

На правах рукописи

Климов Валентин Вячеславович

**МОДЕЛИ, АЛГОРИТМЫ И ПРОГРАММНЫЕ СРЕДСТВА
ПОИСКА И КОМПОЗИЦИИ ВЕБ-СЕРВИСОВ С ИСПОЛЬЗОВАНИЕМ
СЕМАНТИЧЕСКИХ ОПИСАНИЙ**

05.13.11 – математическое и программное обеспечение вычислительных
машин, комплексов и компьютерных сетей.

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Автор:

Москва – 2012

Работа выполнена в Национальном исследовательском ядерном университете «МИФИ».

Научный руководитель: доктор технических наук, профессор
Щукин Борис Алексеевич

Официальные оппоненты: доктор технических наук, профессор
Лебедев Георгий Николаевич
кандидат технических наук,
Кальченко Александр Николаевич

Ведущая организация: Федеральное государственное унитарное
предприятие «Центральный научно-
исследовательский институт экономики,
информатики и систем управления»

Защита диссертации состоится «21» марта 2012 г. в __ час. __ мин. на заседании диссертационного совета Д 212.130.03 при Национальном исследовательском ядерном университете «МИФИ» по адресу: 115409, Москва, Каширское шоссе, 31

С диссертацией можно ознакомиться в библиотеке Национального исследовательского ядерного университета «МИФИ».

Отзывы в двух экземплярах, заверенные печатью организации, просьба направлять по адресу: 115409, Москва, Каширское шоссе, д.31

Автореферат разослан «___» _____ 2012 г.

Ученый секретарь
диссертационного совета
д.т.н., профессор

Леонова Н.М.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследований.

По мере развития информационных технологий возникали разные подходы к написанию программ: модульное, объектно-ориентированное, компонентно-ориентированное программирование. Однако все подходы имеют ряд недостатков, они часто приводят к созданию негибких приложений, приходится создавать множество специализированных приложений для решения задач, которые лишь немного различаются. Кроме того, в последнее время большое значение придается программированию не монолитных систем, а систем, состоящих из большого числа разнородных частей. Поэтому, альтернативой традиционным подходам стала сервис-ориентированная разработка программного обеспечения. Использование данного подхода к программированию IT-инфраструктуры позволяет компаниям получать больше преимуществ при своем дальнейшем развитии. Подход SOA развивался с целью интеграции сложных и разнородных программных систем за счет использования слабо связанных друг с другом повторно используемых компонентов и обеспечения среды для их удобного связывания. Основной трудностью при реализации данного подхода является сложность нахождения веб-сервиса, который предоставлял бы требуемую функциональность. Зачастую при разработке крупных проектов, включающих сотни сервисов, программисту гораздо проще написать свой сервис, чем искать аналогичный среди уже реализованных. Тем самым нарушается главный принцип и преимущество SOA - принцип повторного использования компонентов, что негативно сказывается в дальнейшем на всем процессе разработки приложения.

Поиск и композиция сервисов приводит к повышению эффективности разработки сервис-ориентированных приложений, а также к упрощению процесса повторного использования сервисов. Поэтому данной проблематике в настоящее время уделяется пристальное внимание, следствием которого являются многочисленные научные исследования и публикации.

Значительный вклад в исследование и решение проблем поиска с использованием онтологий и технологий семантического веба внесли как отечественные учёные В.Ф. Хорошевский, Л.Б. Шереметов, В.А. Дерещкий, В.Д. Соловьев, Б.В. Добров, Н.В. Лукашевич и др., так и целый ряд зарубежных учёных A. Brogi, S. Corfini, M. Paolucci, O. Lassila, J. Ziembicki, K. Sacnhes, S. McIlraith, D. Martin и др.

Актуальность темы диссертационной работы определяется возрастающей потребностью индустрии разработки веб-ориентированных информационных систем в средствах поиска и композиции веб-сервисов с использованием семантических описаний в связи с быстрым ростом числа сервис-ориентированных приложений и необходимостью усовершенствовать и сделать более точным процесс поиска и композиции веб-сервисов. Это приведёт к ускорению разработок и снижению трудоёмкости создания SOA-приложений.

Цель диссертационной работы. Разработка и исследование моделей, алгоритмов и программных средств, расширяющих функциональные возможности поиска и композиции веб-сервисов путём учета предусловий, эффектов и отображений входов на выходы в их семантических описаниях.

Для достижения поставленной цели в рамках диссертационной работы были решены следующие задачи:

1. Разработаны модели поиска и композиции веб-сервисов с использованием семантических описаний.
2. Разработаны алгоритмы поиска и композиции веб-сервисов по их семантическим описаниям.
3. Решена поставленная задача оценки релевантности поискового запроса и семантического описания сервиса в случаях недостатка или избытка хранящихся и запрошенных пользователем входов и выходов веб-сервисов.
4. Разработаны структура и механизмы взаимодействия компонентов программной системы поиска и композиции веб-сервисов по их семантическим описаниям.
5. Проведено экспериментальное исследование разработанных моделей, алгоритмов и программных средств.

Объектом исследования является задача поиска и композиции веб-сервисов с использованием семантических описаний.

Предметом исследования являются модели и алгоритмы, используемые при ее решении.

Методы исследования. Исследования проводились с помощью постановки и решения оптимизационных задач (задачи о назначениях) и задач, сводящихся к формальному доказательству в исчислении предикатов первого порядка. Работа проводилась на основе принятых международных стандартов семантического описания веб-сервисов и онтологий предметных областей. При разработке программного обеспечения использовались подходы логического и клиент-серверного программирования, XML, RDF, OWL технологии.

Достоверность результатов подтверждена результатами экспериментального исследования предложенных моделей, алгоритмов, разработанной и практически реализованной системой поиска и композиции веб-сервисов с использованием семантических описаний, публикацией результатов в печати, апробацией на научно-технических конференциях.

