В процедуре поиска информации в Интернет обычно выделяют следующие

этапы.

1. ***Формализация*** ***информационной потребности*** пользователем, например, путем составления списка ключевых понятий предметной области и определения смысловых (семантических) связей между ними. Так, например, для поиска электронных образовательных ресурсов в сети Интернет, пригодных к использованию в процессе обучения студентов высшего профессионального образования, целесообразно при формализации информационной потребности использовать дидактические единицы.
2. ***Выбор информационного ресурса Интернет***, на котором осуществляется поиск (поискового пространства).

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.

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).

Вопросы свободного доступа учащихся к качественным электронным образовательным ресурсам призваны решать Информационная система «Единое окно доступа к образовательным ресурсам» (window.edu.ru), а также портал «Федеральный центр информационных образовательных ресурсов» (ФЦИОР, http://fcior.edu.ru).

ИС «Единое окно» предоставляет свободный доступ к каталогу образовательных интернет-ресурсов и полнотекстовой электронной учебнометодической библиотеке для общего и профессионального образования и является уникальным образовательным проектом в русскоязычном Интернете, объединяя в единое информационное пространство более 27000 электронных ресурсов свободного доступа для всех уровней образования в России.

Каталог ФЦИОР объединен единой информационно-поисковой системой с Интернет-порталами «Единая коллекция цифровых образовательных ресурсов» [(http://school-collection.edu.ru)](http://school-collection.edu.ru/) и «Каталог учебников, оборудования, электронных ресурсов для общего образования» ([http://ndce.edu.ru)](http://ndce.edu.ru/).

Каждый учебный модуль (имеются модули трех типов: информационные, практические и контрольные) автономен и представляет собой законченный интерактивный мультимедиа продукт, нацеленный на решение определенной учебной задачи.

Для эффективного использования ЭУМ в высшей школе необходимо определить соответствие имеющихся модулей направлениям подготовки и учебным планам программ высшего профессионального образования.

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

Поисковые системы интернета представляют собой сложные программы для поиска информации на сайтах сети и web-интерфейс для взаимодействия с пользователем. *Web-интерфейс* – это сайт, где пользователь вводит в специальное окно свой поисковый запрос и получает результат поиска – ссылки на найденные страницы с небольшим анонсом.

Программная часть поисковика занимается сбором информации с сайтов, обработкой текстов с web-страниц, хранением этих текстов в специальной форме, поиском по хранящимся текстам и выводом пользователю ссылок на релевантные страницы в определенном порядке.

Основными критериями качества работы поисковой машины являются релевантность (степень соответствия запроса и найденного, то есть уместность результата), полнота базы, учёт морфологии языка.

Индексация информации осуществляется специальными поисковыми роботами.

По данным компании Net Applications в ноябре 2011 года использование поисковых систем на Западе распределялось следующим образом:

1. Google — 83,87 %
2. Yahoo — 6,20 %
3. Baidu — 4,22 %;
4. Bing — 3,69 %;
5. Yandex — 1,7 %;
6. Ask — 0,57 %;
7. AOL — 0,36 %.

В России, по данным LiveInternet на сентябрь 2012 года, доли поисковых систем распределяются следующим образом:

1. Яндекс — 52,8 %,
2. Google — 33,9 %,
3. Mail.Ru — 8,4 %, 4. Рамблер — 1,2 %.

Не все из перечисленных поисковых систем имеют собственный поисковый алгоритм – так QIP.ru используют поисковый механизм Яндекса, Mail.ru использует Google, а Nigma сочетает в себе как свой алгоритм, так и сборную выдачу от других поисковиков.

1. ***Построение*** ***запроса*** на основе формализованной информационной потребности – выражения на языке запросов информационно-поисковой системы.

*Поисковый запрос – введенные в поисковик слова, например, «список поисковых систем интернета».*

Виртуальному роботу-поисковику проще найти тот или иной документ, если поисковый запрос составлен с помощью специального языка — так называемого «языка запросов» – своеобразного словаря из нескольких десятков символов, куда входят, к примеру, плюс, минус, вертикальная черта, кавычки, восклицательный знак, тильда и т. д. Грамотное использование этих элементов способно значительно ускорить и облегчить поиск.

