

Поиск информации в Интернете

Павел Юрьевич Белкин, преподаватель Московского Центра
Интернет-образования

В настоящее время всемирная сеть Интернет хранит более миллиарда информационных объектов, таких как Web-документы, файловые архивы, архивы телеконференций и т.п. Очевидно, что эффективность использования информационного ресурса такой сложности требует от пользователя Сети определенных знаний, навыков и умений в области организации и проведения поиска информации в Интернет.

Именно поэтому большинством программ обучения, в том числе и школьными программами по информатике, предусмотрена тема "Поиск информации в Интернете" (или тема с похожим названием).

Нередко обучение в рамках этой темы заключается в том, что для формирования комплекса начальных представлений о процедуре поиска и навыков работы с поисковыми системами используются относительно несложные модельные примеры поисковых задач. Например, задачи типа "Найти информацию о родном городе", "Найти сведения о любимом писателе" и т.п. Как правило, целью подобного обучения, является знакомство с несколькими поисковыми системами и, в лучшем случае, - приобщение к работе с языками запросов этих систем. А в качестве проверки знаний от учащихся требуется продемонстрировать знание нескольких адресов и операторов языков запросов. Это хорошо, но этого мало. Основным недостатком такого подхода является то, что главным считается показать, КАК примерно искать, а ЧТО искать - это обучаемый сам для себя решит, когда ему это понадобится. С одной стороны, при этом теряется интерес обучаемых ("попробовали - и ладно"), а с другой - на двух-трех простых примерах нельзя серьезно показать (а тем более - изучить) методику эффективного поиска. Возможно, этот подход используется из-за непонимания до конца того, что же на самом деле представляет собой поиск информации в Интернет.

Иногда процесс обучения получает некоторое развитие: учащимся ставится поисковая задача из какой-нибудь предметной области, а результаты поиска служат базой для подготовки предметной работы (обзора, реферата, дипломного проекта и т.п.). В этом случае можно на практике проверить навыки и умения обучаемых. Интересный пример такого подхода описан в этом номере журнала в статье А. Андреевой. Безусловно, использование конкретных предметных задач весьма и весьма полезно при обучении поиску информации.

Однако применение такого подхода может привести к появлению другой проблемы. Дело в том, что в этом случае очень легко начать оценивать умения учеников по уровню новизны и "интересности" результирующей работы по сравнению со стандартным содержанием учебника. Хотя это и выглядит естественно и нормально при практическом решении прикладных задач, но все же недопустимо при обучении.

Для решения этой проблемы можно привлечь стандартные критерии оценки, например, "как много поисковых систем знает ученик" и "представляет ли, что такое язык запросов". Но это вряд ли может охарактеризовать умение учащегося искать информацию в сети Интернет. Да, конечно, можно поставить хорошую или отличную оценку, если ученик на основе найденных материалов написал прекрасный реферат с одной-двумя иллюстрациями и показал, что знает три поисковых системы. Но чем хуже будет не менее интересная, к тому же богато проиллюстрированная работа учащегося, использовавшего всего одну поисковую систему, с помощью которой он, манипулируя запросами, получил гораздо более представительную выборку документов, чем первый?

По мнению автора, главная проблема интегрирования обучения поиску и предметного обучения заключается в необходимости выработки эффективной формы обучения и комплексных критериев оценки знаний и умений учащихся. Причем оценивать необходимо знания и умения как по предмету, в рамках которого формулируется поисковая задача, так и по теме "Поиск информации в Интернет".

В данной работе мы попытаемся найти базовые составляющие для решения указанной проблемы.

Чтобы понять, что необходимо изучать и контролировать, вспомним, что представляет собой процедура поиска информации в Интернет. Обычно выделяют следующие этапы.

1. Формализация пользователем своей информационной потребности, например, путем составления списка ключевых понятий предметной области и определения смысловых (семантических) связей между ними.

2. Выбор информационного ресурса Интернет, на котором осуществляется поиск (поискового пространства).

3. Выбор информационно-поисковой системы, с помощью которой осуществляется поиск.

4. Построение на основе формализованной информационной потребности запроса - выражения на языке запросов информационно-поисковой системы.

5. Проведение поиска (ввод запроса в информационно-поисковую систему; выборка документов или других объектов, осуществляемая поисковой системой, а также формирование и выдача результата пользователю).

6. Анализ полученных материалов пользователем.