Научная новизна работы заключается в следующем:

1. Разработаны интегрированная модель и алгоритмы поиска процессов веб-сервисов с использованием онтологий предметных областей и общепринятого лингвистического тезауруса WordNet.
2. Разработана расширенная модель семантического описания веб-сервиса за счет введения отображения входов на выходы и, а также алгоритмы её обработки.

3. Разработаны модель и алгоритмы использования предусловий, эффектов и отображений входов на выходы при поиске и композиции процессов веб-сервисов.
4. Разработана структура и механизмы взаимодействия компонентов программной системы поиска и композиции процессов веб-сервисов с использованием семантических описаний.

Практическая значимость. Разработанные модели и алгоритмы поиска и композиции могут быть использованы для:

- Усовершенствования стандартов, связанных с семантическим описанием веб-сервисов.
- Построения реестров, предназначенных для публикации и поиска веб-сервисов по их семантическим описаниям.

Практическую ценность проведённых исследований составляет реализованный комплекс инструментальных программных средств, расширяющий функциональные возможности поиска и композиции веб-сервисов путём учёта предусловий, эффектов и отображений входов на выходы в их семантических описаниях.

Основные научные результаты, выносимые на защиту.

1. Интегрированная модель и алгоритмы поиска процессов веб-сервисов с использованием онтологий предметных областей и общепринятого лингвистического тезауруса WordNet.
2. Расширенная модель семантических описаний веб-сервисов за счет введения отображений входов на выходы, а также алгоритмы её обработки.
3. Модель и алгоритмы использования предусловий, эффектов и отображений входов на выходы при поиске и композиции процессов веб-сервисов.
4. Структура и механизмы взаимодействия компонентов программной системы поиска и композиции процессов веб-сервисов с использованием семантических описаний, функционирующей на основе разработанных моделей и алгоритмов.

Реализация результатов работы. Разработана программная система поиска и композиции веб-сервисов по их семантическим описаниям. Система была использована при разработке и внедрении сервис-ориентированного приложения в ООО «Банковский Производственный Центр», а также в Министерство образования и науки РФ в рамках государственного контракта №П2187 от 9 ноября 2009 года, что подтверждается актами о промышленном внедрении и использовании результатов диссертационной работы.

Апробация работы. Основные положения и результаты диссертации докладывались и обсуждались на следующих конференциях и выставках:

- XVII, XVIII, XIX и XX международные научно-технические семинары «Современные технологии в задачах управления, автоматизации и обработки информации» (Алушта, 2008-2011);
- VIII Курчатовская Молодёжная Научная Школа (Москва, 2010);

- Международный форум «Дни русских инноваций» (Москва, 2009);
- III международная научно-практическая конференция «Информационные технологии в образовании, науке и производстве» (Серпухов, 2009);
- IX и X Всероссийские выставки научно-технического творчества молодежи НТТМ-2009 и НТТМ-2010 (Москва, 2009-2010);
- Всероссийские образовательные форумы «Смена молодых инноваторов» (Селигер, 2009-2010);
- Научные сессии НИЯУ МИФИ 2008-2011 (Москва, 2008-2011);
- X-XIV международные телекоммуникационные конференции "Молодежь и наука" (Москва, 2008-2011).

Разработанная система является победителем восьмого и девятого конкурсов молодежных инновационных проектов технопарка НИЯУ МИФИ, проводимого в рамках 12-й и 13-й Московских международных телекоммуникационных конференций «Молодежь и Наука» (2009-2010 гг.). Система является победителем по программам «Участник Молодежного Научно-Инновационного Конкурса» (У.М.Н.И.К.-2009 и У.М.Н.И.К.-2010), победителем Регионального Инновационного Конвента г. Москвы в 2010 году. Она является победителем конкурса НК-389П в рамках Федеральной Целевой Программы «Научные и научно-педагогические кадры инновационной России на 2009 – 2013 годы». Система заняла второе место в категории «IT-реализация» (специалисты) в номинации «Software» в конкурсе «IT-прорыв» партии «Единая Россия», государственной корпорации «Ростехнологии» и компании Softline. На выставке Научно-технического творчества молодежи (НТТМ-2009) она получила диплом 1 степени и свидетельство за результативное участие. Система награждена дипломом за участие в полуфинале Национальной премии в области инноваций для молодых ученых и специалистов имени В.К. Зворыкина в 2011 году.

Публикации. Материалы диссертации опубликованы в 33 печатных работах, из них 5 статей в журналах, включенных ВАК РФ в перечень ведущих рецензируемых научных журналов и изданий.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения и списка литературы. Общий объем основного текста, без учета приложений — 150 страниц, с учетом приложений — 181. Диссертация содержит 54 рисунка, 25 таблиц и 21 листинг программного кода. Список литературы включает 83 источника.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы диссертационной работы, выделяются объект и предмет исследования, определяются цели и задачи работы, приводится список используемых методов исследования, даются сведения о структуре и объеме диссертационной работы.

В первой главе приводится анализ подходов к описанию семантики веб-сервисов. Анализируются технологии семантического веба, такие как RDF, OWL, OWL-S. Приводится сравнительный анализ различных стандартов: OWL-S, WSMO, BPEL, SAWSDL, CDL. Проводится анализ современных моделей поиска и композиции сервисов, развиваемых в ряде научных работ, а также сравнительный анализ систем поиска и композиции.

Актуальность исследования и разработки модели поиска и композиции веб-сервисов с использованием семантических описаний определяется быстрым ростом числа сервис-ориентированных приложений и необходимостью усовершенствовать и сделать более точным процесс поиска и композиции веб-сервисов.

