

Профессиональные электронные энциклопедии в системе VEDA

Костенко К.И.
КубГУ, Краснодар, Россия
kostenko@kubsu.ru

Левицкий Б.Е.
КубГУ, Краснодар, Россия
bel@kubsu.ru

Назаренко С.В.
КубГУ, Краснодар, Россия,
nazarenko@kubsu.ru

Аннотация

Система VEDA представляет собой универсальную программную оболочку, позволяющую создавать и использовать системы документов, имеющих иерархическую структуру, с основной формой доступа из глобальных компьютерных сетей. Она ориентирована на применение технологий обработки знаний для разных этапов работы с системой: проектирования структур документов и типов значений отдельных полей, создания конкретных документов, построения запросов к базе размещенных материалов и использования содержащейся в ней информации. Одно из применений VEDA связано с созданием профессиональных электронных энциклопедий для конкретных предметных областей.

Электронные энциклопедии в Интернет - это форма представления полей профессиональных знаний специалистов, доступных как для редактирования, так и практического использования. Фрагменты полей знаний задаются описаниями в специальных языках представления знаний и размещаются в отдельных документах энциклопедий или их частях. Семантические связи различных документов в энциклопедии и их частей представляются именованными ссылками. Совокупность статей разной структуры вместе со связями образует структуру профессионального поля знаний, допускающего автоматическую обработку.

Имеется несколько вариантов использования энциклопедий:

- индивидуальной информационной среды профессионалов;
- обучающей среды при подготовке специалистов;
- справочных материалов по предметной области для специалистов в других областях деятельности;
- руководству получения консультаций для

принятия решений классификации и прогнозирования;

- интеграции знаний групп специалистов одной области.

Достоинством электронных энциклопедий является возможность непрерывного авторского сопровождения, представления разных мнений и знаний разного уровня, взаимодействие авторов с пользователями, удаленный сбор и размещение информации.

Специальная форма применения электронных энциклопедий связана с использованием механизмов вывода и других методов обработки знаний специалистов целью автоматизации процесса решения задач или получения ответов на вопросы пользователей. В этом случае пользователи системы имеют возможность формулировать свои задачи и цели и получать комментированные решения, получаемые с помощью обработки знаний, содержащихся в интеллектуальных документах.

Информация, представляемая в системе VEDA размещается в структурах документального и поискового пространств. Документальные пространства материалов в составе энциклопедий реализованы в отдельной подсистеме хранилища документов и содержат системы структурированных документов, размещенных в базах данных вместе с семантическими связями между документами и их частями, а также формами документов и шаблонами структур значений отдельных полей.

Поисковые пространства хранилищ документов предназначены для размещения поисковых заданий в форме паспортов информационных потребностей пользователей вместе с поисковыми образцами и описаниями дисциплины информационного обслуживания, которые обрабатываются потоком входных документов электронных энциклопедий на протяжении времени действия заданий.

1 ПРОСТРАНСТВО ДОКУМЕНТОВ

Документальное пространство всякой электронной энциклопедии образуется содержимым размещаемых в ней документов. Структура такого пространства представляется системой форм, используемых для создания документов, а также семантическими связями документов и их фрагментов.

© Вторая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
26-28 сентября 2000г., Протвино

2.1 Организация представлений документов

Создание документа в VEDA выполняется на основе связанно предварительно созданной формы и состоит в построении общего описания документа и последующего занесения содержимого документа в поля значений формы. Всякая форма документов является либо элементарной и состоит из одного поля, либо задается последовательностью, состоящей из: отдельных составных форм, классов форм, серий однотипных форм и ссылок. Элементарные формы представляются множествами значений, которые могут в них заноситься. Отдельная форма в составе формы может быть произвольной созданной ранее формой и образует узел иерархии, сопоставляемой всей форме. Класс форм это набор форм, для которого предусмотрена возможность выбора лишь одной из них в составе конкретных документов, формы которых содержат данный класс. Серия в структуре документа определяется формой, которая представляется в конкретных документах конечным числом своих заполненных копий. Допускаются серии на основе классов форм. Поля ссылок предназначены для формирования смысловых связей между документами и их частями.

