

## Поиск информации о диабете в Интернете

А.В. Древаль

Московский областной научно-исследовательский  
клинический институт им. М.Ф. Владимирского МЗ РФ

**М**ожно выделить 3 основных способа поиска медицинской информации в Интернете: по ключевому слову (например, «диабет») или фразе («диабетическая стопа»), которые вводятся в так называемую Поисковую Машину (например, представленные по адресу ... или ... ) и она проводит поиск информации в Интернете, содержащей указанное слово или фразу; в предметных каталогах Поисковой Машины, например, в разделах «Эндокринология», «Диабетология» и т.п.; в предметных каталогах, имеющих оценку качества информации, так называемые оцененные предметные каталоги.

### Поиск по ключевому слову

Несмотря на предубеждение среди медиков относительно сложности работы в Интернете, начинающего пользователя обычно удивляет простота поиска нужной информации, осуществляемой с помощью Поисковых Машин. Наиболее популярны в российском Интернете Поисковые Машины, и вход в сеть Интернет осуществляется обычно или через Microsoft Internet Explorer или Netscape Navigator, которые можно отнести к индивидуальным (пользовательским) Поисковым Машинам. В каждой из них имеется кнопка «Поиск» (Net Search), нажав на которую можно выйти на зарубежные Поисковые Машины (например, взаимодействие пользователя с любой Поисковой Машинной в Интернете происходит однотипно. В специальную рамку (бокс) Поисковой Машины вводится ключевое слово или фраза, а затем нажимается кнопка «Поиск» (Search). В течение нескольких секунд появляется перечень сайтов, в которых обсуждаются, например, проблемы диабета, если было введено слово «диабет»).

Такой удобный способ поиска информации в Интернете стал возможен благодаря разработке компьютерных программ под общим названием «Роботы». Задача такой программы состоит в том, чтобы регулярно собирать в Интернете информацию о сайтах и помещать ее в свою базу данных, с которой и работает пользователь, когда обращается к услугам Поисковой Машины. Когда пользователь Интернета вводит в Поисковую Машину (Rambler, например)

ключевое слово, то из базы данных, сформированной роботом, ему предоставляется перечень сайтов, в которых имеется ключевое слово. В настоящее время в Интернете представлено несколько десятков популярных роботов. Но, к сожалению, их обилие создает определенные проблемы для пользователя, поскольку принцип работы каждого робота уникален и поэтому одно и то же ключевое слово (например, «диабет»), введенное в разные Поисковые Машины (Rambler, Yandex и т. п.), базы данных которых сформированы разными роботами, дает различный перечень сайтов как по порядку их представления, так и по охвату. Можно выделить 2 основные причины таких отличий в работе роботов: одни из них собирают все, что они в состоянии найти в сети по заданной теме, а другие выбирают из сайтов только определенные страницы по заданной теме; одни из них успешно идентифицируют такие сложные объекты для поиска как рисунки карт, фреймовые ссылки и страницы, доступ к которым возможен только с паролем, а другие роботы такими возможностями не обладают.

Следует также отметить, что независимо от того, как велика база данных в Поисковой Машине, она не в состоянии детально и автоматически классифицировать всю информацию, представленную в Интернете. Например, ввод ключевого слова «MEDLINE» в поисковую машину Rambler предоставляет пользователю до 18000 сайтов, через которые можно войти в эту базу данных, но ни одна Поисковая Машина, в том числе и Rambler, не классифицирует информацию, расположенную в базе данных MEDLINE. Таким образом, для врача, который ищет определенную информацию в Интернете, остается довольно широкий диапазон для творческого подхода к этому процессу, который и определяет результативность поиска.

Рассмотрим практическую сторону вопроса на примере некоторых популярных отечественных и зарубежных Поисковых Машин.

**Yandex (Яндекс)** является наилучшей Поисковой Машинной в российском Интернете. Эта база данных содержит около 200 000 серверов и до 30 млн. документов, которые Поисковая Машина просматривает в течение нескольких секунд.