Искать веб-сервисы можно на сайтах производителей, однако основным источником информации остается реестр UDDI. Поиск в реестре UDDI может быть выполнен по классификационному коду веб-сервиса, присваиваемому по классификаторам (отраслевому классификатору NAICS, классификатору продуктов и услуг UNSPSC, географическому классификатору ISO 3166) при регистрации веб-сервиса в реестре и/или ключевым словам, приписываемым веб-сервису производителем. Классификационный код и ключевые слова характеризуют веб-сервис в целом без разбиения на выполняемые операции, без четкого разграничения и описания входов и выходов сервиса.

Таким образом, поисковая модель веб-сервиса в UDDI представляет собой двойку $\langle KK, KC \rangle$, где KK – это классификационный код сервиса, а KC – множество ключевых слов, описывающих веб-сервис. Запросы на поиск также строятся в соответствии с этой моделью: либо задается классификационная рубрика, по которой находятся все коды, входящие в эту рубрику, и соответствующие им сервисы; либо задается набор ключевых слов, по которому строится эшелонированная выдача сервисов, релевантных запросу; либо задается и то и другое. В любом случае результаты запроса требуют осмысливания с точки зрения соответствия реальным требованиям.

В настоящее время нет общепризнанного реестра для семантических описаний сервисов, хотя имеются предложения о включении такого описания в стандарт UDDI. Предполагается, что семантическое описание сервиса будет выполняться разработчиком в соответствии с моделью $\langle I, O, P, E \rangle$. Это значит, прежде всего, что описываются не сервисы, а связанные с ним абстрактные процессы, планируемые к выполнению в конкретной предметной области, задаваемой концептуальной моделью в виде OWL-онтологии. I, O – это множества типов данных на входе/выходе процесса, задаваемые классами онтологии; P – это множество логических выражений, отражающих предусловия, при выполнении которых возможно успешное выполнение процесса; E – множество логических выражений, отражающих эффекты выполнения процесса с точки зрения изменения состояния внешней среды.

Очевидно, что и запрос на поиск процессов должен строиться в соответствии с моделью $\langle I, O, P, E \rangle$ - фактически пользователь, составляя запрос, должен сформировать семантическое описание необходимого процесса. Оно возможно, если онтологии, используемые при формировании запроса и процесса, идентичны или приводимы друг к другу. Это самый сложный вопрос семантического веба, не относящийся непосредственно к технологии семантических веб-сервисов, но существенно на нее влияющий.

Поиск процессов на основании соответствия их входов/выходов запросу теоретически – только промежуточный этап в решении проблемы поиска сервисов по их семантическим описаниям. Многие исследователи этой проблемы остановились на этом этапе, так как не могли найти место предусловиям и эффектам в алгоритмах поиска. Поэтому, подавляющее большинство исследований, связанных с поиском процессов семантических веб-сервисов, строится на основе модели запроса в виде $\langle I, O \rangle$. Так как I/O – это множества типов данных на входе/выходе процесса, задаваемые классами онтологии предметной области, то вносится семантическая составляющая в поиск процессов. Этим поиск процессов семантических веб-сервисов отличается от поиска веб-сервисов по ключевым словам.

Современные средства выполнения композиции (оркестровки, хореографии) предоставляют комплекс синтаксических структур, которые позволяют разработчику объединять и выполнять операции, предоставляемые сервисом, в рамках некоторого распределенного процесса. Разработчик либо выбирает опубликованные сервисы, точно понимая функциональность предоставляемых им операций, либо выдает неформализованное задание на их разработку в рамках нового. Семантический подход предполагает формализованное описание операций сервиса в виде атомарных процессов. Это формализованное описание может выступать либо как запрос на поиск среди таким же образом выполненных описаний, либо как задание на разработку требуемых операций в новом сервисе.

Если есть уверенность в том, что формализованное описание сервиса точно представляет функциональность его операций, то при разработке сложного распределенного процесса можно не фиксировать операцию конкретного сервиса в рамках разрабатываемой сетевой структуры, а оставить эту работу на этап выполнения. То есть на этапе выполнения система сама найдет необходимый сервис, опираясь на его семантическое описание и, кроме того, может выбирать среди подобных, опираясь на нефункциональные признаки.

При этом можно не ограничиваться поиском единственного атомарного процесса, а, при необходимости, выполнить автоматическую композицию ограниченного числа атомарных.

Запросы, ограниченные моделью $\langle I, O \rangle$ не могут гарантировать требуемую точность поиска. Возникает вопрос, могут ли гарантировать

требуемую точность поиска запросы, формулируемые по модели $\langle I, O, P, E \rangle$. В диссертации анализируется этот вопрос, предлагается расширенная модель семантического описания и алгоритмы поиска, позволяющие использовать максимум информации из семантического описания процесса.

Во второй главе диссертации проведена формализация и описываются модели поиска и композиции процессов сервисов.

Семантическая модель веб-сервиса, в соответствии с рекомендациями W3C, состоит из множества независимых процессов, каждый из которых определяется следующей четверкой множеств:

$$\langle I, O, P, E \rangle$$

В реальной практике имеют место сервисы, не изменяющие состояния внешней среды. К такого рода сервисам относятся поисковые, справочные сервисы, например сервис, возвращающий курс валют. Соответствующий процесс, в случае отсутствия предусловий, определяется парой множеств:

$$\langle I, O \rangle$$

Если предположить, что эта пара однозначно определяет семантику процесса, то это значит, что предполагается использование онтологий с исключительно тонкой классификацией понятий и, соответственно, огромным числом классов и отношений между ними. Как построение таких онтологий предметных областей так и, что особенно важно, практическое использование их, представляет исключительную сложность. Поэтому в диссертации, чтобы снять неоднозначность в описании процесса, предлагается ввести отображение R входов процесса на его выходы – логическое выражение, связывающее входы и выходы, и определять процесс следующей пятеркой:

$$\langle I, O, P, R, E \rangle$$

Отображение R моделируется предикатом местности $|I|+|O|$, т.е. местности равной сумме мощностей множеств. Понимая онтологию как концептуальную модель предметной области, задаваемую логической теорией первого порядка, можно явно определить этот предикат, записывая логическое выражение, связывающее входы и выходы процесса.

