

*Паньгин Андрей Александрович,
Поздняков Сергей Николаевич,
Посов Илья Александрович,
Рыбин Сергей Витальевич,
Степулёнок Дмитрий Олегович*

ВОЗМОЖНОСТИ ИНФОРМАЦИОННО-ПОИСКОВОЙ СИСТЕМЫ УЧЕБНО-МЕТОДИЧЕСКИХ МАТЕРИАЛОВ SCISEARCH

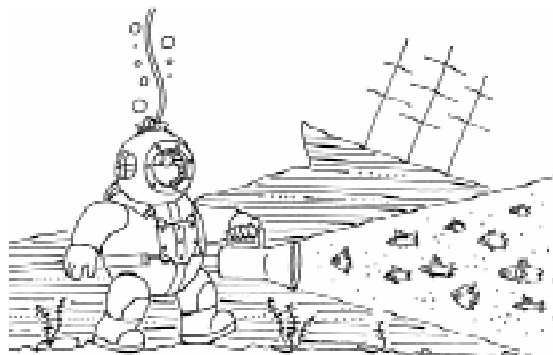
Информационно-поисковая система учебно-методических материалов SCISEARCH – это специализированная система поиска, использующая в своей работе описание ресурса в терминах ключевых и антиключевых слов и анализ результатов на основе существенных для образования признаков.

Традиционно поисковые средства для нахождения интересующей информации разделяют на две категории: каталоги и поисковые машины. Каталоги обычно предусматривают строгое разделение веб-сайтов по рубрикам и пополняются самими владельцами ресурсов, за счет чего оказываются более удобными для направленного поиска в рамках одной тематики, в то время как поисковые машины занимаются пополнением своей базы ресурсов самостоятельно, вследствие чего оказываются более приспособленными к условиям постоянно изменяющегося потока информации в сети, но лишенными возможности фильтровать и структурировать веб-ресурсы. Однако и те, и другие универсальные поисковые средства плохо приспособлены для узконаправленного специализированного поиска с повышенными требованиями к качеству информации, что, в частности, характерно для поиска учебных материалов в заданной предметной области преподавателями школ и вузов. В этой ситуации необходимо применять специальные инструменты поиска, нацеленные на увеличение его эффективности, при условии, что тематика, характер требований, предъявляемых к

искомой информации, и потенциальные пользователи заранее известны.

Система SCISEARCH является именно таким специализированным инструментом поиска, сочетающим в себе возможности организации динамического каталога с возможностями специализированной поисковой системы, основанной на дополнительном анализе ресурсов, полученных предварительными запросами к поисковым системам общего назначения, таким как, например, Yandex, Rambler и Google.

Особенностью каталога-классификатора является его способность к динамическому обновлению, поскольку его наполнение составляют не столько ссылки, сколько алгоритмы для извлечения этих ссылок посредством данной системы и существующих универсальных поисковых машин. Каталог-классификатор построен на основе стандартных рубрикаторов, используемых на порталах, поддерживающих обучение мате-



...необходимо применять специальные инструменты поиска...

матике. Таким образом, систему можно применить и к каталогизации физических ресурсов сети.

Полезным свойством системы является возможность расширять и уточнять динамические каталоги, создавать новые, в том числе жанровые каталоги для одного множества ресурсов.

Перейдем теперь к задаче поиска. Очевидно, что наиболее полно описать требования пользователя, предъявляемые к искомым материалам, может только предложение на естественном языке, однако лишь немногие из этих предложений могут быть достаточно точно выражены в терминах количественных и качественных оценок произвольного Интернет-ресурса. В системе SciSearch фразы человека анализируются на вхождение определенных ключевых слов и конструкций, влияющих на значение некоторых параметров поиска и приоритет сортировки найденных материалов. Такими параметрами могут служить, к примеру, наличие в содержании Интернет-ресурса большого числа иллюстраций или ссылок на файлы для закачивания, либо опция, отвечающая за то, будут ли исключены из результатов поиска задачи без решений или теоремы без доказательств.

