

Применение кластерного поиска в образовании

А.В. Дворецкая,
А.В. Рафаева

В современном мире Интернет стал одним из основных источников информации, и роль этого источника со временем возрастает. В образовательной практике Интернет используется наряду с книгами, учебниками, методическими пособиями и периодическими изданиями. Сейчас уже большинство крупных библиотек и серьезных периодических изданий, как российских, так и зарубежных, имеют собственные электронные ресурсы. Интернет делает доступным удаленные ресурсы (например, материалы зарубежных образовательных сетей или ресурсы российской системы открытых электронных библиотек OPEL), облегчают общение и доступ к информации, в том числе редким и труднодоступным изданиям, а также предоставляет ряд других возможностей.

Умение пользоваться ресурсами и возможностями Интернета, в частности, быстро находить нужную информацию, — стало привычкой, необходимым как учащимся, так и педагогам. Некоторые современные образовательные технологии, например, метод проектов¹ и ТОГИС (технология образования в глобальном информационном сообществе)² не только предполагают, а прямо требуют работы учащихся в глобальном

¹ Метод проектов: Научно-методический сборник / Белорусский государственный университет. Центр проблем развития образования. Республиканский ин-т развития высшей школы БГУ. Минск: РИВШ БГУ, 2003. (Серия «Современные технологии университетского образования», выпуск 2). С. 48–62.

² Материалы по образовательной технологии ТОГИС. Версия 2.0: Сб. статей под общей ред. В.В. Гузеева. М.: Илекса, 2002. (Серия «Системные основания образовательной технологии»).

информационном пространстве. В то же время использование Интернета в образовании затрудняется тем, что его ресурсы каталогизированы лишь в небольшой части, и полностью не могут быть каталогизированы в принципе. Это делает нахождение нужной информации весьма непростой задачей. Вследствие этого учителю, прежде чем давать задание, требующее использования ресурсов Интернета, приходится предварительно провести весьма трудоемкий поиск и предварительную оценку ресурсов по выбранной тематике. Конечно, это сильно затрудняет работу.

С другой стороны, не оценить предварительно, какие именно ресурсы могут быть найдены по заданной тематике, много их или мало, как лучше построить поиск, с каких ресурсов его начать, особенно при работе с недостаточно подготовленными учащимися, невозможно. Таким образом, возникает необходимость оптимизации поиска и предварительной оценки информации.

В предлагаемой статье рассмотрены следующие вопросы:

- стратегии поиска текстовой информации в сети Интернет;
- особенности систем поиска с кластерным анализом результатов поиска;
- результативность использования кластерного поиска в преподавании;
- план прохождения темы «Поиск в Интернете».

Статья посвящена использованию общедоступных поисковых сервисов Интернета. Для проведения занятия по предлагаемому сценарию необходимо иметь компьютерный класс с доступом в Интернет и программное обеспечение (Интернет-обозреватели

или браузеры), установленное на компьютерах учащихся. Лекционная часть курса сопровождается показом видеопрезентаций, поэтому компьютер учителя должен быть обеспечен соответствующим программным обеспечением и подключен к проектору. Таким образом, вопросы использования специализированного программного обеспечения, предназначенного для оптимизации поиска и устанавливаемого на персональный компьютер пользователя, мы не рассматриваем.

Стратегии поиска информации в сети Интернет

Структура глобальной информационной сети весьма неоднородна: в Интернете можно найти как массу полезной, а иногда даже уникальной информации, так и бесполезные и вредоносные ресурсы. Поэтому важно выбрать оптимальную стратегию поиска. Рассмотрим стратегии поиска.

Поиск по ссылкам и поиск по известному адресу. Сайты в сети Интернет связаны между собой посредством **гиперссылок**. Большинство специализированных сайтов содержат ссылки на другие страницы или сайты, посвященные этой же тематике. Этот метод поиска довольно трудоемкий, однако, часто приносит хорошие результаты. Даже при использовании других методов поиска имеет смысл посмотреть, нет ли на найденном сайте раздела ссылок (иногда этот раздел может называться и по-другому, например, Links, Ресурсы, Другие сайты и т.п.). Часто в таком разделе можно найти адреса ещё нескольких ресурсов, тематически связанных с данным, с комментариями автора или ведущего сайта.

