информацию, в рамках установленной периодичности её выпуска, непосредственно на свой электронный адрес с библиографическими данными и аннотациями статей. Эта информация является базой, на основе которой пользователь может решить, нужен ли ему сам первоисточник или нет. Таким образом, знакомство с «информацией об информации» и доступ к самому первоисточнику должны быть неразрывным процессом информационного обслуживания.

Результаты мониторинга в виде электронных информационных бюллетеней должны аккумулироваться, формируя ретроспективную базу данных, доступную пользователю в любое время. Для того, чтобы обеспечить выполнение функций доступа к этой базе, можно воспользоваться внутренней корпоративной сетью института.

Текущие информационные бюллетени и их ретроспективная коллекция, как база данных, позволяют пользователю заказать полноформатную электронную версию или сам первоисточник «в живую». Таким образом, мониторинг в известной мере решает организационную и экономическую проблемы информационной поддержки научных исследований.

Организация мониторинга в предлагаемом виде позволяет оплачивать доступ не ко всему выпуску журнала, а только к тем его статьям, которые заказаны пользователем. Это ведёт к значительному сокращению расходов на доступ к электронным версиям первоисточников и, тем самым, к экономии средств на подписку.

Стержнем ретроспективной базы данных является её поисковый аппарат. Он должен быть достаточно простым и удоб-

ным для пользователя. Элементом поискового аппарата является классификационная схема информатики или, другими словами, рубрикатор. Индексы его рубрик могут использоваться в качестве одного из информационно-поисковых языков, на котором пользователь может общаться с системой.

Для поиска информации в базе нужно предусмотреть возможность использования и других языков. Например, элементов библиографического описания: названий журналов, дат и номеров их выпусков, авторов статей и терминов из аннотации (ключевых слов и их сочетаний). Информационные технологии поиска информации в базе данных по названному выше элементу хорошо известны, многократно описаны и широко используются.

3. Заключение

Предложена методика использования Интернет-ресурсов на основе тематического мониторинга «ядра» зарубежной периодики. С организационной и финансовой точек зрения, эта методика доступна для практической реализации в любом институте Российской академии наук. Результаты мониторинга позволяют оперативно информировать исследователей о последних публикациях их коллег за рубежом.

Создаваемая на основе мониторинга ретроспективная база данных может рассматриваться как уникальное, тематически направленное собрание информации, в максимальной степени соответствующее тематике исследований, проводимых в институте. База данных, составленная на основе рассмотренных принципов, содержит минимальный «информационный шум», с точки зрения тематического спектра, соответствующего научным интересам специалистов института.

Предложенная модель использования англоязычных информационных ресурсов Интернет позволяет максимально сократить расходы на доступ к научной информации, предоставляемой на платной основе. При этом, работая с оптимальным количеством журналов, в наибольшей степени соответствующих тематике института, можно адресно удовлетворять информационные потребности пользователей.

Расходы на проведение тематического мониторинга Интернет — информации и ведение ретроспективной базы данных окунаются вполне оправданными, если пользователи будут в полной мере использовать эту информацию.

Список литературы

1. Гаврилов А.В. Модель ассоциативного мышления. — <http://emark.cs.nstu.ru>.
2. Ларичев О.И. Теория и методы принятия решений, — М.: Наука, 2000.
3. Миркин Б.Г. Проблема группового выбора. — М.: Наука, 1974.
4. Михайлов А.И., Чёрный А.И., Гиляревский Р.С. Основы информатики. 2-е изд. — М.: Наука, 1968. — С. 92–95.
5. Принципы построения интегральной информационной сети для информационной поддержки НИР в институте РАН. Отчёт. Фонды ИПИ РАН. — М., 2005.
6. Computer Science Journals. — <http://www.informatik.unu-trier.de/~ley/db/journals/index-old.html>.