

Поисковые системы в удовлетворении отраслевых информационных потребностей (На примере агропромышленного комплекса)

Высокие темпы появления разнообразных по содержанию и формам представления информационных ресурсов и внедрение электронных информационно-коммуникационных технологий меняют классические представления об удовлетворении информационных потребностей специалистов. В области агропромышленного комплекса под воздействием таких перемен существенно изменилась практика удовлетворения информационных потребностей специалистов. Процесс перехода большинства потребителей к Интернет-ресурсам как основному источнику информации, необходимой для полноценного функционирования в профессиональной среде, в области сельского хозяйства, как и в других отраслях, происходит стихийно и чрезвычайно активно. Вследствие этого особую остроту приобретают средства ориентации в Интернет-пространстве, вызвавшие к жизни целую отрасль, включающую специальные поисковые системы и сервисы: справочники (directories) и поисковые системы (search engines).

В практике обслуживания информационных потребностей в области сельского хозяйства, по оценкам специалистов, в разной степени используется около пятисот российских и зарубежных поисковых систем, некоторые из которых имеют мировой охват, другие ограничиваются региональными рамками или нацелены на отражение конкретной отраслевой тематики и оказание определенных видов информационных услуг. Одни поисковые системы используют поисковые машины, которые работают по принципу ответа экспертов на вопросы типа: «кто был...» или «что такое...», - базирующихся на анализе статистики пользовательских запросов. Для многих поисковых задач такие поисковые машины очень удобны, но в русском секторе интернета таких машин пока нет, зато в англоязычном секторе имеются подобные машины, способные воспринимать вопросы на русском языке. Другие поисковые системы реализуют навигацию на основе специальных указателей, представляющих собой тематические «деревья», построенные на основе классификаций. Од-

нако регистрация веб-сайтов и веб-страниц в таких системах осуществляется вручную, поэтому объем базы данных систем классификационного типа невелик по сравнению с информационной емкостью всего интернета. Поиск можно вести по конкретному адресу или по так называемым справочным системам (каталогам или рубрикорам), имеющим четкую иерархическую систематическую или логико-тематическую структуру как в систематическом каталоге библиотеки. Работа со справочниками позволяет ориентироваться в Интернет-ресурсах в пределах одной отрасли знания от общего к частному, меняя иерархические ветви. К таким поисковым системам относятся справочники всемирного масштаба: Yahoo! (<http://www.yahoo.com>), Open Directory (<http://dmoz.org>), About (<http://www.about.com>), а также российские: Aport (<http://www.apor.ru>), List.ru (<http://list.mail.ru>), Weblist (<http://www.weblist.ru>) и др.

Итак, состав поисковых средств Интернет-пространства чрезвычайно разнообразен, что ставит перед любым пользователем закономерный вопрос выбора. Статистика виртуального пространства показывает, что в практике на протяжении пяти последних лет из всего множества поисковых систем выделяется небольшая группа, к услугам которой обращается основная доля российских пользователей. Разные источники относят к этой группе четыре или пять систем, в числе которых Google (<http://www.google.com>), безоговорочный мировой лидер; Aport (<http://www.apor.ru>); Search mail.ru (<http://search.mail.ru>); Rambler (<http://www.rambler.ru>); Yandex (<http://www.yandex.ru>). Кроме того, в России имеется поисковая система, которая осуществляет мета-поиск, то есть дает пользователю возможность искать ресурс одновременно по трем ведущим поисковым системам - Yandex, Rambler и Aport. Возможно подключение к поиску еще двух ведущих систем (адрес: <http://search.freeland.ru>).

Сравнение этих систем по эффективности стандартного поиска в области сельского хозяйства показывает следующее. На Яндексе фактическая релевантность первых ссылок может быть и высокой, и крайне низкой. Так, по запросу «Вступление России в ФАО» системой были выданы две копии одной и той же статьи на серверах «agronews.ru» и «gazeta.ru», страницы «ссылки» с разных сайтов. Еще хуже показатели у Рамблера: система выдает большое количество ссылок, которые никак не упорядочены. По умолчанию они сортируются по релевантности, то есть по степени соответствия запросу. Однако очень часто первые страницы дают результаты, весьма далекие от требуемых. Например, запрос «Вступление России в ФАО» был обслужен следующим образом: первые пять ссылок - страницы новостей с сайтов «fishers.ru», «agronews.ru», «lenta.ru», «product.ru», затем ссылки, вообще не имеющие отношения к теме, например, под номером 6 шел сайт «windowslive.ru». Если в случае с Яндексом использование синонимов, а также обращение к каталож-

