

УДК 004.75:004.451.82



## ВЕБ-СЕРВИСЫ В ЛИНГВИСТИЧЕСКИХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

А. Р. Ерошенко

УМИФ, г. Киев, Украина, demode@ukrpost.ua

В статье рассмотрены принципы работы веб-сервиса. Автором выделены основные компоненты сервис-ориентированной архитектуры (SOA) и обоснована технология взаимодействия компонентов веб-сервисов. Проанализирована возможность применения технологии CORBA к задаче создания и использования лингвистических ресурсов.

ВЕБ-СЕРВИСЫ, ИНТЕРНЕТ, CORBA, SOA, JAVA, XML, SOAP, WSDL

### Введение

С ростом популярности технологии WEB 2.0 в последнее время значительно увеличилось количество ресурсов Интернет, предоставляющих не только доступ к информации, но и позволяющие решать гораздо более широкий круг задач. Совсем недавно использование Интернет для поиска информации и поиска файлов было вполне достаточным. Сегодня же естественной является функция создания и редактирования посредством Интернет-браузера сложных документов, графических файлов, управления научными, производственными и бизнес-процессами. Кроме того, неуклонно возрастает роль лингвистических технологий в создании информационных систем, особенно для класса интеллектуальных. В связи с этим особо актуализируются задачи создания лингвистических ресурсов и возможности применения компьютерных технологий в лексикографии.

В отечественной научной литературе уделяется определенное внимание компьютерным технологиям лексикографической деятельности. Прежде всего, это научные разработки В. Широкова [1], А. Костышина [2], А. Рабульца [3], К. Якименко [4]. В этих работах ставится и решается целый ряд задач: перевод в цифровую форму традиционного лексикографического наследия, разработка инструментальных систем для создания электронных лингвистических продуктов, проведение разноплановых лингвистических исследований.

Существующие на сегодняшний день лексикографические исследования и проекты, целью которых является создание лингвистических ресурсов, выполняются значительными коллективами исследователей, зачастую распределенных территориально. Следовательно, стоит задача создания инструмента, позволяющего выполнять большие лингвистические проекты в режиме сетевого взаимодействия. Такая постановка задачи является совершенно новой для прикладной лингвистики.

Целью данного исследования является обоснование возможностей применения технологии CORBA для создания и использования лингвистических ресурсов. Реализация поставленной цели обусловила необходимость решения таких задач:

- описание принципа работы веб-сервиса;
- выделение основных компонентов сервис-ориентированной архитектуры;
- обоснование технологии взаимодействия компонентов веб-сервисов и операций, которые они выполняют;
- описание алгоритма взаимодействия информационных технологий и лингвистических ресурсов.

### 1. Теория веб-сервисов

Современный веб-сайт можно представить как совокупность взаимодействующих между собой сервисов. Обмен информацией может происходить также с сервисами, расположенными на других сайтах. Таким образом, мы получаем взаимодействующую гетерогенную среду, построенную по технологии, которая называется *CORBA*<sup>1</sup> [5].

*CORBA* является интегральной технологией взаимодействия веб-сервисов. Одной из ее главных составляющих является технология сервисно-ориентированной архитектуры *SOA* (*Service-Oriented Architecture*) [6]. На основе *SOA*-технологии нами создан веб-сервис, дающий возможность пользователям Интернет в режиме просмотра работать с реестром нескольких украинских словарей. Уникальность данной системы состоит в том, что пользователь может получить доступ к любому слову или словосочетанию словарной статьи, тогда как большинство мировых лингвистических систем позволяют выполнять данную процедуру только с реестром слов.

По своей сути *SOA* является новым направлением развития информационных технологий, представляющих собой создание глобальной сети распределенных вычислений. Наиболее подходящей средой функционирования данной системы являются интернет-сети. Инструментом для создания *SOA*-приложения могут быть многие популярные системы разработки: *.NET*, *PHP*, *Java*, и так далее. Этот список с каждым годом будет все более расширяться. Данный инструментальный средств

<sup>1</sup> *CORBA* (*Common Object Request Broker Architecture*) — общая архитектура брокера объектных запросов. Это технологический стандарт написания распределённых приложений, продвигаемый консорциумом *OMG*.