Поиск по тематике. Этот вид поиска используется, как правило, в тех случаях, когда известно лишь направление поиска (например, *образование, музыка*) и нужно составить общее представление об имеющихся ресурсах на заданную тему. В таком случае можно воспользоваться каталогами ресурсов и тематическими подборками поисковых сайтов.

В пособии «Практический Интернет для библиотекарей» рекомендуется также попробовать набрать адрес www.»тема«.ru, где «тема» — транслитерация или английский перевод русского названия. Довольно часто этот приём приводит на российский сервер, посвящённый выбранной тематике. Однако возможности приёма не следует переоценивать, поскольку сетевой адрес принадлежит тому, кто его первым зарегистрировал. Например, федеральный портал Российское Образование можно найти по адресу www.edu.ru, а по адресу www.education.ru находится сайт Центра Обучения и Тестирования специалистов «САМАН-МАТИ».

Поиск по ключевым словам. Для поиска по ключевым словам используются поисковые машины, число которых в настоящее время достаточно велико. После выполнения запроса поисковая машина выдаёт список всех страниц, на которых встречается запрашиваемое слово или совокупность слов. Этот вид поиска — самый гибкий и мощный, он позволяет искать в Сети информацию, посвящённую самой разнообразной, в том числе, узкоспециальной, тематике. В то же время выбор ключевых слов, позволяющих быстро найти именно то, что нужно, достаточно сложен.

Поиск по ключевым словам можно выполнять также с помощью мета-

поисковой системы или при помощи поисковых систем с кластеризацией результатов.

Каждая из перечисленных выше стратегий поиска имеет свои достоинства и свои недостатки. Рассмотрим их с этой точки зрения.

По известному адресу. Поиск информации по известному адресу эффективен, но не всегда есть адрес необходимого ресурса. В строгом смысле слова этот вид работы в Интернете нельзя назвать поиском: предполагается, что в данном случае пользователь работает с конкретной страницей (в ряде случаев — с сайтом, который содержит множество страниц). Таким образом, речь идёт скорее о навыках работы с одним сайтом, а не о поиске в Интернете в целом.

С помощью тематического каталога — тематические каталоги довольно удобны в работе, многие аннотированы. Недостатком каталогов является их неполнота, в частности, они почти не содержат ссылки на многие личные страницы и персональные сайты. Кроме того, тематические каталоги, особенно созданные энтузиастами в какой-то области, достаточно быстро устаревают. Основное достоинство тематических каталогов — это то, что такие каталоги создаются людьми и, следовательно, отбор сайтов, на которые в каталоге помещаются ссылки, производится не по формальным, а по содержательным признакам.

По ссылкам, начиная с известного адреса — такая стратегия поиска требует большого опыта работы в сети Интернет. В частности, одна из рекомендаций для тех, кто прибегает к этому виду поиска, — не увлекаться и помнить, ради чего, собственно, поиск

производится. В общем случае поиск по ссылкам можно рекомендовать в качестве дополнения к другим видам поиска.

С помощью поисковой машины по ключевым словам — если адрес (искомого ресурса, тематического каталога и т.п.), с которого можно начать поиск, неизвестен, применяется стратегия поиска по ключевым словам. Однако каждая из поисковых машин имеет свою «зону предпочтения», свои особенности языка запросов и вывода найденной информации. Вследствие этого многие пользователи, освоив одну-две поисковые машины, в дальнейшем пользуются в основном ими. Это сужает охват ресурсов, в то же время не облегчая отбор и оценку результатов поиска. В целом при обращении к поисковым машинам желательно учитывать особенности различных поисковых машин.

При занятии в группе, посвящённой некоторой конкретной теме, учащиеся составляют поисковый запрос с использованием одних и тех же ключевых слов, вследствие чего получают одинаковый результат. В результате вся группа работает с 5–10 сайтами (из нескольких тысяч или даже миллионов) по изучаемой теме. Такой ситуации можно избежать, устроив предварительное обсуждение ключевых слов для поиска, однако, подобное обсуждение требует опыта работа в Интернете и может быть рекомендовано для группы продвинутого уровня.