Формализованный запрос на поиск тематического ЭОР, построенный с использованием дидактических единиц, при необходимости может быть уточнен и конкретизирован предметной областью задачи.

1. ***Организация поиска и получение результата***. Онтологический подход1.

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.

Так как при обращении к поисковой системе пользователь должен иметь возможность получить в ответ ресурсы, релевантные смыслу запроса, то их поиск должен быть семантически ориентированным. Для этого средства поиска соответствующей запросу информации предлагается организовать на основе онтологии, содержащей описания семантики ресурсов.

1Россеева О.И., Загорулько Ю.А. *Организация эффективного поиска на основе онтологий*. Российский

НИИ Искусственного Интеллекта, Институт систем информатики СО РАН [http://dialog-](http://dialog-21.ru/Archive/2001/volume2/2_49.htm)

[21.ru/Archive/2001/volume2/2\_49.htm](http://dialog-21.ru/Archive/2001/volume2/2_49.htm)

Известно, что семантика Internet-ресурсов очень разнообразна, следовательно, осуществлять поиск необходимой информации тем проще, чем уже и специфичнее предметная область. Вследствие этого на практике можно ограничиться построением онтологии одной конкретной области.

Для построения онтологии требуется формальное декларативное представление четко организованных конструкций, которые включают в себя словарь терминов тематической области, описание определений этих терминов, существующие взаимосвязи между ними, их теоретически возможные и невозможные взаимосвязи. Описанные таким образом онтологии предлагается применить в качестве посредника между пользователем и поисковой системой (см. рис. 1). Взаимодействие с онтологией предполагается на следующих этапах:

* формирование поискового образа релевантного документа;
* построение запроса к поисковой системе;
* формирование списка релевантных документов.

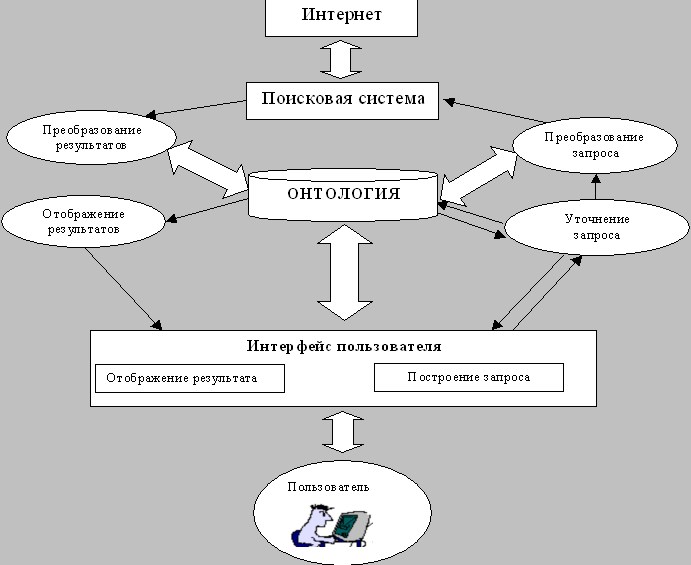


Рисунок 1- Схема организации поиска на основе онтологии

Проблема состоит в том, чтобы сделать поиск динамичным и удобным для пользователя. Для любого типа запроса, возникающего у человека в практической деятельности, должны быть найдены адекватные знания в информационном пространстве Internet. При этом язык для формулирования поискового требования не должен быть слишком сложным. В частности, общение пользователя с поисковой системой можно сделать более простым, приблизив язык запроса к естественному языку.

При такой организации поиска на этапе формирования образа релевантного документа из пользовательского запроса выделяются смысловые структуры: значимые слова и термины предметной области. Эти смысловые структуры затем используются для формирования поискового образа с применением эвристических правил и вывода на онтологии. Образ релевантного документа представляет собой описание желаемого результата работы поисковой системы, которое включает в себя:  набор терминов, которые должны включаться в документ;

* набор характеристик документа;
* набор требований к результату поисковой системы, таких как количество документов и т.п.