разработки может быть интегрирован в различные аппаратные платформы. Главное в том, что столь различные приложения могут обмениваться данными, то есть эффективно функционировать.

Прежде чем приступить к рассмотрению лингвистического веб-сервиса, опишем теорию функционирования распределенных вычислительных ресурсов, которыми в нашем случае являются веб-сервисы. Технологии Web-сервисов — это набор основанных на XML спецификаций, обеспечивающих универсальный метод технического описания сервисов и взаимодействия с ними. Сами сервисы, реализованные в соответствии с этими спецификациями, называют Web-сервисами. Возможно, именно технологии Web-сервисов сыграли свою роль в наименовании сервис-ориентированной архитектуры. На самом деле принципы SOA существуют уже как минимум третье десятилетие, в то время как о Web-сервисах мы узнали не так давно. На протяжении большей части этого времени доминировала объектно-ориентированная идеология построения систем, задачи технической реализации которой сводилась главным образом к возможности вызова удаленных объектов

Концепция SOA подразумевает существование информационных процессов в виде совокупности сервисов, взаимодействующих друг с другом.

Различают следующие три основных компонента сервис-ориентированной архитектуры:

- пользователь сервиса: приложение, программный модуль либо сервис, осуществляющий поиск и вызов необходимого сервиса из реестра сервисов по описанию сервиса, а также использующий сервис, предоставляемый провайдером сервиса, в соответствии с интерфейсом сервиса;
- провайдер сервиса: приложение, программный модуль либо сервис, осуществляющий реализацию сервиса в виде веб-сервиса, прием и исполнение запросов пользователей сервиса, а также публикацию сервиса в реестре сервисов;
- реестр сервисов: библиотека сервисов, предоставляющая пользователям сервиса средства поис-

ка и вызова необходимого сервиса и принимающая запросы провайдеров сервисов на публикацию сервисов.

Современный веб-сервис может быть как компонентом пользователя сервиса, так и сочетать в себе компонент провайдера, что само по себе увеличивает его работоспособность и стабильность работы. Введем обозначения, которые будут нами использоваться при описании принципа работы данной системы (рис. 1):

*WSDL* — язык описания веб-сервисов, основанный на языке XML; определяет веб-интерфейс для любых функций, которые используют удаленный вызов процедур (RPC);

*UDDI* — платформонезависимая структура описания, поиска и интегрирования веб-сервисов в Интернет. Данная структура представляет собой средство регистрации вычислительных ресурсов в формализованном XML-формате;

*SOAP* — протокол обмена структурированными сообщениями в распределенной вычислительной среде. Если раскрыть аббревиатуру, получим *Simple Object Access Protocol*, то есть простой протокол доступа к объектам. SOAP используется для удаленного вызова процедур (функций), а также для обмена произвольными сообщениями в формате XML;

*XML* — расширенный язык разметки, представляющий собой текстовый формат, предназначенный для хранения структурированных данных и обмена информацией между программами. XML может эффективно работать с различными типами данных, и реальной альтернативы ему на сегодняшний день не существует.

Каждый компонент может играть либо лишь одну роль (быть, например, только пользователем сервиса), либо одновременно сразу несколько ролей (например, быть провайдером одних сервисов и пользователем других).