Сравнение этих систем по эффективности стандартного поиска в области сельского хозяйства показывает следующее. На Яндексе фактическая релевантность первых ссылок может быть и высокой, и крайне низкой. Так, по запросу «Вступление России в ФАО» системой были выданы две копии одной и той же статьи на серверах «agronews.ru» и «gazeta.ru», страницы «ссылки» с разных сайтов. Еще хуже показатели у Рамблера: система выдает большое количество ссылок, которые никак не упорядочены. По умолчанию они сортируются по релевантности, то есть по степени соответствия запросу. Однако очень часто первые страницы дают результаты, весьма далекие от требуемых. Например, запрос «Вступление России в ФАО» был обслужен следующим образом: первые пять ссылок - страницы новостей с сайтов «fishers.ru», «agronews.ru», «lenta.ru», «product.ru», затем ссылки, вообще не имеющие отношения к теме, например, под номером 6 шел сайт «windowslive.ru». Если в случае с Яндексом использование синонимов, а также обращение к каталож-

СЛАДКОВА Ольга Борисовна - доктор культурологии,

зав. кафедрой прикладной информатики Московского государственного университета культуры и искусств.

Адрес: 141406, Московская обл., г. Химки, ул. Библиотечная, 7

e-mail: olgasladkova@mail.ru

ПИРУМОВА Лидия Николаевна - кандидат педагогических наук,

зам. директора Центральной научной сельскохозяйственной библиотеки Российской академии сельскохозяйственных наук (ЦНСХБ Россельхозакадемии).

Адрес: 107139, ГСП, г. Москва, Орликов пер., 3, корп. «В»

e-mail: pln@cnsnb.ru

ПИРУМОВ Андрей Александрович - кандидат педагогических наук,

специалист по оборудованию ООО «ДжиИ Хэлскеа».

Адрес: 123317, г. Москва, Пресненская набережная, 10, корп. «С», 12 этаж

e-mail: punya061@gmail.com

ной части системы и функция «искать в найденном» позволяют найти желаемое, то в Рамблере объем базы данных очень велик, а возможности по структурированию информации недостаточны. Расширенный поиск Яндекса - чрезвычайно мощная система, одним из основных преимуществ которой является наличие целого арсенала средств для разбора структуры текста: можно указывать расстояние между словами и их порядок, акцентировать отдельные слова запроса, указывать уточняющие слова, не участвующие в поиске, но меняющие представление Яндекса о релевантности того или иного документа. Также можно указать слова, которые не должны присутствовать в искомом документе, ограничивать поиск только определенными разделами документа, например, заголовками или подписями к картинкам и т. д. Сама форма расширенного поиска позволяет определить словарные фильтры, указать дату модификации документа, ограничить поиск только конкретным сайтом, искать в документах картинки или специальные объекты. Кроме того, удобной возможностью для веб-мастеров и владельцев ресурсов является поиск сайтов, которые ссылаются на указанный. Это роднит систему со знаменитой концепцией PageRank системы Google.

Расширенный поиск Рамблера имеет свои неоспоримые преимущества: широкие возможности по заданию условий поиска, таких как: выбор раздела документа, в котором нужно проводить поиск, выбор языка документа; поиск документа в различных словоформах; поиск документов с определенной датой модификации; исключение из результатов документов, содержащих определенные слова. Кроме того, использование специальных символов в тексте запроса позволяет искать слова без окончаний или слова с пропущенными буквами. Это полезно, если потребитель не уверен в написании термина.

Сравнивая результативность поиска с помощью поисковых систем Rambler, Google, Yandex, Aport и др. на примере запросов сельскохозяйственной тематики, следует отметить: совпадение результатов поиска составляет всего 10%, что заставляет всесторонне учитывать возможности различных поисковых машин в выборе стратегии поиска для решения эвристических задач.

Сложность поиска информации по проблемам АПК в электронных сетях объясняется огромными объемами ресурса, представленного в интернете, рассеянием нужных сведений по сайтам, порталам и т.д. Однако можно попытаться собрать информацию о наиболее ценных в научном и практическом плане сайтах, охарактеризовав состав и качество представленных на них материалов.

Опыт показывает, что часто кажущаяся полнота найденной в интернете информации не соответствует действительности, а сама информация нередко вызывает сомнения в достоверности. В повседневной практике сельскохозяйственные библиотеки активно используют сетевые ресурсы, размещенные как во внутрибиблиотечных сетях (интранет), так и вне библиотеки в интернете. Однако в силу того, что ни одна из рассмотренных поисковых систем не обладает абсолютными преимуществами, представляется целесообразным библиотекам взять на себя функцию анализа и отбора наиболее авторитетных, полных, достоверных и надежных Интернет-ресурсов. Это оградит ученых и практиков АПК от информационного шума, с которым те неизбежно сталкиваются, осуществляя поиск в Интернет-ресурсах с помощью различных поисковых систем.

