

ОБРАБОТКА ИНФОРМАЦИИ. ПРОГРАММНЫЕ КОМПЛЕКСЫ

УДК 004.418

ПРОБЛЕМЫ ВЕДОМСТВЕННОГО ПОИСКА ИНФОРМАЦИИ И ВОЗМОЖНЫЕ ПУТИ ИХ РЕШЕНИЯ

А.О. Дементьев

ГОУ ВПО «Тамбовский государственный технический университет», г. Тамбов

Рецензент д-р техн. наук, профессор В.Ф. Калинин

Ключевые слова и фразы: базы данных; информационная система; информационное общество; парадигмы программирования; язык Python.

Аннотация: Рассматривается вопрос оптимизации поиска и взаимодействия различных баз данных в органах государственной власти и решение данной проблемы при помощи разработанной на языке Python программы, позволяющей интегрировать информацию из разрозненных баз данных в одно целое.

До недавнего времени основные усилия государства в области развития информационных технологий (ИТ) были связаны с набором федеральных программ, самая известная из которых – Федеральная целевая программа (ФЦП) «Электронная Россия» (2002–2010 гг.). Завершилась она с непонятным результатом.

В стране существует также «Стратегия развития информационного общества в РФ» (СРИО), которая была утверждена 27 июля 2007 г. на заседании Совета безопасности под председательством Президента РФ Владимира Путина и подписана в окончательном виде 7 февраля 2008 г.

Представляя СРИО еще в июле 2007 г., Леонид Рейман, тогда министр Мининформсвязи, обещал представить план реализации к концу года, но так и не представил. После подписания СРИО Президент страны говорил о том, что такой план будет рассмотрен и утвержден Правительством РФ до конца 2008 г., но и этого не произошло.

Вскоре за дело информатизации государства взялся новый Президент страны Дмитрий Медведев, человек, как известно, к информационным технологиям равнодушный. Президент провел 17 июля 2008 г. в Петрозаводске заседание президиума Госсовета с повесткой дня: «О реализации

Дементьев Алексей Олегович – аспирант кафедры «Информационные системы и процессы», e-mail: alexey.dementyev@gmail.com, ТамбГТУ, г. Тамбов.

стратегии развития информационного общества в Российской Федерации». Также Президент подписал 11 ноября 2008 г. указ «О Совете при Президенте РФ по развитию информационного общества в РФ» и определил его состав. В совет не вошли представители ИТ-индустрии.

Первое заседание этого совета прошло 12 февраля 2009 г., на нем Дмитрий Медведев подверг критике недопустимо низкий уровень развития информационно-коммуникационных технологий (ИКТ) в России и поставил задачу за два года создать условия для интеграции России в мировое информационное общество. Там же глава государства одобрил инициативу Леонида Реймана по созданию девяти рабочих групп, отвечающих за определенные направления СРИО.

Минкомсвязи разработало, а Президиум Правительства в конце сентября 2010 г. одобрил проект Федеральной целевой программы «Информационное общество». Сейчас проект ФЦП с общим бюджетом в 9,3 млрд рублей проходит последние правки в аппарате правительства.

В нем отсутствует ожидаемый многими участниками рынка проект создания российской национальной поисковой системы. Он заменен на разработку поисковой системы для ведомственных нужд, с помощью которой чиновники будут искать документы в министерских базах данных, а остальные пользователи – находить нужные документы на сайтах госорганизаций.

Основной проблемой поиска в ведомственных базах данных до сих пор остается их разнородность. Каждое ведомство использует свое программное обеспечение, внедренное коммерческой организацией. Сейчас не существует нормативно-технической базы для внедрения программного обеспечения на государственном уровне, поэтому каждое ведомство вправе самостоятельно искать и выбирать подрядчика.

В ФЦП «Информационное общество» (в плане реализации) отмечается, что данную проблему необходимо решать на региональном уровне, тем самым решение этой задачи делегируется ведомствам.

В последнее время, для ликвидации разобщенности информационных систем ведомств используется такой подход: каждое ведомство, для решения задач информационной связи с другим ведомством, нанимает того же подрядчика для реконструкции их программного обеспечения и реорганизации баз данных.

Для решения задач поиска информации в ведомственных базах данных предлагается иной подход – внедрить программно-аппаратный комплекс для связывания разрозненных баз данных и решения второстепенных задач кэширования и централизованного резервирования информации.

В связи с этим было сформировано техническое задание, в котором описаны основные требования к информационной системе.

Основные требования к серверной части.

1. Объектно-ориентированный язык должен иметь объекты и методы, наследование классов, множественные наследования классов, механизм обработки исключений, итераторы, управление памятью, модульную систему.

2. Кроссплатформенность – возможность запускать серверную часть независимо от типа операционной системы.

3. Поддержка модулей доступа к базам данных, основные типы: Microsoft SQL Server, MySQL, Oracle, PostgreSQL, SQLite, Firebird, Informix, Sybase, DB2.

4. Поддержка компиляции модулей в ядро программы.

5. Поддержка sha-, md5-шифрования.

6. Поддержка hash-функций.

7. Поддержка кеширования множественных запросов к базам данных.

8. Возможность обфускации кода программы.

Основные требования к модулю администрирования.

1. Web-интерфейс.

2. Поддержка аутентификации клиента на основе логина + пароля + IP адреса.

3. Простое и удобное управление.

4. Прозрачное взаимодействие с серверной частью.

На рассмотрение были выбраны несколько языков программирования: C++, C#, Java, Python, Ruby, PHP.

1. C++ – компилируемый статически типизированный язык программирования общего назначения. Поддерживая разные парадигмы программирования, сочетает свойства как высокоуровневых, так и низкоуровневых языков. Для разработки подходит по многим параметрам, является очень распространенным языком программирования и поддерживает взаимодействие с любыми базами данных через поставщиков операционной системы.