Необходимо знать различные стратегии поиска и уметь их применять в зависимости от ситуации. Все описанные выше стратегии поиска не-

совершенны; в то же время в ряде ситуаций хотелось бы иметь какое-то средство, позволяющее быстро составить мнение о некоторой проблеме, оценить, какие именно ресурсы могут быть найдены, не тратя на подобный анализ нескольких часов, а то и дней.

Одним из общедоступных способов решить эту проблему являются системы кластерного анализа³, как оснащенные своими поисковыми машинами, так и метапоисковые, т.е. использующие результаты поиска, проведённого несколькими поисковыми машинами. **Поиск информации при помощи поисковых систем с кластеризацией результатов** можно рассматривать как самостоятельную стратегию поиска. Системы кластерного поиска — относительно новое, малоизвестное средство, поэтому на них мы остановимся подробно.

Особенности систем с кластерным анализом результатов поиска

Метапоисковые системы — это поисковые системы, использующие результаты поиска нескольких поисковых машин, они передают один запрос нескольким поисковым системам, позволяя, одновременно искать информацию в нескольких базах данных.

— параметры поиска минимальны: пользователь может задать время ожидания и диапазон (поиск среди англоязычных страниц или общий поиск). Есть возможность выбора поисковых машин из предложенного списка.

³ Баловсяк Н. Поиск совместными усилиями — [электронный ресурс] — http://www.comprice.ru/set/2005-31_2.phtml

Средства управления полученным списком ссылок — одни из лучших. Система предлагает наиболее релевантные результаты поиска, полученные отдельными поисковыми машинами. Система настраиваемая: скачав специальные плагины, можно задать разное представление результирующего списка и внешнего вида страницы.

Обе эти системы англоязычные.

Основной недостаток метапоисковых систем — многократное дублирование найденных ресурсов. Кроме того, одновременное использование нескольких поисковых систем приводит к тому, что дополнительные возможности поиска, предоставляемые каждой из отдельных поисковых систем, остаются невостребованными. В некоторых случаях это обстоятельство несущественно, в других же полностью лишает пользователя возможности найти искомое.

Системы кластерного анализа информации представляют пользователю результаты поиска, уже объединённые в тематические разделы (*кластеры*). Процесс такого объединения называется *кластеризацией*.

Кластеризация результатов поиска производится автоматически. После получения запроса пользователя система кластерного анализа просматривает найденные документы и анализирует их текст, выделяя, какие ещё слова или словосочетания встречаются вместе с искомыми. Те слова или словосочетания, которые встречаются достаточно часто, и служат средством, позволяющим группировать некоторые из найденных ресурсов.

Таким образом пользователь системы с кластерным анализом получает:

— результаты поиска по введённому запросу (как в обычной поисковой или метапоисковой системе);

— результаты кластерного анализа — список слов и понятий, которые встречаются вместе с искомыми ключевыми словами. Этот список позволяет обращаться к отдельным кластерам.

Общий вид такой поисковой системы после выполнения запроса «метод проектов» приведен на рисунке 1. Система выделила следующие основные кластеры (приводится только часть кластеров в том виде, в котором они выданы пользователю):

- полат
- обучения
- технология
- методы управления проектами
- методы оценки.

Мы видим, что количество ресурсов, выделенных в кластеры, сравнительно невелико (в нашем примере в каждом кластере число ресурсов исчисляется десятками, в то время, как общее число документов, содержащих искомые слова — более 14 миллионов). В дальнейшем пользователь может смотреть общие результаты поиска, переходить к выделенным кластерам или уточнить запрос, добавив новые ключевые слова (что рекомендуется для получения более полных результатов по наиболее интересным разделам). В любом случае, однако, можно получить некоторое общее представление о заданной теме и её представлении в Интернете.

Основные свойства систем кластерного поиска:

- Системы кластерного поиска могут иметь свои поисковые машины, а могут пользоваться результатами поиска известных поисковых серверов.