Разработчики системы специально предоставили пользователю средства постановки задачи в заведомо неудобной для восприятия компьютером форме. Таким образом, пользователь системы имеет возможность сформулировать задачу поиска максимально подробно, с расчетом на то, что в

дальнейшем система будет развиваться, пополняясь новыми аналитическими правилами, и тот же самый вопрос в дальнейшем может быть интерпретирован несколько иначе, с учетом новых знаний, извлеченных из текста вопроса. Кроме того, программы обработки запросов в самих поисковых системах тоже постоянно совершенствуются, что, таким образом, приводит к улучшению результатов поиска без каких-либо существенных изменений в программном обеспечении SciSearch.

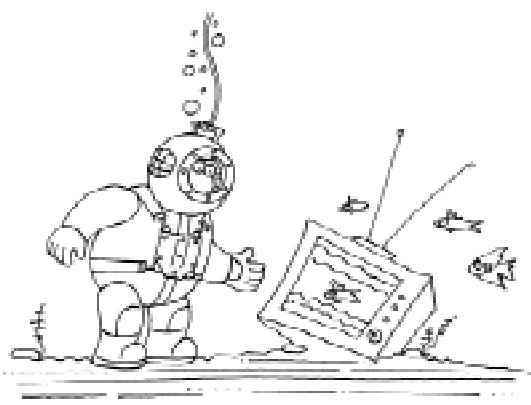
В случае если в процессе поиска система не нашла ресурсы, соответствующие запросу, либо количество ссылок, которое она возвратила, слишком мало, система SciSearch формирует повторный запрос, уменьшая количество слов на единицу. Из запроса исключается самое малоинформативное слово с точки зрения конкретизации поисковой задачи. Так продолжается до тех пор, пока мы не получим более 5 ссылок, либо пока запрос не сократится до одного слова. Разумеется, ссылки, полученные в результате предыдущих запросов, также участвуют в дальнейшем анализе, причем с большим приоритетом, чем остальные.

Такой анализ позволяет отбраковать случайные ресурсы, такие как рекламные материалы или обсуждение в форуме. Найденные по стандартным алгоритмам поиска с помощью поисковых систем общего назначения ресурсы проходят далее фильтрацию на основе контекста, определяющего принадлежность найденного сайта к содержательным ресурсам сети Интернет.

Система имеет два режима доступа: один для экспертов, создающих и поддерживающих каталоги, второй – для пользователей образовательных сайтов. Внешний вид интерфейса системы в экспертном режиме представлен на рис. 1.

Эксперт может запретить появление некоторых ресурсов, ставя соответствующие пометки. Система позволяет давать экспертную оценку (интегрируя оценки разных экспертов) и классифицировать ресурсы по трём основным категориям:

- рекомендуемые ресурсы,
- дополнительные ресурсы,
- не рекомендованные ресурсы.



Такой анализ позволяет отбраковать случайные ресурсы...