Различные варианты работы с документами порождают потребность в использовании нескольких способов визуализации. Например, если документ в составе хранилища содержит профессиональную базу знаний, то визуализация такого документа для автора и пользователя могут быть разными. Визуализация для автора связана с обеспечением удобства редактирования поля профессиональных знаний, представленного в документе, а визуализация пользователя связана с возможностью отображения сведений инициализации задачи и результатов ее решения, выводом комментариев к полученным решениям.

Способы визуализация документов представляются с помощью сценариев визуализации, в которых декларируются интерфейсы отображения и работы с документами.

1.2 Классификация документов

Для документов, размещаемых в составе энциклопедий, применяются синтаксическая и семантическая классификации.

Синтаксическая классификация документов, формируется автоматически и основывается на иерархии форм, образованной отношениями включения специальных множеств, определяемыми для структур форм и типов значений полей.

Семантическая классификация состоит в построении иерархий классов документов, объединяемых смысловой общностью содержащейся в них

информации. При этом допускаются разные семантические классификации одних и тех же документов в составе энциклопедии, что делает семантическую классификацию многозначной. Например, возможны классификации по потребителям информации и областям использования материалов энциклопедии.

Для классов документов, используемых в составе энциклопедий, создаются специальные описания, содержащие сведения о предназначении, способах использования, особенностях ведения документов таких классов.

Возможность автоматического отнесения документов к семантическим классам поддерживается с помощью систем логических и структурных условий, а также механизмов сопоставления документов классам на основе оценки степени их соответствия описаниям классов.

1.3 Типы значений полей документов

Значения элементарных полей форм документов могут быть как стандартных, так и конструируемых типов. Последние создаются в виде комбинаций из логических и структурных условий на значения, помещаемые в отдельные поля. В первом случае условие представляется логической формулой, составленной из стандартных и специальных сравнений значений функциональных выражений, соединяемых логическими связками. Такие условия позволяют выделять части стандартных типов, для которых они выполняются. К стандартным сравнениям в составе условий относятся отношения равенства-неравенства, а также отношения принадлежности и включения. Специальные сравнения являются специфическими для конкретных областей деятельности и включаются в систему сравнений пользователя посредством описания свойств и задания процедур установления выполнимости таких сравнений.

Для задания структурных условий используются:

- простые шаблоны, представляющие собой регулярные выражения, составленные из основных и вспомогательных символов, символов переменных, а также операций сцепления, выбора и итерации;
- сложные шаблоны, представляющие линейную запись параметризованных параллельно-последовательных семантических сетей (иерархических s-p сетей), включающих итеративные и варианты части, параметры которых могут быть простыми шаблонами или принимать значения из заданных словарей.

Такие шаблоны делают возможным структуризацию, как отдельных предложений, так и сверхфразовых единств текстов в полях документов, для которых появляется возможность формальной работы с

содержанием текстов, создаваемых на основе прописанной системы используемых конструкций предложений. Заполнение полей документов, значения которых должны соответствовать структурным условиям, осуществляются с помощью интерфейса визуального конструирования значений.

1.4 Семантические связи между документами

Зависимости документов, размещенных в электронных энциклопедиях, реализуются конструкциями двух видов:

- именованными ссылками из документов на связанные с ними документы или фрагменты документов;
- семантическими сетями, представляющими фрагменты поля знаний предметной области внутри значений, помещаемых в отдельные поля документов.

Связи первого типа формируются при создании документов, определяют структуру документального пространства и обеспечивают возможность перехода между документами при работе с электронной коллекцией. Связи второго типа делают возможным представление семантики информации в отдельных полях.

Семантическая сеть системы документов представляет смысловые зависимости содержащихся в них фрагментов информации. Она формируется как результат анализа содержимого документов и связывает как документы, так и их компоненты. Описания отношений между документами и их частями представляются системами специальных правил, с помощью которых возможно автоматическое связывание содержимого электронных коллекций в семантические сети. В VEDA предусмотрена возможность перемещения по семантической сети связей документов на основе описания путей и окружения документов с помощью регулярных выражений, а также просмотра отдельных материалов вместе с их окружением.