Задача поиска процессов веб-сервисов по их семантическим описаниям, очевидно, предполагает, что и запрос на поиск задается в виде аналогичной пятерки, представляющей некий искомый (абстрактный) процесс. Таким образом, задача сводится к сопоставлению двух описаний процессов реального и абстрактного:

$$\langle I_{real}; O_{real}; P_{real}; R_{real}; E_{real} \rangle \text{ и } \langle I_{abs}; O_{abs}; P_{abs}; R_{abs}; E_{abs} \rangle.$$

При «точном» поиске требуется сопоставление, удовлетворяющее условиям

$$I_{abs} = I_{real} \ \& \ O_{abs} \subseteq O_{real}$$

Реально условия более мягкие: для несовпадающих элементов множеств необходимо проверить нет ли между соответствующими классами отношения эквивалентности или «быть подклассом».

При выполнении этих «необходимых» условий требуется доказать истинность утверждений, приведенных в таблице 1.

Таблица 1. Утверждения, требующие доказательства

$P_{abs} \rightarrow P_{real}$	Реальный процесс должен обязательно давать положительный результат с предусловиями искомого процесса. Это значит, что предусловия реального процесса являются более мягкими или их нет вообще.
$R_{abs} \rightarrow R_{real}; R_{real} \rightarrow R_{abs}$	Отображения входов на выходы реального процесса должны быть идентичны отображениям входов на выходы искомого процесса.
$E_{abs} \rightarrow E_{real}; E_{real} \rightarrow E_{abs}$	Изменения в предметной области, производимые реальным процессом, должны быть идентичны изменениям, требуемым искомым процессом.

Получение положительных результатов при выполнении этих доказательств дает уверенность в том, что функциональность искомого (абстрактного) процесса может быть выполнена с помощью реального.

Задача поиска композиций процессов предполагает их последовательный или параллельный запуск. Реально говорить о композиции только двух процессов. После того как композиция найдена, то есть найден составной процесс, он оформляется как простой и может участвовать в дальнейших процессах поиска.

Рассмотрим последовательную композицию процессов. Входом композиции будет считаться вход первого процесса, а выходом – выход последнего. На рисунке 1 приведена иллюстрация композиции.

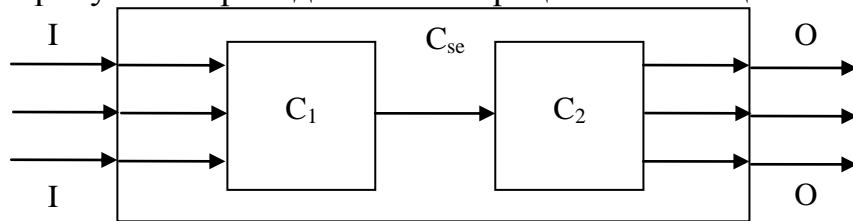


Рис. 1. Последовательная композиция двух процессов.

Теоретико-множественные модели соответствующих процессов:

- $C_1: \langle I_1; O_1; P_1; R_1; E_1 \rangle,$
- $C_2: \langle I_2; O_2; P_2; R_2; E_2 \rangle, I_2 \subseteq O_1;$
- $C_{se}: \langle I_1; O_2; P_1 \& P_2; R_{se}; E_1 \& E_2 \rangle;$

С точки зрения теории баз данных, R_{se} получается с помощью операции эквисоединения по всем входам второго процесса и соответствующим выходам первого с необходимой проекцией. В языке первого порядка это выражается связыванием кванторами существования всех промежуточных переменных. Обозначим $R_{se} = joinOI(R_1, R_2)$

Кандидаты для последовательного соединения находятся из условия:

$$I_1 = I_{abs} \ \& \ I_2 \subseteq O_1 \ \& \ O_{abs} \subseteq O_2;$$

Если такие процессы будут найдены, то необходимо доказать, что

$$P_{abs} \rightarrow P_1 \ \& \ P_2;$$

$$R_{abs} \rightarrow joinOI(R_1, R_2); \ joinOI(R_1, R_2) \rightarrow R_{abs};$$

$$E_{abs} \rightarrow E_1 \ \& \ E_2; \ E_1 \ \& \ E_2 \rightarrow E_{abs};$$

При параллельной композиции процессы выполняются параллельно, при этом их вызов осуществляется одновременно, а результатом служит процесс, входами которого является объединение входов всех процессов, подвергшихся композиции, а выходами – объединение всех выходов. На рисунке 2 отражено ее условное графическое представление.

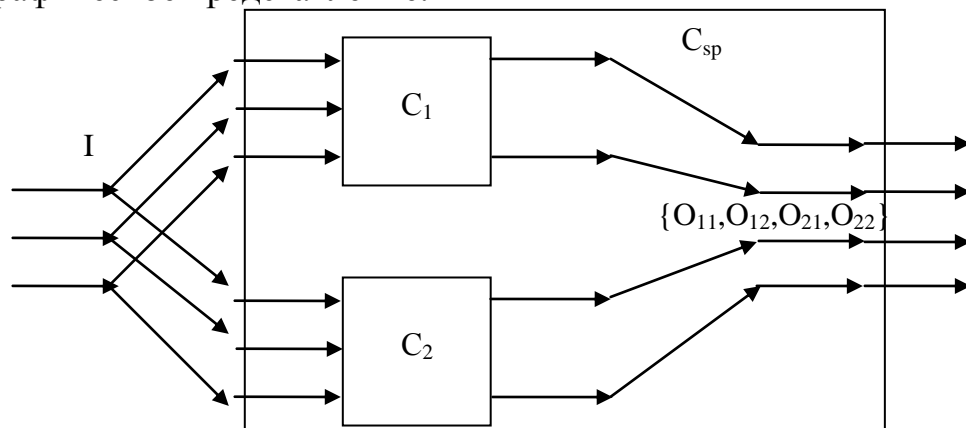


Рис. 2. Параллельная композиция двух процессов.

