

Введение

В настоящее время глобальная сеть Интернет содержит большой объем динамически изменяющейся информации и развивается весьма бурными темпами. Для того чтобы как-то упорядочить этот непрерывный поток данных, а самое главное, дать возможность пользователям Сети находить нужную информацию, были созданы специальные поисковые системы. Развитие данных систем идет по пути автоматизации информационного поиска, и главным критерием при этом является увеличение релевантности найденной информации.

На современном этапе среди проблем поиска информации в Интернете особенно выделяются следующие.

- **Временные затраты:** в результате автоматизированного поиска в Сети согласно запросу пользователь получает большое количество ссылок, просмотр которых занимает много времени, при этом даже простой выбор необходимой информации зачастую представляет собой нелегкую задачу.
- **Экономические затраты:** большие временные затраты невыгодны с экономической точки зрения для пользователей, подключенных в лимитном режиме. Кроме того, некачественные связь и линии также затрудняют работу и увеличивают трафик.
- **Неэффективность поиска** при работе группы пользователей, осуществляющих поиск информации по одинаковой тематике: каждому приходится просматривать и анализировать одну и ту же информацию, при этом тратится время для работы в Сети.
- **Нерациональность и непоследовательность действий** пользователей при поиске информации.

- Трудности при составлении запросов: пользователь часто не может строго определить цель поиска и использует нечетко определенные понятия.

Вместе с тем следует отметить недостатки существующих поисковых систем. Относительно низкая релевантность поиска в результате неучета в системах отношения пользователя к знанию и взаимодействия элементов информации. Неспособность к самообучению (неумение системы адаптироваться к потребностям пользователя и стилю его работы). Пользователю во время процедуры поиска информации необходимо определиться с темой для поиска, согласно которой затем выявить ключевые понятия поиска и их значимость.

В учебном пособии авторы раскрывают особенности поиска информации в Интернете. В рассматриваемой области можно использовать известные результаты: концепции построения поисковых систем, теоретические концепции моделей представления знаний, построения экспертных систем, универсальных алгебр, нечетких множеств и применение теории выбора. Эти вопросы освещены, в частности, в работах Г.С. Поспелова и Д.А. Поспелова [34, 39–41], Э.В. Попова [38], М.Л. Минского [33], Н. Нильсона [35], Т. Саати [43], Л. Заде [11, 12, 16], А. Кофмана [22] и др. Теория поиска информации представлена в работах М.Г. Крейнса [47], С.В. Кузнецова [32], Г.С. Осипова [37] и др.

Использование формального подхода к построению модели предметной области для информационного поиска и описание предметной области (ПрО) с помощью тем является малоизученной проблемой. Решение этой проблемы позволит упростить не только идентификацию свойств модели предметной области, но и облегчить процесс обработки и поиска информации, необходимой пользователю. Математическим аппаратом автоматизации решения упомянутых проблем являются семантические сети, универсальные алгебры и нечеткие множества, позволяющие автоматизировать процесс поиска информации, а также процесс построения и обработки модели предметной области.

Эта книга посвящена изучению основных концепций построения и обработки модели предметной области для систем

обработки знаний, ориентированных на поиск информации, а также подход к созданию интеллектуальных поисковых систем в Интернете. Исследования авторов, которые применяются в книге для решения обозначенных вопросов, базируются на теории универсальных алгебр, теории множеств, теории нечетких множеств, теории принятия решений, методов структурного и объектно-ориентированного программирования.

Назначение и особенности книги

Данная книга представляет собой учебное пособие, включающее в себя разделы по разработке и анализу интеллектуальных поисковых программ в Интернете на основе универсальных алгебр с различными подходами к вопросам интеллектуального анализа данных, а также примеры практического проектирования таких систем. Разделы содержат определения и понятия, приведены требования, предъявляемые к информации в таких системах. Отражена специфика анализа данных в условиях неопределенности. Приведены основные моменты теории нечетких множеств, необходимые для решения задач интеллектуального анализа на основе теории семантических сетей. Книга содержит много примеров и особенностей реализации в предметной области.

В области построения интеллектуальных поисковых систем на основе нечеткой логики целесообразно использовать: определение формальной модели построения предметной области с применением темы для поиска, используемой при создании интеллектуальной поисковой системы; создание моделей пользователей, используемых при поиске информации, и их анализ; описание новых механизмов логического вывода, используемых при поиске информации в Интернете, и основ методики извлечения знаний для формирования персональной модели знаний пользователя с целью повышения релевантности поиска; формирование проектных решений по программной реализации компонентов интеллектуальной поисковой системы.

Учебное пособие основано на лекционном курсе по дисциплинам: «Предметно-ориентированные информационные системы», «Математические методы и модели принятия про-

ектных решений», «Методы оптимизации и теория принятий решений», «Системы поддержки принятия решений в медицине», «Интеллектуальная поддержка принятия решений в системах медицинского назначения», «Проектирование рекомендательных систем», «Проектирование цифровых систем медико-биологического назначения».

Учебное пособие преследует следующие цели.