Не удовлетворяет требованиям серверной программы по п. 2, так как перед запуском программа должна быть скомпилирована под определенную операционную систему, то есть должна иметь модули взаимодействия со всеми операционными системами и их поставщиками данных.

2. C# – язык имеет статическую типизацию, поддерживает полиморфизм, перегрузку операторов (в том числе операторов явного и неявного приведения типа), делегаты, атрибуты, события, свойства, обобщенные типы и методы, итераторы, анонимные функции с поддержкой замыканий, LINQ, исключения, комментарии в формате XML.

Не удовлетворяет требованиям п. 1 – C# не поддерживает множественное наследование классов.

3. Java – объектно-ориентированный язык программирования, разработанный компанией Sun Microsystems. Приложения Java обычно компилируются в специальный байт-код, поэтому они могут работать на любой виртуальной Java-машине независимо от компьютерной архитектуры.

Не удовлетворяет требованиям по п. 2 – для запуска системы требуется виртуальная Java-машина, которая не может быть включена в серверную часть.

4. Python – высокоуровневый язык программирования общего назначения с акцентом на производительность разработчика и читаемость кода. Синтаксис ядра Python минималистичен. В то же время стандартная библиотека включает большой объем полезных функций. Python поддерживает несколько парадигм программирования, в том числе структурное,

объектно-ориентированное, функциональное, императивное и аспектно-ориентированное. Основные архитектурные черты – динамическая типизация, автоматическое управление памятью, полная интроспекция, механизм обработки исключений, поддержка многопоточных вычислений и удобные высокоуровневые структуры данных. Код в Python организуется в функции и классы, которые могут объединяться в модули (которые, в свою очередь, могут быть объединены в пакеты).

Полностью соответствует требованиям, интерпретатор может быть включен в программу, имеет модули для запуска административной части.

5. Ruby – динамический, рефлексивный, интерпретируемый высокоуровневый язык программирования для быстрого и удобного объектно-ориентированного программирования. Язык обладает независимой от операционной системы реализацией многопоточности, строгой динамической типизацией, сборщиком мусора и многими другими возможностями.

Не удовлетворяет требованиям по п. 2 – для запуска системы требуется интерпретатор, который не может быть включен в серверную часть.

6. PHP – скриптовый язык программирования общего назначения, интенсивно применяемый для разработки Web-приложений. В настоящее время поддерживается подавляющим большинством хостинг-провайдеров и является одним из лидеров среди языков программирования, применяющихся для создания динамических Web-сайтов.

Не удовлетворяет требованиям по п. 1 – не имеет средств высокоуровневого взаимодействия.

Для решения проблемы было разработано общее техническое задание на разработку программного обеспечения на языке Python, как отвечающего всем требованиям для реализации поставленной задачи.

1. Определение типа базы данных при подключении.
2. Абстрактная реализация поискового механизма на основе статических методов.
3. Динамическое наследование методов класса.
4. Динамическое создание потоков обработки данных.
5. Автоматическая генерация объектов данных на основе пользовательского ввода.

6. Автоматическое кэширование запросов к базам данных.

7. Автоматическое создание кэшируемых объектов.

8. Определение наследования по типу и параметрам объекта.

На основании технического задания была разработана программа поиска и взаимодействия различных баз данных и вывода информации для пользователя.

Кратко решение поставленной задачи и алгоритм действия программного комплекса кратко можно описать следующим образом:

- 1) аутентификация клиента (логин + пароль + IP адрес);
- 2) определение прав доступа к данным;
- 3) кэширование запроса клиента во внутренней базе данных;
- 4) определение выборки результатов и соотношение с объектами;
- 5) генерирование иерархии объектов выборки;
- 6) динамическое создание объектов, классов, методов выборки;

- 7) поиск результатов и сравнение методов;
- 8) выдача результатов по классам;
- 9) сравнение выдачи результатов;
- 10) выдача результатов по запросу клиента.

Разработанная информационная система полностью справляется с высокими нагрузками, позволяет вести подробный анализ баз данных, автоматически соотносить различные типы данных, определять зависимости в структурах баз данных и т.д.

Информационная система призвана обеспечить прозрачное взаимодействие различных типов баз данных, «слепой поиск» информации, разграничение прав доступа к информации и высокую скорость внедрения продукта на всей территории страны.

Список литературы

1. О Федеральной целевой программе «Электронная Россия (2002–2010 гг.)» [Электронный ресурс] : постановление Правительства РФ от 28.01.2002 г. № 65. – Режим доступа : <http://government.e-rus.ru/>. – Загл. с экрана.

2. Стратегия развития информационного общества в России (проект) [Электронный ресурс] : Совет безопасности РФ. – Режим доступа : <http://www.scrf.gov.ru/documents/87.html>. – Загл. с экрана.

3. Национальной поисковой системы в России не будет [Электронный ресурс]. – Режим доступа : http://www.rian.ru/tech_internet/20101027/-289709777.html. – Загл. с экрана.

4. Информационное общество ФЦП [Электронный ресурс]. – Режим доступа : http://www.tadviser.ru/index.php/Статья:Информационное_общество_ФЦП. – Загл. с экрана.

Problems of Industry Sponsored Information Search and Possible Ways of their Solution

A.O. Dementyev

Tambov State Technical University, Tambov

Key words and phrases: database; information system; information society; programming paradigms; Python.

Abstract: The paper is devoted to the question of optimization of search and interaction of various databases in government authorities as well as to the solution of the given problem by means of the program developed in Python language, enabling to integrate the information from isolated databases into a single whole.

© А.О. Дементьев, 2010