Теоретико-множественные модели соответствующих процессов:

- $C_1: \langle I_1; O_1; P_1; R_1; E_1 \rangle,$
- $C_2: \langle I_2; O_2; P_2; R_2; E_2 \rangle, \ I_1 = I_2;$
- $C_{sp}: \langle I_1; O_1 \cup O_2; P_1 \ \& \ P_2; R_{sp}; E_1 \ \& \ E_2 \rangle.$

Такая композиция позволяет получить простой процесс, осуществляющий отображение, которое можно назвать «соединением по входу»: $R_{sp} = joinII(R_1, R_2).$

Кандидаты для параллельной композиции находятся из условия:

$$I_1 = I_2 \ \& \ I_1 = I_{abs} \ \& \ O_{abs} \subseteq O_1 \cup O_2;$$

Если такие процессы будут найдены, то необходимо доказать, что

$$P_{abs} \rightarrow P_1 \ \& \ P_2;$$

$$R_{abs} \rightarrow joinII(R_1, R_2); \ joinII(R_1, R_2) \rightarrow R_{abs};$$

$$E_{abs} \rightarrow E_1 \ \& \ E_2; \ E_1 \ \& \ E_2 \rightarrow E_{abs};$$

Решение задачи «точного» поиска процессов и их композиций возможно только при «идеальном» положении дел, когда производители сервисов, составляющие семантические описания процессов, так и потребители, составляющие запросы на поиск, руководствуются одними и теми же онтологиями. Это очень сильное предположение.

Кроме того, как показывает анализ доступных OWL-S описаний процессов, понятия предметной области, используемые в описании, выражаются целыми словосочетаниями, что приводит к идентификации классов онтологии кодами, образованными конкатенацией отдельных слов или их частей. Вполне возможна ситуация, в которой как у производителя, описывающего сервис, так и у потребителя, составляющего запрос на поиск, возникнут трудности в выборе некоторых классов для описания типов входов и выходов процесса/запроса и в результате – неточная спецификация как того, так и другого.

Поэтому в качестве результата, отражающего реальное положение дел, предлагается решение, связанное с комментированием классов онтологии, используемых в качестве типов входов/выходов, набором ключевых слов. Это дает возможность произвести оценку релевантности классов и построить модель, позволяющую оценить релевантность процессов.

В этом случае в диссертации вводится понятие «неточного» поиска, который не накладывает никаких ограничений на анализируемые процессы и производится на основании запроса в виде $\langle I_{запр}, O_{запр} \rangle$, который сравнивается с моделью процесса в виде $\langle I_{проц}, O_{проц} \rangle$. Каждый элемент этих множеств представляется структурой, представляющей расширение исходного класса:

Класс OWL-онтологии

{Класс из ближайшего окружения основного класса }*

{Ключевое слово, комментирующее основной класс

{Слово из ближайшего окружения ключевого слова по WordNet}* }*

Чтобы поставить в соответствие классы множества $I_{запр}$ классу $I_{проц}$ приходится решать задачу о назначениях. Предположим, что $|I_{запр}| = N$, $|I_{проц}| = M$; то есть в запросе ищется процесс с N входами, а сравниваемый с ним процесс имеет M входов, причем, в общем случае, $N \neq M$. Вычисляется матрица $Q = \{q_{ij}\}$, где q_{ij} – степень близости $i = \overline{1, N}$ входа процесса $j = \overline{1, M}$ входу запроса. Алгоритм вычисления этой матрицы приведен в диссертации. Требуется i -ому входу процесса назначить j -ый вход запроса, то есть найти булевы переменные x_{ij} , удовлетворяющие ограничениям: $x_{ij} = 1$, если для i -ого входа процесса назначен j -ый вход запроса; $x_{ij} = 0$, иначе. Условия однозначности назначения определяются неравенствами (1) – каждому входу процесса не более одного входа запроса и (2) – каждому входу запроса не более одного входа процесса:

$$\sum_{j=1}^N x_{ij} \leq 1, \quad i = \overline{1, M} \quad (1)$$

$$\sum_{i=1}^M x_{ij} \leq 1, \quad j = \overline{1, N} \quad (2)$$

Из всех возможных вариантов выбирается тот, который определяет назначения с максимально возможной суммой степеней близости:

$$\max F = \sum_{i=1}^M \sum_{j=1}^N x_{ij} q_{ij} \quad (3)$$

Пусть $\overline{x_{ij}}$ есть решение задачи о назначении со значением критерия \overline{F} . Тогда релевантность RI_{UD} можно оценить формулой:

$$RI_{UD} = \frac{\overline{F}}{N + M - \sum_{ij} \overline{x_{ij}}} \quad (4)$$

Проанализируем выражение (4). Когда множество входов запроса $I_{зап}$ совпадает с множеством входов процесса $I_{проц}$, релевантность максимальна и значение $RI_{3П} = 1$. Если же множества полностью различны, то все $q_{ij} = 0$ и, следовательно, $\overline{F} = 0$, что дает $RI_{3П} = 0$. Во всех остальных случаях релевантность оценивается в интервале $0 < RI_{3П} < 1$, так как:

$$\sum_{ij} \overline{x_{ij}} \leq \min(N, M) \quad \text{и} \quad \overline{F} \leq \min(N, M)$$

Для выходов величина $RO_{3П}$ рассчитывается аналогично.

Решение этих задач позволяет оценить релевантности по входам и выходам, полусумма которых дает оценку релевантности процессов.

Результаты «неточного» поиска предназначены исключительно для человека, который и анализирует предусловия, отображение, эффекты, а говорить о формальных доказательствах не имеет смысла.

В третьей главе описываются основные алгоритмы разработанной системы, на примерах дается представление об их работе и описывается структура и механизмы взаимодействия компонентов разработанной системы.