На этапе построения запроса к поисковой системе осуществляется вывод на онтологии. При этом выполняется преобразование пользовательского запроса в соединенный логическими связками набор терминов и понятий, которые будут использоваться поисковой системой.

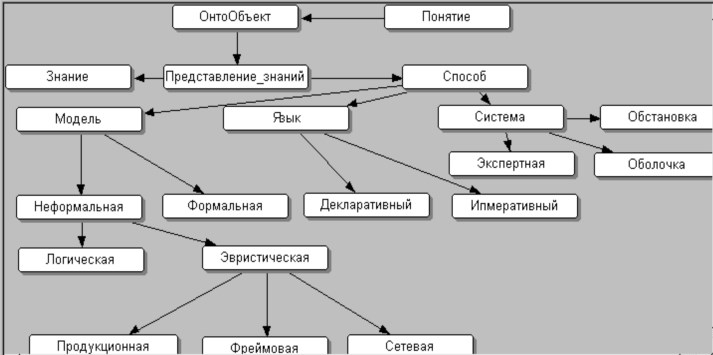


Рисунок 2 - Фрагмент иерархии понятий предметной области

Результатом работы поисковой системы является множество ссылок на HTMLдокументы. Так как среди этого множества ссылок могут содержаться ресурсы, не имеющие ничего общего с запросом пользователя, то на следующем этапе происходит проверка результатов работы поисковой системы на соответствие поисковому образу документа. Анализ соответствия результатов HTML-документов должен основываться на выделении ключевых слов из таких частей документа как заголовки, ссылки и тела документа. Так как результатом работы поисковой системы может быть большое количество ссылок, то анализ документов должен проводиться на основе жестких критериев отбора. После проведенного «отсечения» лишних документов результаты поиска отображаются в удобном пользователю виде*.*

Согласно описанной выше схеме, на каждом шаге движения запроса от пользователя к поисковой системе и, наоборот, от поисковой системы к пользователю происходит оценка релевантности смысла запроса и смысла результатов запроса. Эта оценка осуществляется на основе механизма вывода на онтологии.

На основе представленных в онтологии отношений между описанными терминами осуществляется вывод на следующих этапах:

* построение поискового образа документа, релевантного запросу;
* формирование запроса для поисковой системы на основе построенного образа документа;
* проверка соответствия результатов работы поисковой системы образу релевантного документа.

Реализация описанной схемы организации поиска с использованием онтологий предполагает наличие следующих этапов:

* спецификацию языка запроса на естественном языке;
* построение онтологии;
* выбор средств анализа результатов поиска;
* создание интерфейса между онтологией и пользователем; этот интерфейс должен обеспечивать настройку онтологии на конкретного пользователя, построение и ввод запроса на ЕЯ, просмотр результатов запроса;
* создание модуля взаимодействия онтологии с поисковой системой; данный модуль должен обеспечивать перевод запроса, преобразованного онтологией, в соответствующий формат запроса поисковой системы и отображение результатов выполнения запроса.

6. ***Анализ*** полученных и ***отбор*** ***релевантных материалов*** пользователем.

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.

В современных информационных системах учебные материалы (в технологической терминологии – учебные объекты) хранятся в специально организованном хранилище – репозитарии. Множество хранимых объектов упорядочено иерархически. Элементарные (неделимые) объекты могут быть составными частями сложных объектов. При этом не учитываются связи объектов, относящихся к контекстно связанным предметным областям (например, математика, химия, и т.п.). Не учитываются также иерархические отношения между понятиями внутри одной предметной области (например, математика-алгебра, математикагеометрия и т.д.). В существующих поисковых системах используются, как правило, только формальные признаки объектов: рабочий язык, заголовок, интерактивность сложность, платность[[1]](#footnote-1), то есть в критериях поиска не учитываются специфика содержания учебного материала. А поиск по ключевым словам далеко не всегда отражает реальное содержание понятий в силу произвольности описания и свойств естественного языка (омонимия, синонимия и т.п.).

Таким образом, очевидно, что формальные критерии необходимы, но недостаточны для эффективного поиска и использования объекта в учебном процессе.