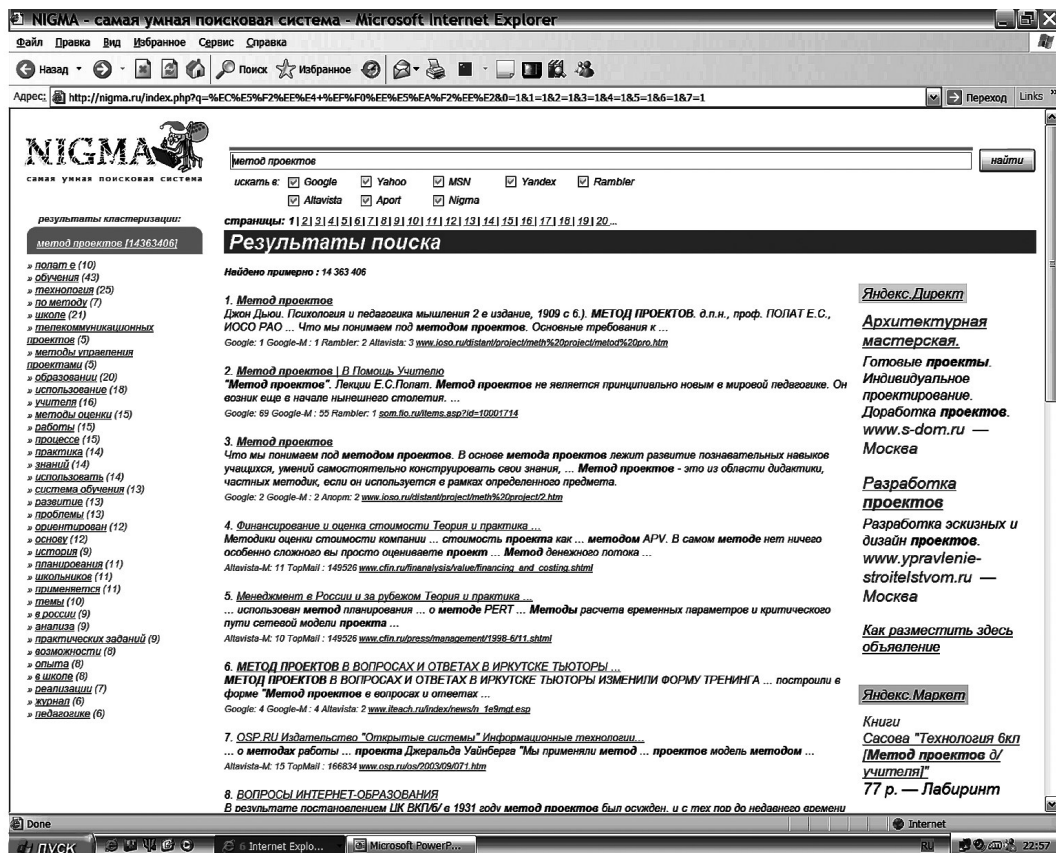


Рис. 1. Кластерный анализ запроса «метод проектов» в системе Nigma

ров. Как правило системы кластерного поиска используют результаты нескольких поисковых машин (т.е. являются метапоисковыми).

- Системы кластерного поиска предлагают пользователю, как минимум, два списка результатов поиска:

- обычный список;
- список кластеров — ключевых слов, по которым сформированы раз-
делы.

- Список кластеров наиболее час-
то представляется в виде дерева разде-

лов, выбор раздела приводит к отобра-
жению списка соответствующих ссылок.

- Многие системы кластерного
поиска дают возможность предвари-
тельного просмотра найденной ин-
формации.

Приведем адреса и краткое опи-
сание некоторых систем кластерного
поиска:

— интерфейс содержит логиче-
ские операторы, позволяет осущест-
вить предварительный просмотр со-
держимого сайта. Поиск можно осу-

ществить в следующих разделах: The Web; Downloads; Dictionary; Headline News; Thesaurus; Images; More; предусматривается расширенный поиск;

— важной особенностью является представление результатов поиска в виде карты поиска и списка кластеров. К сожалению, не отображает символы кириллицы;

— метапоисковая система, одна из первых, использующих кластеризацию. Система строит дерево кластеров, в котором размещены не только ключевые слова, но и типы документов (статьи, новости). Если в кластерах опять встречаются повторяющиеся комбинации, то создаются подкластеры. Корректно работает с запросами, написанными кириллицей;

— наиболее важными особенностями этой системы являются возможность настраивать список поисковых машин, а также то, что результат поиска содержит не только название, ссылку, но и уменьшенное отображение страницы. Это также немаловажно для быстрой оценки результатов поиска;

— один из немногих сервисов, который в полученном списке ссылок предлагает ссылку на страницу соответствующей поисковой машины с её результатами поиска. Предлагает предварительный просмотр. Корректно работает с кириллицей;

— разработана студентами и сотрудниками МГУ им. Ломоносова, используется для поиска наиболее популярные поисковые машины Google, Yahoo, MSN, Yandex, Rambler. Проект сравнительно новый, динамично развивается (в частности, улучшается механизм кластеризации информации). Первая отечественная разработка.