Рассмотрим подробнее эти этапы с целью анализа необходимых знаний и умений.

1. ФОРМАЛИЗАЦИЯ ИНФОРМАЦИОННОЙ ПОТРЕБНОСТИ

Все начинается с того, что у пользователя (обучаемого) появляется необходимость найти сведения по какому-либо вопросу. На языке теории информационного поиска это называется информационной потребностью.

Чтобы "объяснить" поисковой службе, что ему необходимо найти, пользователь вынужден некоторым образом формализовать свою информационную потребность. Наиболее очевидным способом формализации является выявление набора понятий и терминов, характеризующих информационную потребность.

Однако, получение простого списка понятий предметной области - это далеко не все, что необходимо сделать на данном этапе. Необходимо также провести анализ выбранных терминов, поскольку, с одной стороны, они могут оказаться слишком широкими, а с другой - они, как и любые слова языка, могут означать различные понятия (т.е. оказаться многозначными словами или омонимами).

Следовательно, во-первых, с каждым ключевым понятием необходимо связать уточняющую информацию. В роли такой уточняющей информации могут выступать:

- конкретизирующие термины (например, для понятия "физика", конкретизирующим термином может оказаться "ядерная");
- сведения о людях, которые связаны с появлением, разработкой и проч. данного понятия;
- сведения о странах, местностях, населенных пунктах и т.п., связанных с данным понятием;
- хронологические сведения.

Следует особо отметить, что формируемый список ключевых терминов со связными понятиями не является закрытым к моменту начала поиска - он может и должен пополняться на основе результатов итераций поиска.

Во-вторых, каждое понятие должно быть проанализировано с точки зрения таких лексических явлений, как полисемия (многозначность) и омонимия (наличия для данного термина омонимов - слов, совпадающих с ним по написанию и звучанию, и омографов - слов, совпадающих по написанию, но различающихся по звучанию). Кроме того, для расширения объема получаемых данных, необходимо для каждого понятия построить синонимический ряд. В свою очередь, все это может послужить базой для построения практикумов по предметам лингвистического цикла (родной и иностранные языки).

В качестве примера рассмотрим ту же задачу, что и в упомянутой выше работе А. Андреевой - поиск материалов по теме "Произведение А.С. Пушкина "Капитанская дочка"". Конечно, в такой постановке задача выглядит достаточно общей, но, тем не менее, попробуем построить список ключевых понятий.

Итак, мы имеем дело с литературным произведением - сущностью, с которой связаны, как минимум, автор, название, сюжет и персонажи. Т.е. в список основных терминов попадут:

- "Пушкин";
- "Капитанская дочка";
- "Крестьянское восстание под предводительством Пугачева";
- "Гринев", "Швабрин", "Пугачев", "Маша" и т.д.

В качестве конкретизирующих понятий можно использовать:

- понятие "персонаж", связав его с соответствующими фамилиями;
- имена, позволяющие уточнить людей, представленных фамилиями, например, связав "Александр Сергеевич" с фамилией "Пушкин", можно отсеять возможные материалы о В.Л. Пушкине.

В качестве персоналий, связанных с литературным произведением, можно рассматривать

- литературоведов и критиков;
- прототипов персонажей.

К уточняющим географическим сведениям можно отнести места, которые связывают автора с созданием произведения, а также - которые связаны с сюжетом. В нашем примере - это Оренбург.

Аналогично, хронологические сведения - это период создания произведения (1833-1836) и время действия сюжета.

Естественно, что на начальном этапе поиска такие конкретизирующие данные могут быть не известны, если пользователь (ученик) не достаточно хорошо знаком с предметной областью. В этом случае их можно попытаться найти. Кстати, решение поисковой подзадачи типа "Прототипы героев повести" позволит, с одной стороны, расширить список ключевых понятий, а с другой, – найти часть материалов, относящихся к основной (и более общей) задаче.

Далее каждое понятие должно подвергнуться языковому анализу. Здесь, например, следует вспомнить, что А.С. Пушкин может упоминаться в документах по-разному: "Александр Сергеевич Пушкин", "А.С. Пушкин" или "Александр Пушкин".

В качестве отдельных упражнений можно предложить ученикам составить список синонимов для терминов "персонаж" или "литературовед". Может быть, для русского языка такие упражнения покажутся неинтересными, но если поисковая задача связана с иностранными языками, то их можно рассматривать и как контрольные.