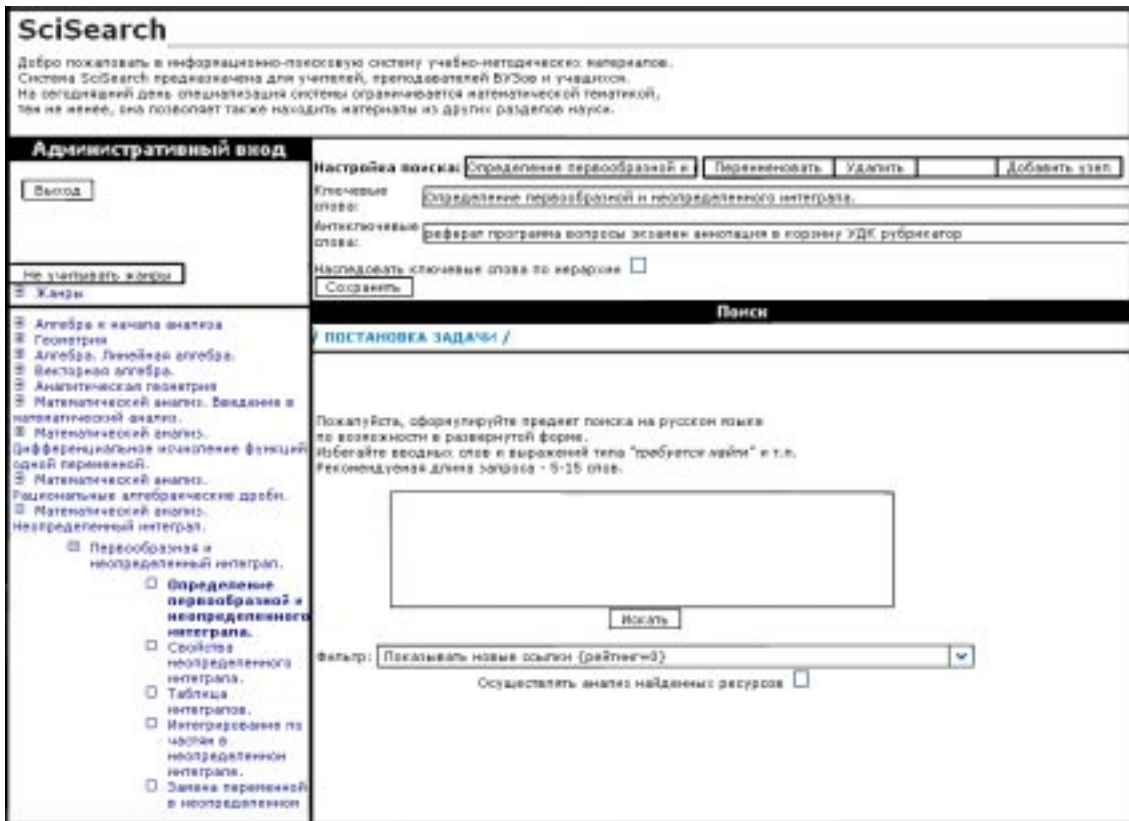


Рис. 1

Последнее особенно важно в связи с тем, что многие сайты, интересные, но не защищённые достаточным образом, часто «взламываются» и на них размещается посторонняя, а подчас и опасная информация (лож-

ные ссылки, вирусы и пр.). Пример работы эксперта представлен на рис. 2.

Отдельный доступ системы для обычных пользователей позволяет отбирать полезные ресурсы в разумном количестве



Рис. 2

(один-два десятка) по их качественному описанию и предоставлять отдельный доступ ко всем отмеченным выше категориям (рекомендованные ресурсы, дополнительные ресурсы, все ресурсы). При этом по желанию пользователя может быть осуществлён анализ отобранных системой ресурсов; возможность отключать или включать эту опцию полезна, так как анализ ресурсов требует просмотра соответствующих сайтов и требует времени, использование рекомендованных ресурсов может производиться без их анализа.

Рассмотрим список (далеко не полный) свойств документа, принимаемых во внимание системой SciSearch для оценки найденного ресурса:

- Объем текста – характеризует, сколько текстовой информации содержит документ.
- Принадлежность ресурса к области тематики. Ресурсы из других областей отбрасываются.
- Уровень сложности текста.
- Количество баннеров и рекламы. Страницы с малым количеством текста, но большим числом баннеров считаются неинформативными.
- Надежность источника.
- Наличие ссылок на файлы для скачивания.
- Сопровождение страницы сайта демонстрационными программами.
- Ресурс представляет собой форум.
- Коммерческая направленность ресурса. Ссылки на такие ресурсы исключаются.
- Неинформативные листинги (сюда относятся список вопросов, программа курса, план урока, оглавление книги, список литературы и т. п.).
- Принадлежность крупному ресурсу.

Литература

1. Информационные технологии в системе непрерывного педагогического образования (проблемы методологии и теории). Монография. Под общей ред. проф. В.А. Извозчикова. СПб.: «Образование», 1996.
2. Башмаков М.И., Поздняков С.Н., Резник Н.А. Информационная среда обучения. СПб.: «Свет», 1997.

