

## **НЕКОТОРЫЕ ВОПРОСЫ ПОСТРОЕНИЯ АРХИТЕКТУРЫ ИИС АДМИНИСТРАТИВНО-ПРОИЗВОДСТВЕННОГО НАЗНАЧЕНИЯ В СРЕДЕ ДАННЫХ РАЗНОЙ СТЕПЕНИ СТРУКТУРИРОВАННОСТИ**

Отсутствие в настоящее время стандартов и рекомендаций к структурированности размещаемой в Интернете разнообразной информации приводит к известным проблемам и ограничениям для поиска информации в этой среде, где используется, в основном, контекстный поиск.

В корпоративных сетях порождаются информационные потоки кроме структурированных данных, представленных в виде реляционных моделей, также слабоструктурированные, в виде целостных документов и форм с относительно часто изменяемым набором параметров и неструктурируемой в виде аудио- и видео- данных, которые размещаются на серверах различного вида. Сюда можно отнести документы государственных фискальных и местных контролирующих органов, структура которых меняется ежемесячно или по-квартально, что приводит к необходимости доработки как самих форм так и базы данных, где они размещаются.

В статье рассматриваются вопросы построения информационных систем корпоративного уровня для решения задач административного и производственного назначения, которые интегрируют и обрабатывают разнородную по структуре информацию различными группами приложений.

Для повышения эффективности этих систем рекомендуется определенная декомпозиция узлов обработки в прикладной сети, т.к. для этого требуются различные технологии и шаблоны, в частности, шаблоны для поиска во внешних источниках и внутри корпоративной сети, шаблоны для работы в библиотечных сетях и т.п.

Поисковые технологии может реализовывать поисковый робот или поисковая машина, как специализированные программно-технические средства в корпоративной сети. Поисковый робот выполняет запросы к Web-ресурсам Интернета и к корпоративным Web-источникам (Web-сервера структурных подразделений и удаленных филиалов, например). Поисковому роботу могут задаваться как ключевые слова для поиска, так и другая, дополнительная информация в виде области поиска, URL-адресов, тематических рубрик и т.п. в результате поиска создаются начальные поисковые образы найденных информационных материалов, которые проходят первичную индексацию с формированием первичной индексной БД. Найденные поисковые образы могут отфильтровываться

© В.И. Костюк, Ю.А. Тимошин, Т.Г. Шемсединов, 2007

с помощью специальных алгоритмов или с участием пользователя, который отбирает из найденного списка требуемые объекты, которые образуют, в конечном счете, базу индексированных поисковых образов.

Для такой базы индексов после ее обновления или дополнения требуется оптимизировать структуру индексов для улучшения быстродействия работы базы. Нами использовалась объектно-иерархическая СУБД UOD (Universal Object Database), разработанная в НТУУ “КПИ” в НИИ системных технологий [1].

Для более полного учета данных о Web-источниках Интернета, в соответствии с запросами внутренних клиентов, с Web-серверов подразделений оперативно снимается и добавляется информация о поисковых образах для их учета в описанной базе индексов и для ускорения первоначального поиска в этой базе до выхода с поисковым запросом в Интернет через HTTP-сервер.

Для связи HTTP-сервера и объектной СУБД был написан специальный CGI-скрипт и создан протокол USP (Universal Servise Protocol).

На рисунке 1 показана схема взаимодействия этих компонентов между собой.

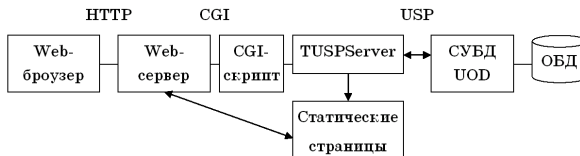


Рис. 1 – Структура компонентов доступа к данным через CGI-скрипт

Работа в этой технологической цепочке происходит следующим образом.

Из Web-браузера запрос пользователя направляется на Web-сервер, который, в свою очередь, вызывает CGI-скрипт. Скрипт проверяет защищены ли требуемые данные или они доступны для публичного использования. Если они для ограниченного использования, то проверяются права пользователя согласно базе пользователей, размещенной в UOD-базе. После этого формируется соответствующий запрос к USP-серверу. Функция GetHTML выполняет заданные в БД правила по оформлению объекта, который был запрошен. Результатом работы функции есть готовая HTML-страница, которая передается CGI-скрипту, который с соответствующими HTTP заголовками возвращает ее через Web-сервер пользователю. Кроме обычных текстовых данных и правил их оформления, в базе UOD могут размещаться также бинарные файлы (рисунки, архивы и т.п.), доступ к которым также обеспечивается через CGI-скрипт, который автоматически распознает тип данных и отдает их с соответствующими HTTP заголовками. Правила, по котрым генерируется страничка функцией GetHTML, фактически являются языком UODML.

Такая структура дает возможность публикации на Web информацию,

сформированную на основе поисковых образов, хранящихся в объектной БД. TUSPServer это разработанный на Delphi специализированный сервер с рядом реализованных на языке UODML функций и тегов, которые предназначены для создания пользователем собственных Web-приложений, которые обрабатывают информационные объекты со слабовязанной структурой.

Перечень задач по сканированию и индексации ресурсов IP сети включает следующие - протокол ICMP (пинг одного IP-адреса, сканирование диапазона IP-адресов, сканирование роутинга) / сканирование HTTP-сервиса (индексация доменных имен, полученных в ответах HTTP-протокола, индексация путей, слов, выражений, метаданных, рекурсия по ссылкам и т.п.) / сканирование FTP -сервиса / протокол NetBios / использование информации от DNS -сервиса / сканирование портов (сервисы и порты TCP, сервисы и порты UDP).