1.5 Жизненные циклы документов

Для документов в коллекциях имеется возможность исполнения сценариев времени существования. Сценарии позволяют автоматизировать работу с отдельными документами или группами документов. Они представляются системами событий, связанных с наступлением заданных моментов времени, выполнением условий на документы и их окружение, а также операциями, выполняемыми над документами. События представляют собой сообщения, размещаемые и обрабатываемые в базе координатора событий. Реакции на события состоят в рассылке сообщений, проверке появления новых документов, создании,

изменении и удалении документов, выполнении других специальных действий.

2 ПОИСКОВОЕ ПРОСТРАНСТВО

Поисковое пространство VEDA состоит из двух частей: базы информационных заданий, относящихся к пользователям и семантическим классам документов, и базы логических зависимостей между запросами и их фрагментами. Первая из этих частей формирует внешнее представление поискового пространства, а вторая - используется для организации эффективной обработки системы запросов потоком документов, размещаемых в хранилище.

2.1 Пространство ордеров заявок

Потребители информации, размещаемой в системе электронных документов, могут создавать заявки на информационное обслуживание, выполняемые в течение заданного периода. Такие заявки размещаются в паспортах заказов клиентов и содержат выбранные значения параметров обслуживания и формы поискового задания. К параметрам обслуживания относятся: регулярность выполнения, форма предоставления найденных материалов, действия в нестандартных ситуациях. Поисковые формы содержат системы условий на отдельные документы в коллекции или системы связанных документов, соответствующие структурам форм документов и типам значений полей в них.

2.2 Структурные поисковые шаблоны

Построение поисковых заданий выполняется методом проектирования сверху вниз. Задание имеет имя и является либо элементарным, либо составным. Составное задание представляется в виде последовательности именованных заданий, каждое из которых имеет вес и значение обязательности выполнения. В задание дополнительно включается условие на суммарный вес выполненных подзапросов. Задание считается выполненным, если выполнены все его обязательные подзадания, а взвешенное значение суммы выполненных подзаданий удовлетворяет условию на суммарный вес.

Это позволяет рассматривать задания как особую форму конструкций, ориентированных на установление соответствия или ассоциации между ними и фрагментами документального пространства хранилища системы VEDA. При этом обеспечивается возможность оценки величины такого соответствия.

Логические условия в виде формул, состоящих из элементарных запросов и логических связей, являются частным случаем поисковых заданий. Элементарные условия это либо шаблоны структур значений полей документов, либо сравнения функциональных

выражений, составленных из обозначений операций, констант, имен полей и связей между документами.

2.3 Ассоциативный и логический поиск

Поиск информации в документальном пространстве, основанный на выделении связанных систем значений полей документов, соответствующих поисковому шаблону, использующий схемы согласования и унификации, делает возможным обработку и использование информации с помощью механизмов установления подобия, аналогии и сходства. Такие механизмы создают дополнительные возможности для анализа, как содержимого информационного пространства энциклопедий, так и процессов его обновления.

2.4 Технология распознавания связанных систем знаний

Особый вид информационных потребностей пользователей и классификаторов документов связан с выделением фрагментов семантической сети пространства документов. Для представления таких потребностей применяются поисковые формы, имеющие вид регулярных выражений, содержащих структурное описание распознаваемых фрагментов, условий на содержание отдельных документов в составе фрагментов и компонент, выделяемых в качестве результата. Такие запросы делают возможным поиск информации, на основе рассмотрения окружения документов с помощью цепочек семантических зависимостей между документами и их содержимым.

2.5 Организация обработки входного потока документов

Размещаемые впервые или исправленные документы образуют входной поток для коллекции документов энциклопедии, из которого выделяются ссылки на отдельные документы и фрагменты документального пространства, удовлетворяющие поисковым запросам пользователей. Система ссылок на документы используется для передачи сведений клиентам в соответствии с параметрами обслуживания. Процесс обработки потока размещаемых документов допускает использование унификации и оптимизируется с помощью:

- сжатых таблиц решений;
- конструируемых деревьев поиска;
- сети логических зависимостей компонент разных запросов.

2.6 Технологии использования интерактивных документов

Интерактивные документы – это документы, содержащие знания специалистов, представленные в формальных моделях, доступ к которым предполагает использование знаний для автоматического решения задач пользователя и состоит в выполнении действий, определяемых сценарием визуализации интеллектуальных документов для решения задач пользователей, включающих следующие этапы:

- общения с клиентом с целью заполнения опросных форм документа;
- обработки содержимого полей знаний документа;
- создания итогового документа и пересылка его клиенту.