В ходе взаимодействия друг с другом компоненты сервис-ориентированной архитектуры выполняют следующие основные операции (рис. 2.):

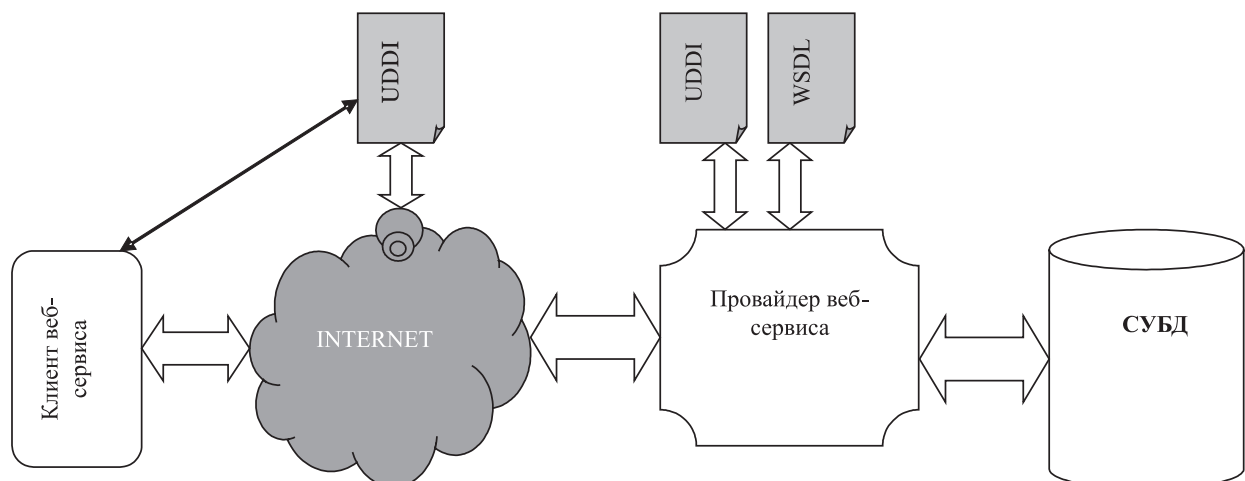


Рис. 1. Принцип работы веб-сервиса

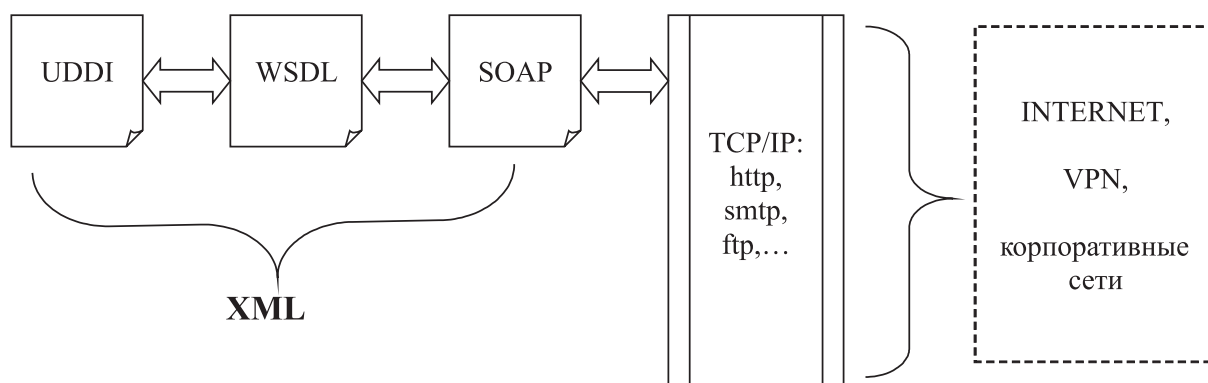


Рис. 2. Технология взаимодействия компонентов веб-сервисов

- публикация: для того, чтобы сервис был доступным (вызываемым) пользователям сервиса, необходимо сделать его интерфейс известным им;
- поиск: пользователь сервиса должен иметь возможность найти в реестре сервисов необходимый сервис, удовлетворяющий заданным критериям;
- связывание и вызов: после получения описания сервиса пользователь сервиса должен иметь возможность вызвать и использовать сервис в соответствии с его описанием.