Итак, на выходе этого этапа получается список основных терминов со связанными понятиями примерно такого вида:

Исходное понятие

Связанные понятия

...

"А.С.Пушкин"

"Александр Сергеевич Пушкин"

"Александр Пушкин"

"Оренбург"

"1833-1836"

...

"Гринев"

"персонаж"

"герой"

"Швабрин"

...

Как уже упоминалось, составленный в самом начале список может (и должен) пополняться при проведении поиска. Например, если на какой-либо итерации ищались (и были удачно найдены) сведения о прототипах героев, то в этот список будут внесены дополнения:

...

"Гринев"

"персонаж"

"герой"

"Михаил Шванвич"

"Михаил Александрович Шванвич"

"М.А. Шванвич"

"Швабрин"

...

Построение такого списка со связанными понятиями оказывается весьма важным для обучения, т.к.:

- с точки зрения поиска информации, служит важным подготовительным этапом проведения поиска;
- с точки зрения предмета, в рамках которого сформулирована поисковая задача, способствует научению работе с понятиями, а также может расширить знания в предметной области.

Сам по себе список, может, с одной стороны, рассматриваться как результат, позволяющий контролировать знания по предмету, а с другой - служить базой для построения терминологического справочника учащегося, который может стать наряду с отчетом о знаниях также и шпаргалкой.

Хотелось бы еще заметить, что сравнение начального состояния списка с состоянием после окончания поисковых работ может дать преподавателю почву для анализа качества проведения поиска (особенно, если для вносимых по результатам выполнения поиска понятий будет указан запрос, результаты которого задействовались). Это оказывается полезным в тех случаях, когда учащийся по каким-либо причинам не справился с написанием итоговой предметной работы.

Справедливости ради надо сказать, что выполнение подобной аналитической работы полезно и для преподавателя. Это позволяет избежать постановки слишком общей или абстрактной поисковой задачи, заменив ее более узкими подзадачами, определившимися в процессе анализа.

2. ВЫБОР ИНФОРМАЦИОННОГО РЕСУРСА ИНТЕРНЕТ

На данном этапе пользователь должен выбрать информационный ресурс и, следовательно, - службу (сервис), которая обеспечивает работу с данным ресурсом.

Этот этап достаточно важен, поскольку он определяет дальнейшую технологию поиска. Однако в подавляющем большинстве программ обучения, в том числе и программах школьного обучения, дело не идет

дальше наиболее популярного и весомого ресурса World Wide Web. Поэтому в данной работе будет предполагаться, что обучение проводится применительно к WWW (хотя, с точки зрения обучения использованию Интернет, на данном этапе можно проконтролировать знание учащимися сетевых сервисов).

И все же, знание того, что существует WWW, в рамках которого будет осуществляться поиск на данном этапе, недостаточно. Здесь важно понимать, что собой представляет этот ресурс как информационное пространство и каковы его особенности.

Как известно WWW - это сервис, одной из главнейших концепций которого является концепция "универсальной читаемости" (Universal readership). Это означает, что с помощью всемирной паутины можно осуществлять доступ к большому числу разнородных информационных объектов: web-документам, файлам графических изображений, аудио- и видеороликам, документам Microsoft Office или PDF-документам, компьютерным анимациям и т.д. Естественно, что по различным, в том числе и техническим, причинам, не все объекты могут быть получены в процессе проведения поиска. Поэтому, на данном этапе обучаемый должен четко ориентироваться в том:

- какого вида материалы могут быть найдены с помощью выбранного ресурса;
- какие из них могут представлять интерес с точки зрения поисковой задачи, т.е. что следует искать.

Ответы на эти вопросы должны определить схему проведения следующих трех этапов и, возможно, - скорректировать список ключевых понятий.

Итак, вернемся к нашему примеру. В настоящее время наиболее просто на пространстве WWW можно осуществлять поиск следующих объектов:

- Web-документов (файлы в формате HTML);
- текстовых файлов;
- файлов изображений (в форматах GIF, JPEG и PNG);
- звуковых файлов (в форматах RealAudio и MP3).

С точки зрения поставленной поисковой задачи, имеет смысл искать первые три типа объектов. Поиск Web-документов и текстовых файлов осуществляется обычным, хорошо известным способом. А вот для поиска иллюстраций часто требуется строить запрос специальным образом, указывая, например, что ключевые слова должны присутствовать только в подписях к рисункам или связывая с ключевыми словами определенные уточняющие понятия. Это необходимо для того, чтобы снизить число выдаваемых информационно-поисковой системой документов, не относящихся к представлению иллюстративного материала.