На первом этапе осуществляется уточнение задачи клиента и начальных данных или условий для ее решения. Эта информация и содержимое баз профессиональных знаний обрабатываются назначенным документу механизмом вывода, результаты работы которого затем komponуются в отчет, пересылаемый пользователю.

Предусмотренная в VEDA стандартная форма представления интерактивных документов позволяет размещать следующую информацию в качестве значений полей документов: данные инициирования задач пользователей вместе с информацией о их назначении, заполнении и использовании, профессиональные базы знаний, представленные в допускаемых к применению моделях представления, знаний с известными механизмами решения задач, системами пояснений к знаниям и форматам предоставления клиентам результатов процесса решения задач. Такие документы делают возможной удаленную работу с профессиональными экспертными системами, для которых имеется возможность непрерывного авторского сопровождения.

Система моделей знаний, используемых для создания и применения интерактивных документов, формируется в составе энциклопедии на основе добавления к предопределенным описаниям новых типов значений для полей профессиональных знаний, процедур и методов, реализующих механизмы доступа и обработки знаний интерактивных документов.

3 ФОРМЫ ДОКУМЕНТОВ ДЛЯ ЭЛЕКТРОННЫХ ЭНЦИКЛОПЕДИЙ

Для электронных энциклопедий в различных областях деятельности можно разработать несколько универсальных форм документов. Такие формы могут применяться в конкретных энциклопедиях либо непосредственно, либо модифицировано.

Использование в стандартных формах энциклопедий систем предопределенных полей и конструкций

значений позволяет разрабатывать универсальные методы редактирования и использования содержимого документов пользователями.

3.1 Терминологические словари

Разработчики, размещающие свои материалы в энциклопедии, имеют возможность создания вспомогательных документов, применяемых при заполнении форм статей и других материалов, предназначенных для внешних пользователей, а также при формировании поисковых запросов. К таким вспомогательным документам относятся словари, содержащие структурированные разъяснения смысла и описания зависимостей между профессиональными терминами предметной области электронной энциклопедии.

Всякий словарь и входящие в него описания отдельных понятий, представляют собой класс документов, создаваемых на основе единой формы. К словарям специалистов, размещающих документы в энциклопедии, относятся также словари семантических связей или зависимостей между документами и их частями.

Словари разработчиков создаются, используются и сопровождаются так же, как и другие классы документов энциклопедии.

3.2 Классификаторы

В классификаторах содержится информация о разбиении документального пространства энциклопедии на группы близких по содержанию или смыслу статей. В системе VEDA реализуются возможности организации ручного или автоматического соотнесения размещаемых документов в выбранной системе классификаторов.

3.3 Справочная документация

В формах справочной документации представляется разнообразная информация, позволяющая получить представление о работе над энциклопедией, источниках и методах создания материалов. В структуре справочных документов предусмотрены такие компоненты как: планы создания энциклопедии, схемы сопровождения и согласования размещаемых материалов, используемые источники и процедуры верификации содержимого документов.

3.4 Специалисты

Для размещения информации о создателях энциклопедии используется специальная форма документов, содержащих разнообразные сведения об авторах статей и других материалов, размещенных в энциклопедии, рецензентах, экспертах их участии в

деятельности творческого коллектива, биографии, описание имеющегося опыта, почтовые адреса и т.д.

3.5 Словари статей по разделам классификатора (событиям, персоналиям, предметам, процессам и т.д.)

Основная компонента энциклопедии, содержащая системы статей, представляющих знания специалистов по предметной области. Полный перечень таких словарей и их структура уточняются отдельно для каждой предметной области. Специальный словарь в данном разделе связан с открытыми вопросами и разыскиваемой информацией.

3.6 История энциклопедии

В документах на основе формы история размещаются материалы, отражающие процесс развития энциклопедии: добавления новой и модификации уже имеющейся информации. История энциклопедии связана с архивом нереализованных или устаревших версий материалов.

4 ФОРМЫ ДОКУМЕНТОВ ПОЛЕЙ ПРОФЕССИОНАЛЬНЫХ ЗНАНИЙ

Основное назначение полей профессиональных знаний специалистов связано с возможностью автоматического применения, представленных в них знаний с целью решения задач пользователей. Это могут быть задачи обучения, консультирования, тестирования, диагностики, прогнозирования. Такие поля могут создаваться для узких предметных областей, связанных с конкретной деятельностью референтов, консультантов, экспертов, технологов.