Для удаленного взаимодействия с веб-сервисами используется *SOAP* [7]. Он обеспечивает взаимодействие распределенных систем независимо от объектной модели, операционной системы или языка программирования. Данные передаются в виде *XML* документов особого формата. Согласно определению *W3C*<sup>2</sup>, веб-сервисы – это приложения, которые доступны по протоколам и являются стандартными для Интернет. Нет требования, чтобы веб-сервисы использовали какой-то определенный транспортный протокол. Спецификация *SOAP* определяет, каким образом связываются сообщения *SOAP* и транспортный протокол (рис. 3).



Рис. 3. Структура *SOAP* пакета

Как мы видим, веб-сервисы состоят из клиентской и серверной части, а часто и сами имеют клиентский и серверный модули. Дело в том, что современные приложения, ввиду их сложности, требуют наличия системы обработки ошибок, которые возникают при работе в интернет-сети. Данные системы обеспечивают стабильность работы приложения. Например, если клиент веб-сервиса перестает получать информационные пакеты от серверной части приложения, он может динамически изменять тайм-ауты ожидания сигнала, либо производить поиск другого источника информации. Таким образом, клиентская часть приложения уже перестает быть таковой фактически, а превращается в некое подобие серверной.

Этапы взаимодействия сервера и клиента будут выглядеть следующим образом:

- Клиентское приложение создает экземпляр объекта *SOAPClient*.
- *SOAPClient* читает файлы описания методов веб-сервиса (*WSDL*). Эти файлы могут храниться и на клиентской части приложения.
- Клиентское приложение, используя возможности позднего связывания методов объекта *SOAPClient*, вызывает метод сервиса. *SOAPClient* формирует пакет запроса (*SOAP Envelope*) и отправляет на сервер. Возможно использование любого транспортного протокола, но, как правило, используется *HTTP*.
- Пакет принимает серверное приложение *Listener* (может представлять собой *ISAPI* приложение или *ASP* страницу), создает объект *SOAPServer* и передает ему пакет запроса.
- *SOAP-Server* читает описание веб-сервиса, загружает описание и пакет запроса в *XML DOM* дерева.
- *SOAP-Server* вызывает метод объекта/приложения, реализующего сервис.
- Результаты выполнения метода или описание ошибки конвертируются объектом *SOAPServer* в пакет ответа и отправляются клиенту.
- Объект *SOAPClient* проводит разбор принятого пакета и возвращает клиентскому приложению результаты работы сервиса или описание возникшей ошибки.

<sup>2</sup> World Wide Web Consortium – организация, разрабатывающая и внедряющая технологические стандарты для всемирной сети Интернет

Помимо передачи данных, протокол *SOAP* предусматривает реакцию веб-сервиса на различные ошибки. Например на неверный вызов методов. Информация об ошибках в протоколе *SOAP* передается при помощи специальных дескрипторов.

В том случае, если необходима конфиденциальность обмена, возможно использование *Secure SOAP (SSOAP)*, в котором предусмотрено шифрование данных.

## 2. Особенности применения веб-сервисов в лингвистических системах

Проанализируем возможность применения технологии *CORBA* к задаче создания и использования лингвистических ресурсов.

В нашей модели лингвистическим ресурсом является интегрированный программный комплекс «Словники України», функционирующий на платформе операционной системы *Windows*, взаимодействовать с которым можно из локальной сети и через Интернет. Интерфейс программного взаимодействия через интернет-канал реализуется при помощи веб-сервиса, работающего на одной платформе с лингвистической системой.

Процедура доступа и взаимодействия с веб-сервисом согласно с технологией *SOA* одинакова для любых аппаратных платформ и операционных систем, что значительно упрощает процесс написания программ-клиентов.

Таким образом, веб-сервис опубликован в реестрах *UDDI*, что позволяет его найти и идентифицировать, имеет *WSDL*, в котором представлены функции программного интерфейса и теперь остается написать программу-клиент, используя подходящий инструментарий (*Perl, PHP, C#, Java, ...*).

Схема работы клиент-серверного модуля системы «Словники України» представлена на рис. 4.