В качестве иллюстраций для итоговой предметной работы по указанной теме могут выступать портреты автора, портреты прототипов героев, портреты героев, иллюстрации произведения. Следовательно, возможно проведение поиска вхождения выражения "А.С. Пушкин" в

подписях к рисунку, и это придется учитывать на следующих этапах, а также - поиска выражения "А.С. Пушкин" в связи с термином "портрет", что требует коррекции списка ключевых понятий.

3. ВЫБОР ИНФОРМАЦИОННО-ПОИСКОВОЙ СИСТЕМЫ

Это наиболее известный этап проведения поиска. По опыту автора, большинство пользователей Интернет считают даже его первым этапом. Именно число (и список) известных поисковых систем является наиболее популярным (а порой - даже единственным) критерием оценивания учащихся.

Что же необходимо знать на данном этапе? Главное - это типы и соответствующие возможности систем, а также адреса конкретных систем. Эта тема достаточно подробно разработана в литературе, поэтому хочется обратить внимание лишь на проблему контроля знаний.

Очевидно, что вопрос о том, "какая система является лучшей" является бессмысленным. Равно как и попытка оценить высоко работу, если материалы были получены с нескольких систем, и низко - если с одной-двух, поскольку какие-то системы могут при использовании определенных запросов не дать удовлетворительного результата.

Следовательно, можно осуществлять контроль знаний на этом этапе отдельно от других. Но в этом случае нельзя напрямую связывать его с полученным результатом. Гораздо удобнее использовать для контроля комплексный (для этого и двух следующих этапов) отчет выполнения контрольной поисковой задачи. Об этом пойдет речь ниже.

4. ПОСТРОЕНИЕ ЗАПРОСА

Это тоже достаточно проработанный этап. Здесь прививаются, проявляются и проверяются умения и навыки использования языков запросов для выражения установленных на этапе формализации связей между ключевыми понятиями, а также умение учитывать особенности поиска объектов различных типов, список которых определен на втором этапе (например, для WWW - Web-документов, файлов изображений и т.п.).

Остановившись на проблеме контроля знаний, хотелось бы отметить, что оценка должна осуществляться комплексно с учетом предыдущего этапа, поскольку хорошие знания языка запросов одной системы могут не привести даже к мало-мальски существенному результату, если данная система не сможет выдать удовлетворительный результат по предъявляемым запросам.

5. ПРОВЕДЕНИЕ ПОИСКА И ПОЛУЧЕНИЕ РЕЗУЛЬТАТА

С одной стороны, этот этап наиболее простой, поскольку требует от пользователя (обучаемого) знаний на темы "где написать" и "на что нажимать". С другой же стороны - относительно сложен, т.к. при получении результата нужно отсеять лишние (нерелевантные) материалы (шум), а на основе остальных - принять решение, как корректировать запросы и стоит ли менять поисковую систему для следующей итерации поиска, или же следует

детально проанализировать найденную информацию и принимать решение потом.

Следовательно, научить необходимым для этого навыкам, с точки зрения технического обучения, достаточно просто, а с точки зрения работы с документами, - нелегко, поскольку от учащегося требуется хорошо понимать предметную проблему, в рамках которой проводится поиск - а это уже задача учителя-предметника.

При контроле знаний очень важно отделять технические умения и умение работать с информацией. Именно поэтому и вводится еще один этап поиска - этап анализа результатов.

На этом же этапе было бы целесообразно подводить черту под результатами третьего и четвертого этапов, что продемонстрировало бы именно техническую сторону знаний.

Автору представляется весьма удобным построение комплексного отчета (в виде таблицы), содержащего по каждой итерации поиска следующую информацию:

- название (адрес) информационно-поисковой системы;
- использованный запрос;
- тип объектов поиска (документ, иллюстрация и т.п.);
- число найденных объектов;
- число отобранных из них для дальнейшего детального анализа.