4.1 Терминологические словари

В составе полей знаний специалистов, словари содержат структурированные представления как смысла отдельных понятий, применяемых в описаниях профессиональных знаний и методов их обработки, так и смысловых зависимостей понятий.

4.2 Постановки задач и методы их решения

С помощью данной формы осуществляется создание документов, относящихся к метауровню процессов создания и обработки систем выделяемых профессиональных знаний. В структуру формы входят компоненты, позволяющие размещать описания постановок задач, решаемых в конкретной области деятельности. В такие описания входят классификация

задач представления ситуаций и исходных данных, ссылки на методы решения, консультации специалистов, приводятся схемы применения рекомендуемых методов и практические примеры использования. Информация в документах на основе формы постановок задач предназначена как для визуального просмотра, так и управления процессами применения документов: анализа постановок задач, выбора метода решения.

4.3 Системы правил решения задач

Соответствующие документы содержат структурное описание процессов использования профессиональных знаний отдельных специалистов, сопоставляемые отдельным типам задач, вместе с авторскими комментариями примерами использования, эвристиками, советами по организации работ.

5 ОРГАНИЗАЦИЯ ДОСТУПА И ОХРАНА АВТОРСКИХ ПРАВ

Документы в составе электронной среды энциклопедии имеют авторов и владельцев, а также описания профилей доступа для работы с документами. Пользователям может быть предоставлена возможность удаленного размещения материалов, в том числе предназначенных авторам. Регистрация пользователей связана с отнесением их к некоторому классу, определяющему множество доступных документов и виды работы с ними. Регистрация документов, размещаемых в системе, также связана с созданием классификатора пользователей, по видам доступных операций над этими документами.

6 СТРУКТУРА СИСТЕМЫ VEDA

Система VEDA состоит из серверной части реализуемой на платформе Oracle 8i с помощью компонент Oracle Server 8i (R2) и Oracle Application Server 4 и содержит:

- подсистему хранилища документов, решающую задачи администрирования, регистрации, размещения и извлечения документов;
- базу поисковых запросов, позволяющую размещать и извлекать информационные заявки, обрабатывать поток входных документов хранилища, формируя списки ссылок на документы, соответствующие запросам, размещенным в базе;
- модули работы с интерактивными документами, обеспечивающие возможность решения задач пользователей с помощью механизмов обработки профессиональных знаний, содержащихся в документах;

- Web-интерфейс Archivarius, предназначенный для удаленной навигации по санкционированным ресурсам организации удаленного доступа, ввода и визуализации документов, работы с семантической классификацией документов.

Построение описаний типов и структур значений полей, форм и сценариев визуализации документов, создание, заполнение и классификация документов осуществляются с помощью клиентского приложения мастера форм.

Составление и редактирование паспортов информационных заданий клиентов, размещение и извлечение паспортов из базы запросов выполняется с помощью клиентского приложения мастера запросов.

Работа выполнена при финансовой поддержке РФФИ грант №00-07-96005

Список литературы

[1] К.И.Костенко Ассоциативные модели сетевых полей знаний //Библиотечное дело на пороге XXI века: Тезисы докладов Межд. Конф. (Москва 15-16 апр.1998 г.) ч1.-М.,1998.-с.136-137.

[2] К.И.Костенко, Б.Е.Левицкий Концепция системы документальных ресурсов библиотечных и информационных учреждений Краснодарского края: Вторая Междунар. конф. ИОЛ, (Ст.-Петербург, 28 февраля - 3 марта 2000 г.), с. 371

[3] К.И.Костенко, С.В.Назаренко Интеллектуальные технологии для библиотек в Интернет. Вторая Междунар. конф. ИОЛ, (Ст.-Петербург, 28 февраля – 3 марта 2000 г.), с. 372-373.

[4] М.М. Денисов, К.И. Костенко Представление и обслуживание регулярных информационных запросов для хранилищ документов на базе системы VEDA. Междунар. конф. ТЕЛЕМАТИКА, (Ст.-Петербург, 28-30 мая 2000 г.), стр.28-29.