При объектном подходе учебный материал разбивается на учебные объекты (ОУ), каждый из которых может многократно использоваться как отдельно, так и во взаимодействии с объектами на других программных платформах, поддерживающих единые международные технологические стандарты. При этом появляется возможность композиции (агрегирования) сложных объектов из совокупности отдельных, включения объектов в логически упорядоченные последовательности. В виде самостоятельных объектов могут представляться также информационные ресурсы стандартных форматов: тексты, рисунки, анимация, аудио- и видеофрагменты, webстраницы, web-сайты, документы, PowerPoint- презентации и т.п. Любой учебный

объект может многократно использоваться для достижения различных учебных целей.

Для организации ***автоматизированного*** ***поиска ЭОР*** содержание ОПОП и метаданных ЭОР должно быть описано в одних и тех же ключевых понятиях (ключевых словах) предметной области (дисциплины).

При разработке ЭОР автор должен строго придерживаться ключевых понятий (ключевых слов), содержащихся в утвержденных примерных основных образовательных программах. Цель и задачи ЭОР, указанные в соответствующих полях метаданных ЭОР, должны быть сформулированы с использованием ключевых понятий, соответствующих утвержденным ОПОП.

Также для автоматизированного поиска ЭОР необходима эффективная технология индексирования информации, содержащейся в ЭОР и ОПОП и составление онтологий обоих объектов (онтологии ЭОР и онтологии ОПОП), при этом множество ключевых понятий онтологии ЭОР должно быть подмножеством множества онтологии ОПОП.

При составлении профиля метаданных потребительских характеристик ЭОР, в частности, полей «Аннотация», «Реферат», автор-разработчик ЭОР должен строго придерживаться требований стандартов РФ, в частности:

* ГОСТ 7.23 Издания информационные. Основные виды, структура и оформление;  ГОСТ 7.59 Индексирование документов. Общие требования к систематизации и предметизации;
* ГОСТ 7.66 Индексирование документов. Общие требования к координатному индексированию;
* ГОСТ 7.9 (ИСО 214-76) Реферат и аннотация.

Должен быть разработан эффективный инструмент сопоставления онтологий ОПОП и ЭОР, результатом работы которого будет являться заполненная форма представления оснащенности ОПОП открытыми ЭОР по каждой дисциплине каждой специальности ВПО.

В рамках решения текущей задачи может быть предложена следующая форма представления оснащенности ООП ВПО открытыми ЭОР, содержащая следующие обязательные поля: 1. Уровень образования

1. Специальность
2. Квалификация (степень)
3. Дисциплина
4. Индекс и наименование дисциплины по ФГОС
5. Раздел дисциплины
6. Подраздел дисциплины /Тематический раздел
7. Ключевые слова
8. Ссылка на ЭОР
9. Ссылка на ЭОР
10. Ссылка на ЭОР

………………

N. Ссылка на ЭОР

Поля формы спроектированы в виде «дерева» объектов, что дает возможность организации информационного поиска по иерархическому признаку. Поля 1-8 соответствуют ООП ВПО и заполняются автором формы. Затем форма передается в поисковую машину для заполнения полей 9-N..

Важной задачей является также необходимость самообучаемости интеллектуального ассистента в непрерывном поиске и определении «удельного качественного веса» появляющихся (новых) на образовательном рынке ЭОР, замены менее «качественных» ЭОР на более «качественные» в тематической сети ОПОП. Таким образом, получаем динамичную тематическую сеть дисциплины, с непрерывно повышающимися качественными характеристиками содержания набора ЭОР.

Подобный инструмент, естественно, также требует решения задачи индивидуализации для каждого конкретного пользователя, основываясь на его типе личности, представлении мира (онтологии), психолого-педагогических характеристиках и т.д., при этом:

* пользователь должен иметь возможность анализировать степень полезности предлагаемых ему сетью данных (ЭОР) с возможным ранжированием по релевантности/полноте/доступности изложения и т. п.

система должна уметь анализировать оценки, полученные от пользователя и ранжировать ЭОР-ы в индивидуальные тематические сети по дисциплине.

1. См. IMS Metadata. … [↑](#footnote-ref-1)