Поиск информации задается введением ключевого слова («диабет») в специальную рамку и нажатием кнопки «Найти» справа от рамки. Результаты поиска появляются в течение нескольких секунд, причем ранжированные по значимости — наиболее важные документы размещаются в начале списка. Ранг найденного документа определяется тем, в каком месте документа находится ключевое слово (в заглавии важнее, чем в другом месте) и числом упоминаний ключевого слова (чем больше упоминаний, тем ранг выше). Сайты, расположенные на первых местах в списке, являются ведущими не с содержательной точки зрения, а по отношению к частоте упоминания ключевого слова. Не следует ограничиваться просмотром первого десятка предложенных Поисковой Машиной сайтов. Содержательную часть сайта можно косвенно оценить по краткому его описанию, размещаемому Поисковой Машиной под адресом сайта.

Поиск сайтов по ключевому слову «диабет» даст список, в начале которого будут представлены Web-страницы (сайты), в заголовках которых содержится эта фраза. Если в найденных Поисковой Машиной документах эта фраза будет содержаться более одного раза, то машина ранжирует остальные документы по частоте встречаемости заданной фразы. Некоторые авторы сайтов для повышения вероятности появления своей Web-страницы на первых местах Поисковой Машины, умышленно включают в документ бессмысленные повторы ключевого слова. Но как только Поисковая Машина обнаруживает такой «замусоренный» документ, она автоматически исключает его из своей базы данных.

Даже ранжированный список документов, предлагаемый Поисковой Машиной в ответ на ключевую фразу или слово, может оказаться практически необозримым. В связи с этим в Яндекс (как и других Поисковых Машинах) предоставлена возможность в рамках первого списка выбрать документы, которые точнее отражают цель поиска. Например, на ключевое слово «диабет» Яндекс предлагает список из 34 899 Web-страниц, где первый российский сайт по диабету занимает первое место. После ввода в команду «Искать в найденном» ключевого слова «нефропатия» список сокращается до 750 страниц, а после ввода в эту команду слова «микроальбуминурия» этот список сокращается до 130 Web-страниц.

Под каждым сайтом указана команда — «Похожие документы», и если ее запустить, Поисковая Машина предложит перечень сайтов с близким по теме содержанием.

**Alta Vista** является лучшим поисковым инструментом в Интернете. Эта база данных содержит более 100 млн. Web-страниц, которые Поисковая Машина

может просматривать в течение нескольких секунд.

Поиск информации задается введением ключевого слова («diabet») в специальную рамку и нажатием кнопки «Search» («Поиск») справа от рамки. Результаты поиска появляются в течение нескольких секунд, наиболее важные документы размещаются в начале списка. Alta Vista также позволяет ограничить результаты поиска определенным языком, а через сервис «Advanced» (дополнительно) найти страницы, которые созданы в определенные промежутки (годы, месяцы и т.п.). Например, на ввод ключевого слова «diabetes» с указанием Russian, Alta Vista представляет 103 русскоязычных Web-страницы, в котором российский специализированный по диабету сайт занимает 4-е место, а на 1-м, — где слово «диабет» входит в рекламу экзотического респиратора для лечения диабета. Специальная синтаксическая поисковая функция Alta Vista обеспечивает не только поиск изображений, но и информацию по Web-страницам со сходной информацией («Related pages»).

**Northen Light.** Эта Поисковая Машина содержит в своей базе данных около 315 млн. Web-страниц, а также собственную специализированную коллекцию полных текстов журналов и газетных статей. Northen Light обеспечивает врачей обширной, а порой и уникальной информацией.

Поиск задаваемой пользователем тематической информации осуществляется как в Интернете, так и в специальной коллекции этой Поисковой Машины. Коллекция состоит из 25 млн. отобранных источников, которые не являются общедоступными в Интернете. Что касается диабета, то эта коллекция формируется на основе журналов Diabetes, Diabetes Care, Lancet, British Medical Journal и др. К сожалению, российские источники не представлены. Другим ограничивающим использованием этой специальной коллекции обстоятельством является платность услуги: чтобы получить статью, следует заплатить от 1 до 4\$ США. Но если окажется, что оплаченная статья не подходит, то можно за эту же сумму заказать новую статью. Поисковая Машина Northen Light сортирует найденную информацию по отдельным категориям, что упрощает поиск нужного материала. В специальной коллекции по диабету выделяются категории «Диабет», «Инсулин», «Болезни сердца», «Генетика» и др. Внутри начальных категорий выделяются подкатегории, что позволяет существенно сузить и ускорить поиск нужной информации. Web-страницы и сайты, не включенные в специальную коллекцию, кроме того что ранжированы в общем списке, также разделены на категории. В этих категориях выделен раздел Russian Site, а первый отечественный сайт выделен в отдельную категорию с перечнем ряда его страниц.