Работа системы происходит следующим образом. С помощью редактора формируется OWL-S описание искомого (абстрактного) процесса в формате $\langle I_{abs}; O_{abs}; P_{abs}; R_{abs}; E_{abs} \rangle$. Из описания искомого процесса формируется расширенное $\langle I, O \rangle$ описание, представляемое в виде XML-документа, пример которого представлен листингом 1.

Листинг 1. Пример поискового образа процесса

```
<process processID="MONOGRAPH_AUTHOR_PROCESS" processType="Real">
<inputs qty="1">
  <input inputID="_MONOGRAPH">
    <inputType ontoClass="http://127.0.0.1/ontology/books.owl#Monograph">
      <equivalentClass>
        http://xmlns.com/wordnet/1.6/book.owl#Treatise
      </equivalentClass>
      <subClass>http://127.0.0.1/ontology/books.owl#Publication</subClass>
      <superClass>http://127.0.0.1/ontology/books.owl#Book</superClass>
    </inputType>
    <inputKeyWords qty="1">
      <keyWord wrdnWord="Monograph">
        <synonym>Treatise</synonym>
        <hyperonym>Publication</hyperonym>
        <hyponym>Book</hyponym>
      </keyWord>
    </inputKeyWords>
  </input>
```

```

</inputs>
<outputs qty="1">
  <output outputID="_AUTHOR">
    <outputType ontoClass="http://127.0.0.1/ontology/books.owl#Author">
      <equivalentClass>
        http://downloads.dbpedia.org/3.2/en/dbpedia-ontology.owl#Writer
      </equivalentClass>
      <subClass>http://127.0.0.1/ontology/books.owl#Person</subClass>
    </outputType>
    <outputKeyWords qty="1">
      <keyWord wrdnWord="Author">
        <synonym>Writer</synonym>
        <hyperonym>Communicator</hyperonym>
        <hyponym>Coauthor</hyponym>
      </keyWord>
    </outputKeyWords>
  </output>
</outputs>
</process>

```

Оценивается релевантность совпадения множеств входов (I_{abs} , I_{real}), (таблица 2), а также выходов (O_{abs} , O_{real}), что в общем случае требует решения задачи о назначениях, и на основе их оценивается релевантность процессов.

Таблица 2. Оценка близости классов

<i>Abs/Real</i>	<i>Class</i>	<i>Equivalent Class</i>	<i>SubClass</i>	<i>SuperClass</i>
<i>Class</i>	1	1	1	0,5
<i>Equivalent Class</i>	1	1	1	0,5
<i>SubClass</i>	0,5	0,5	0,5	0,5
<i>SuperClass</i>	1	1	1	0,5

Выполнение сложных процедур доказательства соответствия предусловий, отображений и эффектов целесообразно только при условии, что сопоставляемые пары вход/вход, выход/выход и т.д. имеют типы данных, определяемые или одним и тем же классом или сопоставимыми классами одной онтологии, другими словами, оценка релевантности равна 1. В диссертации этот поиск назван «точным». К большому сожалению, это несколько идеализированный вариант. В реальности, одна или несколько пар таковы, что их элементы имеют типы, определяемые классами разных онтологий. Чтобы всё-таки оценить возможность использования реального процесса, в разработанной системе продолжается оценка релевантности пар вход/вход и выход/выход, путем сопоставления комментариев к типам данных, задаваемых в метке <label> OWL-S описания входов/выходов. В результате оценивается релевантность процессов, которая принципиально будет меньше 1, поэтому этот поиск в диссертации назван «неточным». «Неточный» поиск упорядочивает реальные процессы по убыванию оценок их релевантности к искомому. Он не ориентирован на композицию, следовательно, анализируются только пары вход/вход, выход/выход.

Рассмотрим процесс доказательства соответствия предусловий, отображений и эффектов искомого и реального процессов на примере доказательства соответствия предусловий ($P_{abs} \rightarrow P_{real}$). Исходно заданы два OWL-S описания предусловий, выполненные разными людьми, в разное время

и, в общем случае, на разных языках. Сначала требуется сформировать из этих описаний синтаксически правильное утверждение для доказательства, выраженное, например, на логико-математическом языке первого порядка. Далее из онтологии предметной области требуется выбрать все определения, связанные с понятиями, используемыми в предусловиях. Наконец, все это надо представить в виде данных для некоторого «автомата», используемого для автоматического доказательства. В системе в качестве такого автомата используется интерпретатор языка Пролог, следовательно, всю информацию требуется представить в виде синтаксически правильной Пролог-программы. Если выполнение этой программы заканчивается успешно, то соответствие по предусловиям считается выполненным.

На рисунке 3 представлен общий алгоритм работы системы.