Для реализации этих задач были разработаны три основных скрипта: а) скрипт доступа к данным индекса (поисковый скрипт по ключевым словам); б) скрипт первичной индексации; в) скрипт обновления индексации по существующим страницам.

Кроме поискового робота, который работает исключительно с Web-контентом для внутреннего корпоративного поиска в среде БД, информационно-справочных систем и приложений, тематических серверов и т.п., может использоваться также Поисковая машина для обработки структурированной информации, через SQL-запросы, в реляционной БД, а через USP-запросы - в объектной БД. При этом в обоих вариантах используются только типовые запросы.

Поисковая машина не только хранит типовые запросы к системе, но также ведет реестр активированных информационных объектов и источников в сети. Кроме того, Поисковая машина хранит поисковые шаблоны и метаданные описаний информационных объектов, а также метаданные по структуре тематических коллекций и описаний электронных документов.

На рисунке 2 показана обобщенная структура разработанной интегрированной информационно-поисковой системы университета, подробное описание которой можно найти в [2].

В такой структуре SQL-сервер обеспечивает доступ к каталогу ресурсов и ведет его. Для обеспечения доступа клиентских приложений к каталогу используется специальный тонкий клиент, который через ODBC-источник подключается на время сеанса непосредственно к соответствующей РБД.

Пользователь через Web-браузер может осуществлять со своей рабочей станцией прямой поиск в среде Интернет-источников с использованием поисковиков Интернета.

Для упрощения взаимодействия клиента с поисковым роботом и поисковой машиной был разработан коммуникатор, через который пользователи системы могут формировать или задавать области поиска, критерии и стратегию поиска, определять новые тематические рубрики и т.п.

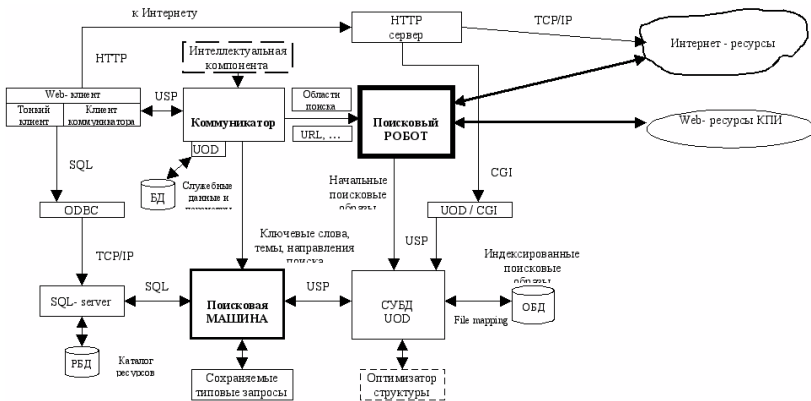


Рис. 2 – Структура взаимодействия компонентов Интегрированной информационно-поисковой системы университета

Коммуникатор имеет свою объектную базу данных, где хранятся служебные данные (например, метаданные и алгоритмы для обеспечения поиска по различным критериям), сведения о классификаторах и справочниках системы, URL-адреса источников, данные о настройках фильтров, др. В этой БД хранятся различного вида библиотеки (поисковых образов, алгоритмов интерпретации, критериев релевантности, моделей анализа, программ расчета статистических параметров и т.п.), таблицы настроек коммуникатора, сетевых адресов источников информации, тематические рейтинги, которые используются для поддержки различных видов поиска.

Коммуникатор имеет серверную и клиентскую части, которые взаимодействуют по USP-протоколу. Серверная часть имеет интеллектуальную компоненту, которая функционально и программно осуществляет поддержку принятия решений пользователем по выбору или созданию новых поисковых критериев, описания целей поиска и формированию алгоритмов стратегий поиска, а также поддерживает вычисление статистических оценок работы поискового робота и поисковой машины.

Клиентская часть коммуникатора устанавливается на рабочей станции, другими компонентами которой могут быть Web-браузер и тонкий клиент. Последний обеспечивает пользователю непосредственный доступ к каталогу ресурсов.

Дальнейшее функциональное расширение коммуникатора может быть связано с обеспечением взаимодействия и созданием управляемых шлюзов компьютерной сети с удаленными и мобильными пользователями, поддержкой среды распределенного проектирования и т.п., что описано в [3].

## **Выводы**

Предлагаемый подход к решению проблем интеграции информации в корпоративной сети с разной степенью ее структурированности путем разделения задач формирования, хранения и обработки потоков данных, позволяет создать более точные модели и разнородных информационных объектов, в большей степени формализовать их, повысив эффективность обработки, значительно уменьшить “поисковый шум”, обеспечить более глубокую структурированность информации и тематических источников, снизить потоки данных в корпоративной прикладной сети.

## **Литература**

1. Universal servise protocol (USP), Universal object database (UOD)-спецификация, методика использования, цели (<http://niist.ntu-krp.kiev.ua/USP>).
2. Интегрированная информационно-поисковая система мониторинга деятельности технического университета. Отчет по НИР.- К., УКРН-ТИИ, 2005, ном. гос. регистр. 0103U000243.
3. Разработка среды распределенных виртуальных лабораторий для образовательных и научных исследований. 2-й Международный Форум "Прикладная электроника. Состояние и перспективы развития", МРФ-2005, т.3, Харьков, 2005.