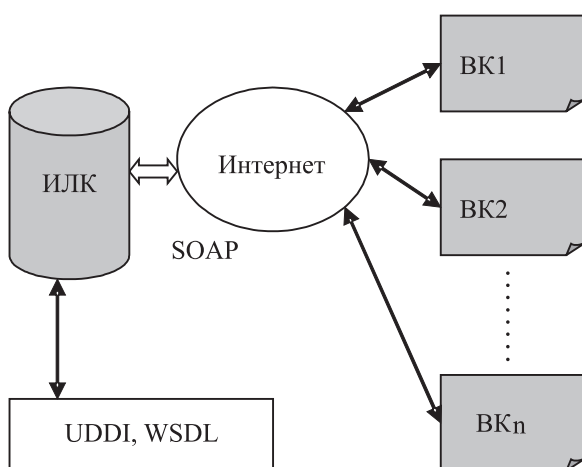


Рис. 4. Схема работы клиент-серверного модуля системы «Словники України»: ИЛК – интегрированный программный комплекс «Словники України»; BK1 – BK<sub>n</sub> – множество веб-клиентов с различными операционными системами и аппаратной платформой

В результате мы получаем гибкий программный инструмент, позволяющий выполнять, например, задачи такого рода:

- выбор словаря;
- получение всего реестра словаря или его части;
- поиск по заданному слову;
- получение словарной статьи;
- доступ к любому слову словарной статьи;
- получение скалярных величин, касающихся того или иного словаря.

## Выводы

Таким образом, проведенное исследование обосновывает необходимость использования технологии сервисно-ориентированной архитектуры *SOA* для создания лингвистических ресурсов и работы с ними. Преимущество веб-сервисов, разработанных с помощью *SOA*-технологии, дают возможность пользователям Интернет в режиме просмотра работать с реестром нескольких украинских словарей.

**Список литературы.** 1. Широков В.А. Элементы лексикографии. – К.: Довіра, 2005. – 304 с. 2. Костышин А. Использование сервис-ориентированной архитектуры и Web-сервисов для построения распределенных лексикографических сред: концепции и технологии // Прикладна лінгвістика та лінгвістичні технології: MegaLing-2006. – К.: Довіра, 2007. – С. 176-183. 3. Рабулець О.Г. Інтегровані лексикографічні системи: Автореф. дис. на здобуття наук. ступеня канд. техн. наук. – К., 2002. – 18 с. 4. Корпусна лінгвістика / В.А. Широков, О.В. Бугаков, Т.О. Грязнухіна та ін. – К.: Довіра, 2005. – 471 с. 5. *Common Object Request Broker Architecture*: <http://ru.wikipedia.org/wiki/CORBA>. 6. Фундамент для SOA: Инновации в технологиях и бизнесе. – Стратегия в области программного обеспечения IBM. – 2006. – № 1. 7. *World Wide Web Consortium, W3C*: <http://ru.wikipedia.org/wiki/W3C>.

Поступила в редколлегию 29.08.2008

УДК 004.75:004.451.82

Веб-сервіси в лінгвістичних інформаційних системах / О.Р. Єрошенко // Біоніка інтелекту: наук.-техн. журнал. – 2008. – № 2 (69). – С. 32-35.

У статті розглянуто принципи роботи веб-сервісу. Автором виділено основні компоненти сервіс-орієнтованої архітектури (SOA) і обґрунтована технологія взаємодії компонентів веб-сервісів. Проаналізована можливість застосування технології CORBA до задачі створення і використання лінгвістичних ресурсів.

Л. 4. Бібліогр.: 7 найм.

UDK 004.75:004.451.82

Web-services in the Linguistic Informative Systems / A.R. Yeroshenko // Bionica intellecta: Sci. Mag. – 2008. – № 2 (69). – P. 32-35.

Principles of work of web-service are considered in the article. An author selects the basic components of Service-Oriented Architecture (SOA) and technology of co-operation of web-services components is grounded. Possibility of application of the CORBA-technology to the task of creations and uses of linguistic resources is analysed.

Fig. 4. Ref.: 7 items.