Исследование такого отчета может показать, как учащийся анализировал и решал задачу, какой путь он выбрал - усложнение и расширение запросов для одной системы или же использование возможностей нескольких систем, а также какие типы объектов он пытался искать и нашел. Например, в нашем случае, отчет может выглядеть примерно так:

№	Поисковая система	Запрос	Тип объектов	Найдено	Отобрано
...					
29	Rambler	"Капитанская дочка" прототип	документы	4	1
30	Yandex	"Капитанская дочка" прототип	документы	22	

...

Используя такую форму, появляется возможность сравнивать решения разных учеников. В частности, в рассматриваемом примере возможны два решения, учитывающих вариативность написания имени Пушкина: можно использовать оператор "логическое ИЛИ" для объединения выражений "Александр Сергеевич Пушкин" и "Александр Пушкин", а также можно использовать оператор расстояния, допускающий возможность употребления одного слова между терминами "Александр" и "Пушкин". Если учащийся выбрал второй вариант (который явно экономичнее по написанию) в системах Yandex и Aport (где имеются четко определенные операторы расстояния) и первый - в системе Rambler (где оператор расстояния NEAR на самом деле не позволяет четко задать расстояние), то это, безусловно, плюс. А вот если ученик вообще не использовал операторы расстояния, существенно перегружая запросы логическими операторами, то это уже повод призадуматься о его знаниях.

6. АНАЛИЗ МАТЕРИАЛОВ

Данный этап предназначен для окончательного отбора материалов, которые будут использоваться при создании предметной работы. Помимо умения ориентироваться в предметной области на данном этапе обычно требуется проанализировать источник информации, т.е. понять, что собой представляет сайт, на котором опубликован материал, и кто является его автором. С этой целью удобно с каждым из документов, который отбирается для дальнейшего использования, по возможности, связать следующую информацию:

- краткая характеристика сайта: общая тематика, сведения о разработчиках;
- наличие на сайте других материалов, связанных с темой поиска;
- сведения об авторе или авторах материала, тематика других работ этих авторов (если возможно найти).

Можно на основе этих данных также составить итоговый отчет, связав его с отчетом предыдущего этапа.

Например:

...

Статья **"Шванвич** **Михаил** **Александрович"**
[\(http://www.orenburg.ru/culture/encyclop/tom2/tom2_fr.htm\)](http://www.orenburg.ru/culture/encyclop/tom2/tom2_fr.htm)

Найдено при обработке запроса 29

Сайт: Оренбургская пушкинская энциклопедия
[\(http://www.orenburg.ru/culture/encyclop/tom2\)](http://www.orenburg.ru/culture/encyclop/tom2)

Сайт представляет собой энциклопедический проект Оренбургского государственного университета. Инициатор и руководитель проекта заслуженный деятель науки РФ Л.Н. Большаков. Авторы проекта доктор исторических наук Р.В. Овчинников и академик Международной академии

гуманизации образования Л.Н. Большаков. Сайт полностью посвящен произведениям А.С. Пушкина "Капитанская дочка" и "История Пугачева".

...

Подобный отчет оказывается полезным, т.к.:

- учащийся приучается анализировать найденную информацию;
- снижается вероятность использования непроверенных материалов;
- в случае получения предметной работы низкого качества это позволяет понять причину - "технически плохо искал" или "недостаточно владеет предметными знаниями";
- в случае получения хорошей или удовлетворительной итоговой работы это позволяет понять степень самостоятельности работы: действительно ли ученик рассмотрел несколько материалов и произвел переработку и обобщение, или все списано из одного-двух источников.

В частности, в примере, приведенном в статье А.Андреевой, одна из учениц не учла, что найденный материал - это текст некорректного сочинения. А причина не ясна: не учла, потому что забыла проверить источник, или проверила, но не смогла сделать вывод о корректности? Наличие такого отчета могло бы дать ответ на этот вопрос.

ВМЕСТО ЗАКЛЮЧЕНИЯ

Итак, при обучении поиску информации в сети Интернет, по мнению автора, важно:

- рассматривать процесс поиска поэтапно, что позволяет с одной стороны четко определить навыки и умения, необходимые для выполнения каждого этапа, а также их влияние на результаты поиска;
- фиксировать результаты проведения каждого этапа, поскольку это может дать материал для анализа знаний учащихся, а, следовательно, - как для их оценивания, так и для коррекции методики преподавания.

Вместе с тем понятно, что эта тема может быть легко связана с различными темами предметов средней школы, а процесс обучения - рассматриваться как интегрированный, способствующий формированию и проверке как технических, так и предметных знаний.