Вместе с тем, исследователи отмечают, что качественная селективная архивация сетевых ресурсов не под силу одной организации. Отражение архивированных сетевых ресурсов в электронных каталогах библиотеки имеет некоторые особенности. В качестве адресов архивированных ресурсов приводится адрес ресурсов на сервере его производителя и в библиотеке. Если ресурс не поддерживается производителем, то его просмотр обеспечивается ссылкой на библиотечный ресурс. В интересах библиотеки развивать систему ссылок на внешние ресурсы. Ссылки собираются в соответствии с ее профилем. Накопление их в большом количестве ставит вопрос об эффективности их организации. Указатели, списки, созданные каждой библиотекой, могут быть со временем объединены и соединены ссылками друг на друга. Крупные библиотеки, как, например, ЦНСХБ Россельхозакадемии, проводят анализ Интернет-ресурсов, составляют списки «Полезных ссылок». В этой ситуации библиотеки вынуждены активизировать ссылки, поскольку, как уже говорилось, многие сайты

либо не обновляются, либо содержат информацию устаревшую, либо прекращают свое существование, и тогда ссылка не работает. Проблема решается по-разному: возможна ручная проверка или регулярный запуск программ-роботов, обходящих сайты по ссылкам и проверяющих, существуют ли они по указанным адресам (новое направление на основе применения стандарта Open URL). Анализ Интернет-ресурсов по глубине и разнообразию содержащейся на них информации, развитости их структуры, количеству и качеству предоставляемых продуктов и услуг пользовательского назначения необходим, так как в отличие от интернета, который не несет ответственности за достоверность информации, библиотека такую ответственность несет и обязана представлять пользователю качественную информацию, потому что она включает Интернет-ресурсы в свои информационные продукты, в том числе в электронные каталоги, справочники и т. д.

Отраслевое информационное пространство формируется децентрализованно, различными учреждениями, деятельность которых не координируется. Такая ситуация не дает возможности оценить реальные размеры, структуру, тематическую наполняемость и информационную ценность его компонентов. Поэтому библиотеки прибегают к созданию собственных путеводителей. Задача создания такого путеводителя на протяжении последних двадцати лет дискутировалась в профессиональной среде, и опыт создания подобных справочников имеется как в нашей стране, так и за рубежом.

Функциональное назначение такого навигатора прежде всего в том, что он используется как средство ориентации в сетевом информационном пространстве. Конечно, навигатор в прямом смысле не дает возможность координировать деятельность тех, кто создает отраслевые информационные ресурсы, но он позволяет пользователю избежать обращения к источникам, дублирующим друг друга. Отсеивает неинформативные объекты, обращает внимание пользователя на самые ценные информационные продукты, таким образом выполняя роль справочного пособия в виртуальной информационной среде.

Кроме того, навигатор позволяет устанавливать структурные связи в Интернет-пространстве, поскольку

с его помощью можно выявлять создателей и владельцев сетевых ресурсов, определять отрасли, недостаточно обеспеченные сетевыми ресурсами, констатировать преобладание одних и нехватку других информационных продуктов и услуг для информационной поддержки аграрной науки, обозначать страны-лидеры в создании отраслевых Интернет-ресурсов, наиболее заметные научные школы и языки, на которых представлены ресурсы. Это позволяет расценивать навигатор как системообразующий инструмент, который в определенном смысле может выполнять функции координации в деятельности производителей сетевой информации.

У ЦНСХБ Россельхозакадемии имеется опыт создания подобных путеводителей: с начала 2000-х годов ведется систематическое наблюдение за отраслевыми сетевыми информационными ресурсами, и в 2009 году был создан навигатор по сельскохозяйственным сайтам (<http://www.cnsheb.ru/AKDiL/navig>).

Поэтому мы можем поделиться некоторыми рекомендациями по его созданию. Во-первых, перед создателями стоит проблема изменчивости и недолговечности Интернет-ресурсов. По разным источникам, период жизни Интернет-ресурсов колеблется от 2 до 4-х месяцев, а затем либо они исчезают вообще, либо существуют и обновляются через 1-2 года. Считается, что существование сайта в течение 6 недель (1,5 месяца) в «живом» состоянии дает основание называть его относительно стабильным. Однако опыт ЦНСХБ Россельхозакадемии доказывает, что для проверки надежности сайта «испытательный» срок его жизни должен быть увеличен, как минимум, до 6 месяцев. Анализ агровеб-сайтов показывает, что вновь возникшие сайты часто перестают развиваться, замирают, не обновляются. Вряд ли следует такой ресурс рекомендовать пользователям.