Исходя из свойств описанной системы, можно рекомендовать ее для применения в следующих областях:

- Система может быть рекомендована преподавателям, не имеющим большого опыта в аккуратной формулировке запросов и его последовательного уточнения. Система позволит им более эффективно пользоваться ресурсами.
- Использование системы в учебном процессе педвуза. Система даст хорошую практику будущим педагогам в создании динамических каталогов полезных ресурсов, позволит создать тематический сборник распределенных в сети материалов.
- Система может быть полезной на сайтах образовательных центров информационных технологий, так как с ее помощью можно организовать совместную работу педагогов по целесообразному использованию сети Интернет: поиску и каталогизации полезных ресурсов.
- Система может оказаться полезной в системах дистанционного обучения, так как позволит естественным образом подключать распределённые ресурсы (блокируя случайные или малосодержательные ссылки).
- Систему можно рекомендовать методистам и экспертам, ведущим собственные каталоги, которые можно создавать непосредственно на сайте, где установлена система; это полностью соответствует тенденциям развития сети Интернет.

Работа выполнена при финансовой поддержке РГНФ в рамках проекта создания информационных систем («Информационно-поисковая система учебно-методических материалов для поддержки работы учителей школ в сети Интернет, основанная на технологии контекстных фильтров»), проект № 05-06-06271а.

3. *Поздняков С.Н.* Информационная среда как новый фактор обучения информатике // Всероссийская конференция «Математика и общество. Математическое образование на рубеже веков», Дубна, сентябрь 2000.
4. *Поздняков С.Н.* Конструирование общего информационного пространства: сервер «Петербургская Интернет-школа». Вторая международная конференция «Интернет. Общество. Личность – ИОЛ-2000. Новые информационно-педагогические технологии». 28 февраля – 3 марта 2000 г.
5. *Поздняков С.Н.* Конструирование общего информационного пространства. Материалы IX Междун. конфер.–выставки. ИТО99: 9–12 ноября. 1999 г. Тезисы докладов. М: МИФИ, 1999.
6. *Иванов С.Г., Поздняков С.Н.* Как должна быть устроена компьютерная поддержка преподавания математики. Материалы I съезда учителей и методистов Северо-Запада и СПб. Секция «Информатизация образования», 5 марта 2003. СПб., 2003.
7. *Ландэ Д.* Поиск знаний в Internet. М.: Диалектика, 2005.
8. *Браславский П.И., Вовк Е.А., Маслов М.Ю.* Фасетная организация интернет-каталога и автоматическая жанровая классификация документов. Диалог'2002. Протвино: 2002.
9. *Кормен Т., Лейзерсон Ч., Ривест Р.* Алгоритмы: построение и анализ. М.: МЦНМО, 2000.
10. *Кураленок И.Е., Некрестьянов И.С.* Оценка систем текстового поиска «Программирование», 2002, № 4.
11. *Браславский П.И.* Построение запросов к машине поиска Internet с помощью тезауруса. Материалы 3й Всероссийской научной конференции. «Электронные библиотеки», Петрозаводск: 2001.
12. *Бузикашвили Н.* Поисковое взаимодействие: мифы, реальность и базовые принципы. Материалы 6й Всероссийской научной конференции. «Электронные библиотеки», Пушкино, 2004.

***Паньгин Андрей Александрович,
Sun Microsystems, старший
инженер-программист,***

***Поздняков Сергей Николаевич,
профессор кафедры ВМ-2
СПбГЭТУ (ЛЭТИ),***

***Посов Илья Александрович,
аспирант математико-
механического факультета СПбГУ,***

***Рыбин Сергей Витальевич,
кандидат физико-математических
наук, доцент кафедры ВМ-2
СПбГЭТУ «ЛЭТИ»,***

***Степулёнок Денис Олегович,
аспирант кафедры АСОИУ
СПбГЭТУ «ЛЭТИ».***



**Наши авторы, 2007
Our authors, 2007**