**Достоинства поиска по ключевому слову (фразе).** По ключевому слову поиск организуется узконаправленно, точно отражая потребность в информации пользователя; обширный перечень найденных документов гарантирует в определенной степени полноту охвата информации; легко организовывать поиск по базе данных.

**Недостатки поиска по ключевому слову (фразе):** неоднозначность выбранных для поиска слов и использование одних и тех же слов в различных предметных областях приводит к тому, что в список включается много документов, не имеющих прямого отношения к запрашиваемой информации (например, на фразу «инсулиновая ручка» в список документов, представленных Поисковой Машиной, могут войти не только инсулиновые, но и другие ручки); ресурсы (Web-страницы, сайты), которые включаются в базу данных Поисковой Машины роботом, специально не оцениваются. Для адекватного отражения сайта в Поисковой Машине детальное ее описание нужно самостоятельно разместить в поисковые сайты Яндекс, Alta Vista, Northern Light или в любом другом.

**Инструменты мета-поиска.** В настоящее время в Интернете представлены серверы, которые позволяют осуществлять одновременный поиск по нескольким базам данных. Поисковая машина Dogpile обеспечивает поиск по 14 базам данных Интернета, представленным в одном экране. Эти базы данных включают Alta Vista, Yahoo, Lycos и др. Сторонники такого мета-поиска полагают, что использование большого перечня Поисковых Машин существенно улучшает информативность. Теневой стороной этого утверждения является то, что поиск информации осуществляется не по объединенной базе данных всех Поисковых Машин. В итоге пользователь нередко набирает, с одной стороны, практически необозримое число Web-страниц, а, с другой, в каждом из списков, которые представляет отдельная Поисковая Машина, оказывается много документов, представляемых другими Поисковыми Машинами.

В связи с вышесказанным наиболее целесообразно использовать этот подход к поиску лишь в тех случаях, когда необходимо найти какую-то редкую и необычную информацию и поэтому удобнее организовывать поиск с какой-то одной Web-страницы в Интернете, где представлен набор ведущих Поисковых Машин, чем разыскивать отдельные Поисковые Машины в сети.

## Предметные каталоги

Поиск информации по ключевому слову (фразе) в Интернете аналогичен поиску информации о ка-

ком-то слове или фразе в книге, когда для этого используется так называемый «Предметный указатель», где против слова или фразы указана страница текста, на которой можно найти его разъяснение. Если нужно получить общее представление о том, чему посвящена книга, следует обратиться к ее содержанию. В Интернете можно искать медицинскую информацию, пользуясь в Поисковых Машинах выделенными тематическими разделами (предметными каталогами). Например, если необходимо найти информацию о сахарном диабете, следует обратиться к разделу «Медицина», в нем выбрать раздел «Эндокринология», а в последнем — раздел «Сахарный диабет», где должны быть сгруппированы материалы по этой тематике.

Информация по медицине и здравоохранению наилучшим образом представлена в Поисковой Машине Yahoo, а также на сайте World Wide Web Virtual Library.

**Yahoo.** Стартовая точка сбора медицинской информации в Интернете для начинающих. База данных этой Поисковой Машины разделена на тематические разделы, один из которых (Health, Поисковая Машина Yahoo!) в ответ на запрос представляет новое «меню» для выбора в виде более детальной рубрикации раздела Health, включающего Diseases and Conditions (болезни и состояния). Выбрав Diseases and Conditions, вы получаете новое «меню», из которого можно вызвать раздел Diabetes, если нажать букву D в алфавите страницы. В разделе Diabetes выделены категории (книги, компании, инсулинорезистентность), наиболее популярные сайты и перечень Web-страниц по диабету. Поскольку Yahoo! удерживает в процессе поиска иерархическую структуру разделов, всегда можно вернуться на предыдущий вышестоящий уровень, если по какой-то причине вызванный текущий раздел не оправдал ожиданий.

**Virtual Library** — старейший каталог ресурсов сети. В отличие от других сайтов такого типа, он содержит обстоятельный каталог, с помощью которого можно выйти на раздел по своей специальности. Предметный каталог Virtual Library дает представление о том, что опубликовано в Интернете и какой материал в настоящее время доступен. Если в каталоге представлен раздел, соответствующий направлению поиска, тогда можно отыскать такие работы, которые в других поисковых системах практически недоступны.