Другая проблема - неинформативность сайтов. Среди них встречается множество рекламных, обращение к которым только отнимает время ученых и не дает полезной информации. Можно потратить время, обращаясь к сайту, название которого подходит по искомой теме, но содержание не оправдывает ожидания. С другой стороны, достойные в содержательном смысле сайты носят порой неинформативные названия и т. д. Поэтому создание навигатора связано с критериальным подходом отбора сайтов. Критерии неизбежно включают как содержательные, так и формальные характеристики.

Обращает на себя внимание и проблема определения принадлежности сайта. Адрес веб-сервера в пространстве IP-адресов, как правило, служит указателем на географическое размещение сайта, но бывает, что русскоязычные сайты размещены на серверах другой страны (например, Великобритании, США, Японии), поскольку это не запрещено правилами. Таким образом, адрес сайта может ввести в заблуждение пользователя. Однако кодировка, в которой веб-сервер предлагает документы, является надежным указателем, позволяющим прояснить языковую и географическую принадлежность документа.

Очевидно, что в условиях интернета можно говорить только об относительной полноте любого навигатора. Однако если поставить цель собрать наиболее информативные сайты, то она практически достижима.

В удовлетворении информационных потребностей отечественных специалистов в области АПК приоритетными являются национальные сетевые ресурсы, соответственно они занимают значительное место в навигаторе. Однако реальная ситуация, сложившаяся в Интернет-пространстве, показывает, что большую часть в нем занимают ресурсы, составленные международными организациями, а также информационными и исследо-

вательскими отраслевыми учреждениями развитых стран, поэтому их информационный сетевой продукт также необходимо отражать в навигаторе.

Для удобства пользователя навигатор должен быть структурирован в соответствии с логикой развития данной области знания, для чего можно использовать устоявшиеся средства формализованной лексики: рубрики и классификаторы.

Любой список Интернет-ресурсов быстро устаревает: каждый день появляются новые сайты с новыми данными, а старые также быстро исчезают. Следовательно, навигатор должен периодически обновляться, а значит, необходимо проводить постоянный мониторинг отраслевых сетевых ресурсов. Это подтверждается опытом работы с навигатором в ЦНСХБ. В результате мониторинга сетевых ресурсов по проблематике АПК в 2010 г. было выявлено 65 новых сайтов. Кроме того, было установлено, что 18 сайтов, присутствующих в навигаторе, изменили адреса, 20 сайтов оказались недоступны. Все изменения и добавления внесены в навигатор.

Итак, при использовании множества поисковых систем для выявления отраслевой информации пользователь затрачивает больше времени, но не может быть уверен в точности и полноте найденной информации, чем при профессиональном поиске, который могут предложить ему библиотеки, используя специально созданные путеводители-навигаторы по сетевым информационным ресурсам данной отрасли. Навигатор ориентирован на реальные запросы пользователей данной отрасли, содержит достоверную и полную информацию о ресурсе, постоянно актуализируется. Опыт ЦНСХБ Россельхозакадемии убеждает в эффективности использования навигатора по сельскохозяйственным сайтам для удовлетворения информационных потребностей, формирующихся в данной отрасли.

Литература:

1. Берестова Т.Ф. Поисковые инструменты библиотеки / Т.Ф. Берестова // Библиогр. - 2006. - № 6. - С. 18-30.
2. Галеева И.С. Справочник «Информационные ресурсы интернета»: опыт создания электронного путеводителя / И.С. Галеева // Библиотечные и информационные ресурсы в современном мире науки, культуры, образования и бизне-

са: материалы 14-й междунар. конф. «Крым-2007», 9-17 июня 2007. - Су-дак, 2007. - С. 267-269.

3. Литвинова Н.Н. Организация интегрированной среды пользователя для поиска в сетевых удаленных ресурсах / Н.Н. Литвинова // Науч. и техн. б-ки. - 2008. - № 6. - С. 5-13.

4. Навигатор по удаленным сетевым ресурсам по проблематике АПК / Центр. науч. с.х. б-ка Рос. акад. с.х. наук / Сост. С.А. Дубров-

ская, А.А. Пирумов, Л.А. Свирская. - М.: ЦНСХБ Россельхозакадемии, 2009. - 176 с.

5. Пирумов А.А. Электронные источники в системе информационных ресурсов по сельскому хозяйству / А.А. Пирумов // Вестник МГУКИ. - 2009. - № 5. - С. 201-205.

6. Шрайберг Я.Л. Роль библиотек в обеспечении доступа к информации и знаниям в информационном веке / Я.Л. Шрайберг // Науч. и техн. б-ки. - 2008. - № 1. - С. 7-44.