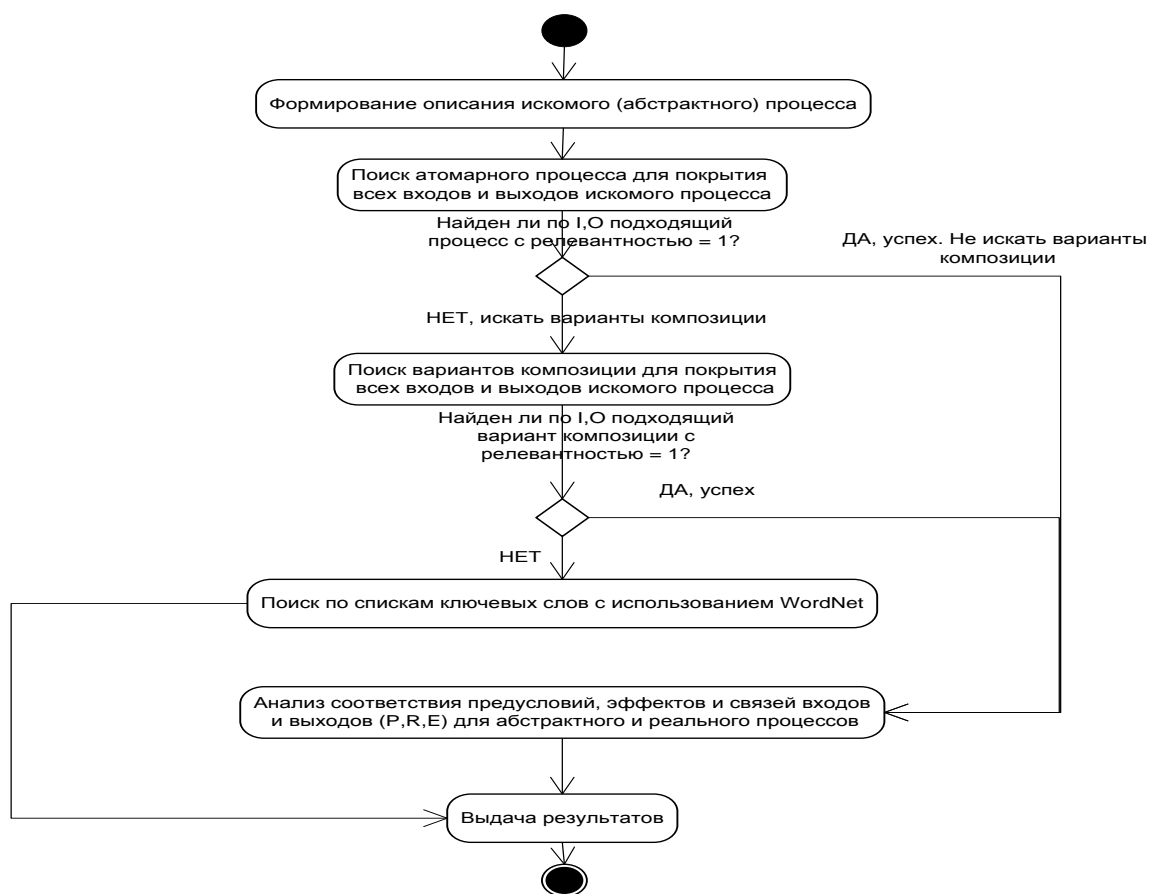


Рис. 3. Общий алгоритм работы системы поиска и композиции

Рассмотрим, для примера, укрупненный алгоритм «точного» поиска, который предполагает нахождение процессов, удовлетворяющих условиям

$$I_{abs} = I_{real} \& O_{abs} \subseteq O_{real}$$

Реально условия более мягкие: для несовпадающих элементов множеств необходимо проверить нет ли между ними отношения owl:equivalentClass, rdfs:subClassOf или отношения diss:superClassOf.

1. Проверяется соотношение мощностей множеств:

$$|I_{abs}| = |I_{real}| \text{ и } |O_{abs}| \leq |O_{real}|$$

При отрицательном результате – переход к анализу следующего процесса.

2. Проверяется совпадение и включение множеств:

$$I_{abs} = I_{real} \text{ и } O_{abs} \subseteq O_{real}$$

При отрицательном результате – переход к шагу 3, при положительном – сохранение идентификатора реального процесса и переход к анализу следующего процесса.

3. Не выполняется требование $I_{abs} = I_{real}$. Строится матрица соответствия входов (таблица 3), например, для 4 входов будет:

Таблица 3. Матрица соответствия входов

I_{abs}/I_{real}	I^1_{real}	I^2_{real}	I^3_{real}	I^4_{real}
I^1_{abs}	0	0	0	1
I^2_{abs}	0	0	0	0
I^3_{abs}	1	0	0	0
I^4_{abs}	0	0	0	0

1 – совпадение классов, 0 – нет.

Вычеркиваем строки и столбцы, где есть 1. Остается матрица (таблица 4):

Таблица 4. Матрица после вычеркивания строк

I_{abs}/I_{real}	I^2_{real}	I^3_{real}
I^2_{abs}	0	0
I^4_{abs}	0	0

Дополняем классы I^2_{real} и I^3_{real} классами, которые находятся с ними в отношении owl:equivalentClass, rdfs:subClassOf или в отношении diss:superClassOf. Предположим, что заключительная матрица будет иметь вид, представленный в таблице 5:

Таблица 5. Заключительная матрица

I_{abs}/I_{real}	I^2_{real}	I^{2e}_{real}	I^{2s}_{real}	I^3_{real}	I^{3e}_{real}	I^{3s}_{real}
I^2_{abs}	0	0	0	0	0	1
I^{2e}_{abs}	0	0	0	0	0	0
I^4_{abs}	0	0	0	0	0	0
I^{4e}_{abs}	1	0	0	0	0	0

Первая строка этой матрицы показывает, что $I^2_{abs} = I^{3s}_{real}$ то есть при описании абстрактного (искомого) процесса в качестве типа данных этого входа был использован класс, который находится в отношении rdfs:subClassOf с классом I^3_{real} . Это допустимо, поэтому выбирается значение 1 из таблицы 1.

Четвертая строка этой матрицы показывает, что $I^{4e}_{abs} = I^2_{real}$ то есть при описании абстрактного (искомого) процесса в качестве типа данных этого входа был использован класс, эквивалентный которому совпадает с классом I^2_{real} . Это допустимо, поэтому 1.

Если хотя бы в одной строке каждого дополнения к основным классам в заключительной матрице содержится 1, то условия по входу выполняются, и следует перейти к проверке условий по выходам (шаг 4). Если нет - переход к анализу следующего процесса.

4. Не выполняется требование $O_{abs} \subseteq O_{real}$. Строится матрица, аналогичная матрице по входам и выполняются аналогичные манипуляции.

5. Если процессов для анализа нет, то поиск заканчивается.

В диссертации подробно расписаны алгоритмы поиска вариантов композиций процессов и алгоритмы «неточного поиска».