**Преимущества предметных каталогов.** Поскольку предметный каталог позволяет выделять из огромных ресурсов сети нужную информацию, то исчезает необходимость задавать в Поисковой Машине специфические для поиска ключевые слова (тем самым облегчается поиск). Организация информации в

виде каталога является логически очевидной и удобной для начинающего пользователя Интернета.

**Недостатки предметных каталогов.** Поскольку каталог формируется не автоматически, а авторским коллективом, то перечень сайтов и Web-страниц оказывается менее полным, а материалы менее свежими, чем в аналогичных разделах автоматизированных Поисковых Машин, работающих по ключевым словам или фразам. Важные ресурсы могут быть упущены вследствие размещения ресурса в неожиданном разделе каталога.

### Оцененные предметные каталоги

Необходимую медицинскую информацию можно найти и по третьему пути, если воспользоваться растущим числом каталогов, где дается оценка качества представленной в Интернете информации. Это «оцененные предметные каталоги». Такой каталог специально формируется для профессионалов службы здравоохранения, которым необходимы только надежные и авторизованные источники информации.

**Medikal Matrix.** Одной из наиболее авторитетных медицинских поисковых служб является Medical Matrix Project, задачей которой являются размещение, аннотация и постоянное обновление клинических медицинских Интернет-ресурсов без ограничения к ним доступа. Материалы для каталога собираются рабочей Интернет-группой от Американской медицинской ассоциации по информатике. В настоящее время Matrix содержит более 4000 оцененных по качеству ссылок на сайты, которые размещены в иерархически организованном предметном каталоге. Семь начальных разделов каталога включают направления «Специальность», «Болезни», «Клиническая практика» и «Образование», «Литература», «Здравоохранение», «Компьютеры, Интернет и Технологии». Внутри каждого раздела ресурсы разделены на подразделы, что позволяет пользователю сужать направление поиска до нужной темы. Например, в разделе «Эндокринология» выделены подразделы «Новости», «Инструменты принятия решений», «Справочники и Практические руководства». Представленные ссылки позволяют выйти на первоисточник информации в Интернете, каждому ресурсу дается детальная оценка относительно типа содержащейся в нем информации. Сайты, включенные в Matrix, отвечают определенным основным критериям качества, сайты, которые в разделе явля-

ются наилучшими или лучшими по данной специальности, отмечены звездочками качества.

**OMNI (Organising Medical Networked Information)** Данные, которые можно получить на OMNI, предназначены профессионалам-медикам. Этот сайт анонсирует себя как высококачественный биомедицинский информационный ресурс Великобритании. Особенностью этого сайта является возможность поиска информации в базе данных по трем направлениям: а) по темам, сгруппированным в алфавитном порядке; б) по темам каталога; в) с использованием метода медицинских тематических заголовков. Последний прием автоматически связывает искомую информацию с узкими и широкими медицинскими темами. Когда происходит поиск информации по теме «Нарушения аппетита» (Eating disorders), то предоставляется возможность на сайте получить информацию сразу по двум направлениям — более узкому (Anorexia nervosa) и более широкому (Mental disorders), а также по смежной тематике.

**Health on the Net.** Организация «Здоровье в Сети» (The Health on Net Foundation) — некоммерческая организация, которая занимается реализацией преимуществ Интернета и смежных технологий в области медицины и здравоохранения.

**Заключительные замечания по предметным каталогам.** Хороший обзор и общее представление по предмету поиска медицинской информации дает Medical Matrix. Если нужно выделить британские информационные ресурсы, то следует пользоваться OMNI. Когда необходимо найти определенный сайт или получить информацию по отдельному географическому региону, целесообразно использовать Health on Net.

**Преимущества оцененных предметных каталогов.** Благодаря тому, что перед размещением информации на сайте она оценивается профессионалами, такого рода каталоги дают возможность быстро и легко получить медицинские данные. Краткое описание предложенных Поисковой Машиной сайтов позволяет получить предварительное представление о сайте, что избавляет от необходимости просмотра ненужных источников информации.

**Недостатки оцененных предметных каталогов.** Так как каталоги составляются не автоматически, то потенциально возможны упущения важных информационных ресурсов; большинство каталогов такого рода находится на начальной стадии развития и потому содержат ограниченный перечень информационных ресурсов.