К сожалению, ни одна система не включает все поисковые машины.

На данный момент времени существует более десяти систем кластерного поиска информации. Выбор той или иной системы во многом определяется задачами и предпочтениями пользователей. Например, для поиска в русскоязычной зоне Интернета имеет смысл использовать систему Nigma, но она не предоставляет возможности предварительного просмотра найденной информации. Такая возможность предоставляется, например, системой KillerInfo, которая корректно работает с запросами на русском языке, но предоставляет гораздо меньший выбор ресурсов.

Что даёт кластерный поиск в преподавании

Преподавание на курсах повышения квалификации образовательных технологий «Метод проектов» и ТОГИС предусматривает проведение со слушателями Интернет-практикумов, во время которых слушатели должны найти в Интернете информацию по определённой тематике (чаще всего связанной с профессиональными интересами слушателей). Наш опыт преподавания на курсах повышения квалификации работников образования в АПКППРО показывает, что использование кластерного поиска уместно как во время подготовки к занятию, так и во время его проведения. Рассмотрим использование кластерного поиска в обоих случаях.

Во время подготовки к занятию позволяют:

1. Существенно сократить время на подготовку к занятию (часть ресурсов можно исключить из рассмотрения, а остальные ресурсы уже предварительно сгруппированы).

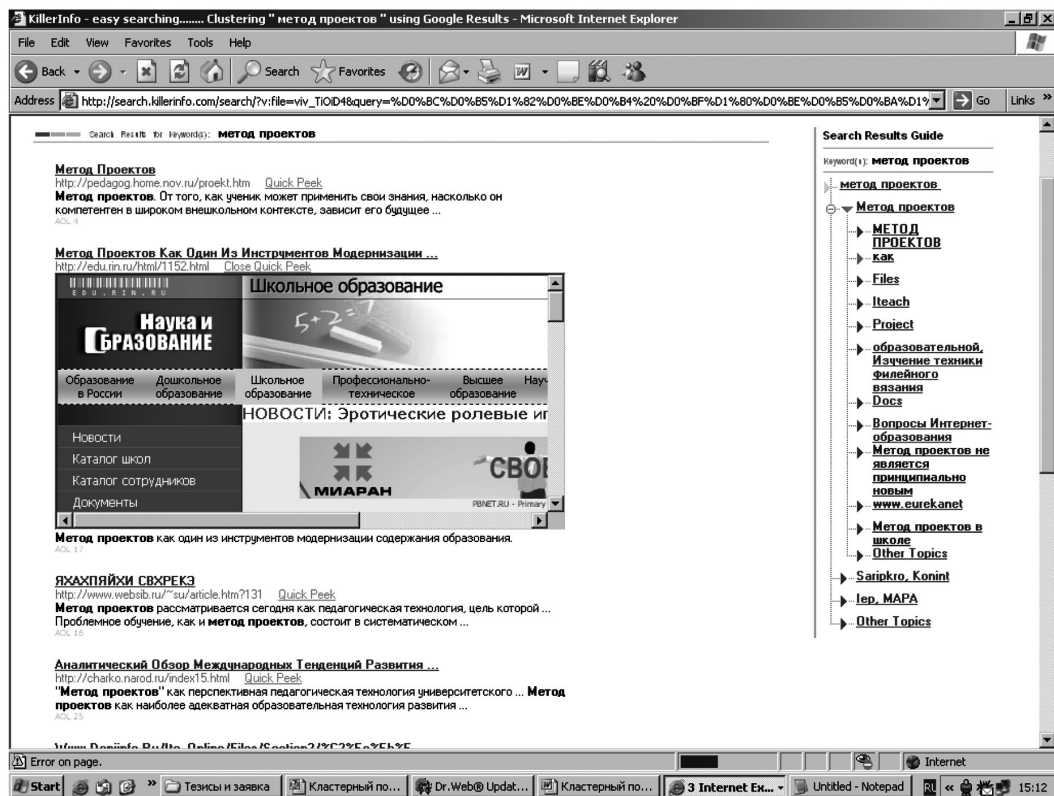


Рис. 2. Пример поиска с помощью системы KillerInfo

2. Достаточно быстро оценить ресурсы по выбранной тематике (за счёт ключевых слов, по которым сформированы кластеры).