- Ознакомить читателя с базовыми понятиями в теории проектирования интеллектуальных поисковых систем и пояснить терминологию в этой области.
- Ознакомить с различными подходами к реализации поисковых систем и областями их применения.
- Рассмотреть основные этапы разработки интеллектуальных поисковых систем, а также применение различных теоретических моделей на каждом из этапов в условиях неопределенности и неполноты исходной информации.
- Сформировать навыки построения и оценки качества полученной модели знаний (модели предметной области).
- Ознакомить с конкретными примерами реализации поисковых информационных систем для заданной предметной области.

Структура книги

Учебное пособие содержит 12 разделов. В тексте имеется теоретическая информация по теме, приведено большое количество примеров и конкретных реализаций в предметной области. Благодаря этим свойствам каждый из разделов может рассматриваться как конспект лекции по определенной теме. В конце каждого раздела приведены основные выводы.

Первый раздел посвящен описанию проблем и задач поиска информации в Интернете. Рассмотрены этапы поискового процесса. Приведена общая архитектура механизма поиска информации, описаны ее компоненты.

Второй раздел посвящен анализу существующих поисковых систем, их преимуществ и недостатков. Приведены виды поисковых систем. Представлены общие недостатки существующих поисковых систем.

В третьем разделе приводится анализ направлений развития интеллектуального поиска в Интернете, приведен пример

существующих интеллектуальных поисковых систем и выявлены их узкие места. Сформулированы требования к проектируемой системе с учетом преимуществ и недостатков существующих поисковых систем. Описано применение интеллектуальных поисковых систем, проанализированы проблемы их построения. Описаны функции экспертных систем. Рассмотрены основные способы построения экспертных систем как элемента интеллектуальной поисковой системы.

В четвертом разделе подробно описаны основные модели представления знаний, используемые при построении систем управления знаниями, выявлены их достоинства и недостатки. Приведены свойства знаний. Сделан вывод, что наиболее универсальной является модель представления знаний в виде семантической сети.

В пятом разделе приводится определение универсальной алгебры, используемой для формализации представления модели знаний. Определены нечеткие множества для обработки неопределенности и приблизительных рассуждений при создании модели знаний, описаны достоинства использования нечетких множеств в обработке модели знаний, а также определены нечеткие объекты для описания реальных предметных областей.

В шестом разделе представлено описание семантической сети для построения модели предметной области. Предложено оригинальное построение семантической сети интеллектуальной поисковой системы с использованием универсальной алгебры и нечетких множеств, позволяющее установить взаимосвязь тем для поиска и найденных документов в модели предметной области. Определены операции и отношения между семантическими сетями и между элементами семантической сети интеллектуальной информационно-поисковой системы (ИПС), дающие возможность создавать модели предметных областей, описываемых в терминах семантической сети, разбивать и объединять построенные модели предметных областей.

Седьмой раздел посвящен блочной структуре построения интеллектуальной информационно-поисковой системы на основе семантических сетей. Описан механизм логического вы-

вода, интеллектуального поиска информации, который на основе модели пользователя и модели действий пользователя отыскивает множество наиболее релевантных документов согласно запросу конкретного пользователя и выдает их в ранжированном порядке.

В восьмом разделе приведено описание механизма формирования сетевой модели пользователя, архитектура модели пользователя для интеллектуальной поисковой системы. Описаны основные типы моделей пользователя. Предложено построение модели действий пользователя для интеллектуальной поисковой системы. Описаны две интерпретации модели поведения пользователя: поиск документов и знакомство с предметной областью. Представлена их структура и принципы организации. Модель действий пользователя позволяет увеличить эффективность поиска информации для пользователя, процесса его ознакомления с предметной областью и упростить способ ее описания в интеллектуальных поисковых системах.

В девятом разделе рассмотрен расширенный и обычный способ поиска информации с использованием модели пользователя на основе применения критериально-экстремизационных механизмов выбора и оценочной функции, позволяющие с высокой достоверностью находить наиболее релевантные документы согласно запросу пользователя, а также осуществлять их ранжирование. Описан механизм формирования групп документов согласно теме для поиска, выбранной пользователем, приведен алгоритм нахождения документа согласно теме поиска из модели предметной области, дающий возможности образовывать ранжированные группы документов по тематической направленности.

В разделах 10, 11 и 12 приведен пример системы, которая разработана и внедрена в деятельность нескольких учреждений. Система была спроектирована и реализована на основе изложенной теории, алгоритмов и методов построения модели пользователя и модели действия пользователя на основе нечетких семантических сетей. Описаны требования, предъявляемые к построению интеллектуальной поисковой системы. Приведена архитектура построения интеллектуальной

поисковой системы, определены ее составные элементы. Представлена методика описания предметной области на основе темы поиска и раскрывающего ее списка понятий.

Показана методика извлечения знаний на основе многослойных репертуарных решеток Келли, позволяющая наиболее адекватно представить отношения понятий в предметной области интеллектуальной поисковой системы. Описан способ построения подсистемы поиска информации в Интернете и особенности работы ИПС в Сети со стандартными поисковыми машинами. Приведена схема построения базы данных и описаны ее таблицы. Описаны этапы построения поисковой базы знаний. Представлено построение интерфейса для извлечения знаний эксперта о близости понятий с использованием многослойных репертуарных решеток Келли. Приведена схема работы пользователя с программной реализацией интеллектуальной поисковой системы.

В конце книги приводится библиографический список публикаций по теме интеллектуальных поисковых систем. Читатель в этом списке найдет книги на интересующую его тему.