3. Оценить результативность поиска разных поисковых машин (многие системы дают ссылки на поисковые машины).

4. Отобрать дополнительные ключевые слова для составления запроса к поисковой машине (за счёт кластеризации результатов поиска).

5. Заранее оценить вероятность попадания на ресурсы сомнительного

содержания (названия кластеров и предварительный просмотр).

Рассмотрим пример поиска при помощи системы кластерного поиска по запросу «метод проектов» (рис. 2).

Список кластеров позволяет не просматривать те ресурсы, которые объединены в кластеры с ключевыми словами, не соответствующими тематике разрабатываемого занятия.

Рассмотрев список кластеров, можно отобрать ключевые слова для уточняющего запроса, например, запрос «метод проектов», уточняющий

запрос «метод проектов в школе». Ключевые слова, по которым сформированы кластеры, могут также дать общее впечатление о представлении интересующей проблемы в сети Интернет. По поисковому запросу «метод проектов» система Nigma даёт список более чем из тридцати кластеров, среди них не соответствуют тематике около пяти (рис. 1), т.е. мы можем выделить около тридцати разделов в выбранной тематике.

Большинство систем кластерного поиска при отображении ссылки на найденный ресурс указывают и поисковые машины, которыми этот ресурс найден. Эта возможность позволяет подобрать наиболее результативную поисковую машину в том случае, когда на занятии предполагается использовать именно поисковую машину без возможностей кластеризации.

Возможность предварительного просмотра (или представление скриншотов) сильно облегчает просмотр ресурсов на предмет вероятности попадания на ресурсы сомнительного содержания. Как видно на приведённом примере, вполне серьёзный педагогический сайт имеет ссылку на ресурс, посвящённый эротике. Очевидно, что этот сайт можно использовать при занятиях со взрослыми людьми и не рекомендуется для занятий со школьниками.

Таким образом, использование кластерного поиска позволяет сократить время на подготовку к занятию и одновременно предложить слушателям более разнообразную тематику занятия.

Во время занятия:

- Появилась возможность быстро найти и отобрать информацию или составить общее представление о по-

ставленной проблеме, что существенно экономит время занятия.

- Даже неопытные слушатели смогли быстро найти и выбрать особенно интересную им информацию и были включены в активную работу. Это особенно важно для повышения мотивации при освоении информационных технологий как для слушателей курсов повышения квалификации, так и для студентов гуманитарных специальностей. Фактически освоение и уверенное применение информационных технологий становится побочным эффектом изучения материалов по тематике, интересной для обучающихся.

- Сильно ускорилась и упростилась работа по поиску и отбору нетривиальной информации, что позволило снять у неопытных пользователей психологический страх перед использованием Интернета.

- Использование кластерного поиска при работе в рамках «Метода проектов» в силу наглядности представления найденной информации имеет в качестве побочного результата освоение учащимися принципов эффективного составления запроса.

Дополнительным положительным эффектом от использования систем кластерного поиска как при подготовке к занятиям или урокам с использованием поиска информации в сети Интернет, так и при их проведении является экономия трафика, что очень важно при небольшой пропускной способности или сильной загруженности сети.

План прохождения темы

Преподавание на курсах повышения квалификации образовательных технологий «Метод проектов» и

ТОГИС предусматривает проведение со слушателями Интернет-практикумов, во время которых слушатели должны найти в Интернете информацию по определённой тематике (чаще всего, связанной с профессиональными интересами слушателей). Наш опыт преподавания на курсах повышения квалификации работников образования в АПК и ППРО показывает, что далеко не все слушатели владеют навыками поиска информации в сети Интернет, а часто и навыками работы на персональном компьютере. Поэтому в программу курсов включается модуль «Поиск информации в сети Интернет». Рассмотрим фрагмент этого курса, включающий кластерный поиск. Предложенный фрагмент состоит из двух занятий по 4 академических часа.

Первое занятие.

Цель занятия: познакомить слушателей с основными понятиями и навыками, необходимыми для работы в сети Интернет.

1. Общие сведения об Интернете, основные понятия (Интернет, гипертекстовый документ, сайт «Всемирная паутина», размещение данных в сети Интернет, типы данных в сети, стратегии поиска информации). Проводится в лекционной форме с обязательным сопровождением-презентацией (1 ч).

2. Основы работы с Интернет-обозревателями на примере одной произвольно выбранной поисковой системы (Рамблер, Яндекс, Mail.ru, Google и т.п.). Практическое занятие с консультативной помощью преподавателя (3 ч).

1. Запуск обозревателя. Основы интерфейса пользователя Интернет-обозревателей.

2. Выход на главную страницу поисковой системы: слушатели вводят заданный адрес, например, <http://www.rambler.ru>

3. Составление учащимися произвольного запроса по тематике курсов. Поиск по ключевым словам.

4. Гиперссылки: переход по гиперссылке и возврат.

5. Оценка учащимися результативности поиска.

6. Сохранение результатов.

Это занятие позволяет усвоить основные навыки работы с поисковыми машинами и навигации в Интернете. В конце занятия имеет смысл показать поиск в найденном и другие дополнительные возможности поиска, особенно для слушателей, уверенно владеющих персональным компьютером. В качестве формы контроля на этом занятии можно использовать устный опрос.

Второе занятие. Поиск по заданной тематике при помощи систем кластерного поиска. Тема задаётся преподавателем в рамках общей тематики курсов или выбирается слушателем как название своего проекта.

1. Краткий обзор метапоисковых систем и систем с кластерным анализом результатов поиска. Проводится в лекционной форме с обязательным сопровождением презентацией (1 ч)

2. Поиск информации в сети Интернет при помощи систем Nigma и/или Killerinfo. Практическая работа с консультативной помощью преподавателя.

- Составление первоначально запроса.

- Оценка учащимися результативности поиска и полученных кластеров.

- Переход по гиперссылке в кластер, который кажется наиболее инте-

ресным. Таким образом реализуется уточнение запроса.

- Просмотр найденных ресурсов, отбор, оценка и сохранение информации.

3. Составление отчёта (контроль).

4. Представление отчётов и обсуждение их в группе. Наиболее удобная форма представления отчёта — презентация.

В результате работы на втором занятии слушатели уверенно овладевают навыками: составления первичного запроса; отбора, оценки и сохранения информации.

Также слушатели получают первичные навыки составления уточняющего запроса. После выполнения заданий и обсуждения отчётов снимается психологический страх перед использованием Интернета.

В зависимости от объёма программы и уровня начальной подготовки слушателей представленные занятия можно объединить в одно или, наоборот, дополнить и расширить. Возможные темы следующих занятий:

- Работа с тематическими каталогами Интернет ресурсов.

- Использование поисковых систем общего назначения. Языки запроса, опции расширенного поиска.

- «Интернет-серфинг» — переход по ссылкам, как дополнение к рассмотренным методам поиска.

Этот фрагмент модуля «Поиск информации в сети Интернет» можно рассматривать и вне контекста курсов повышения квалификации по проблеме образовательных технологий, а как самостоятельный курс по поиску информации в сети Интернет для школьников или студентов гуманитарного профиля.

В заключение хотелось бы обратить внимание на следующие моменты:

- кластерный поиск, которому посвящена большая часть статьи, не панацея, а ещё один подход с решением проблемы поиска в учебном процессе, наиболее удобный при необходимости быстро найти и оценить информацию по некоторой теме;

- в некоторых ситуациях использование кластерных и метапоисковых систем излишне (например, известно название искомой статьи или есть точная цитата);

- надо учитывать, что большинство поисковых машин, используемых системами кластерного поиска, не работают с русской морфологией;

- результаты поиска зависят от точности запроса. Надо учить слушателей составлять запрос. В кластерных системах поиска это делать легче, поскольку представление результатов намного более наглядно.