Система поиска и композиции процессов сервисов представляет набор взаимосвязанных подсистем, на каждую из которых возложены определенные функции. Они позволяют разобрать пользовательский запрос на составляющие, осуществить генерацию и выполнение запросов к базе семантических описаний, произвести оценку релевантности и многое другое. На рис. 4 представлены основные подсистемы системы поиска и композиции процессов.

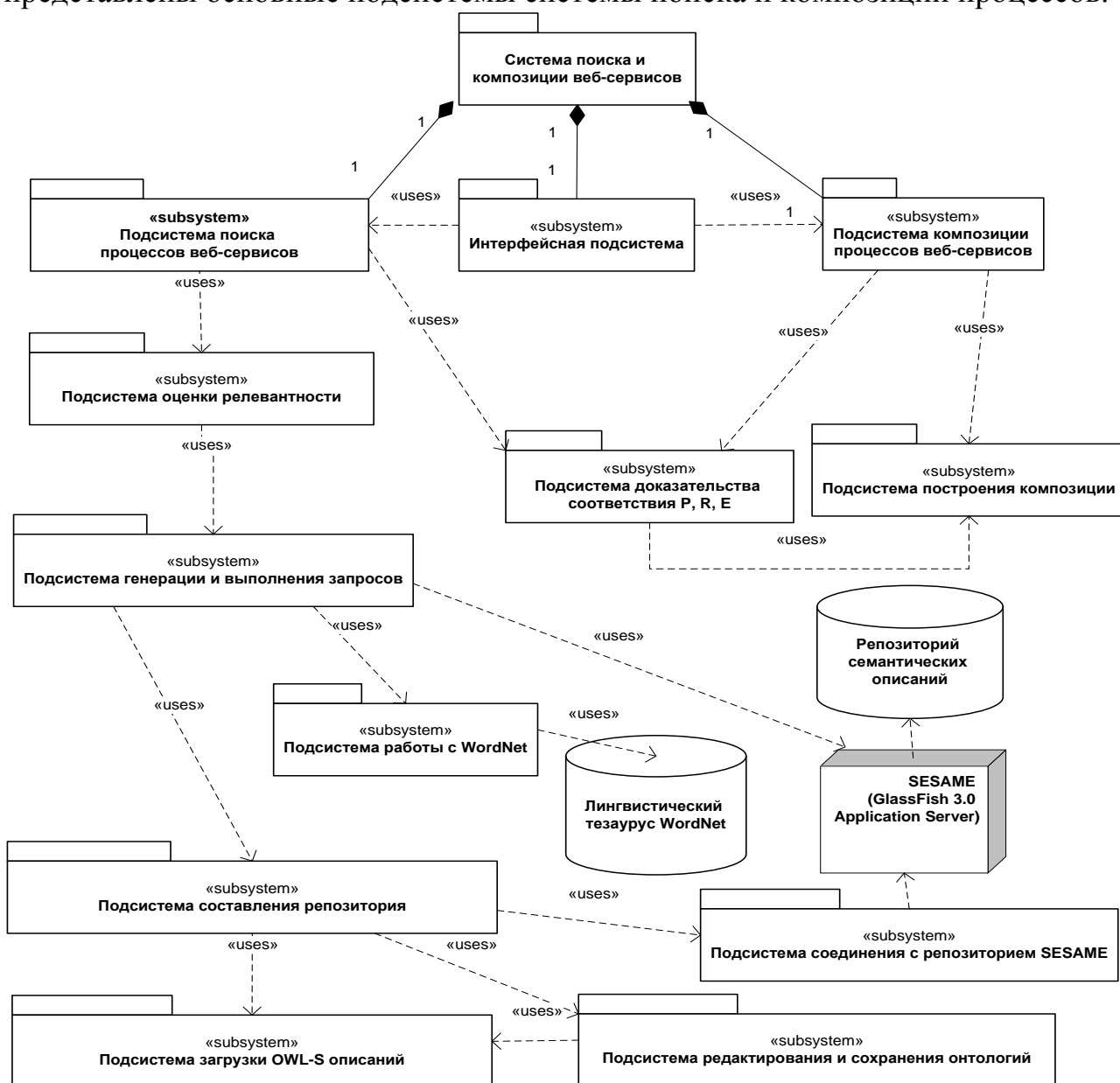


Рис. 4. Общая структура системы поиска и композиции веб-сервисов на верхнем уровне

В четвертой главе приводятся результаты экспериментальной проверки работоспособности разработанной программной системы. Система поддержана грантами «Участник Молодежного Научно-Инновационного Конкурса» (У.М.Н.И.К.), полученными диссертантом в 2009 и 2010 годах, а также грантом по государственному контракту №П2187 от 9.11.2009 в рамках Федеральной Целевой Программы «Научные и научно-педагогические кадры инновационной России на 2009 – 2013 годы», и грантом за второе место в категории «IT-реализация» в номинации «Software» в конкурсе «IT-прорыв» в 2010 году.

Оценка работоспособности системы поиска и композиции производилась на специальном тестовом наборе OWL-S описаний OWLS-TC4 (OWL-S Test Collection Version 4).

Этот набор предназначен для оценки производительности алгоритмов поиска и композиции процессов с использованием OWL-S описаний. Он включает семантические описания сервисов по стандарту OWL-S 1.1 с предусловиями и эффектами, записанными с использованием языка SWRL для 9 различных предметных областей. В него также входит набор из тестовых запросов.

В соответствии с алгоритмами, представленными в главе 3 диссертации, был разработан программный модуль, который содержит 4 метода:

1. Метод, формирующий из OWL-S описаний сервисов списки типов входов/выходов;
2. Метод, который по сформированным спискам типов входов/выходов формирует их ближайшее окружение из классов соответствующих онтологий (в частности, из имеющихся в архиве OWLS-TC4);
3. Метод, формирующий списки ключевых слов, поясняющие типы входов/выходов, из комментариев к описаниям процессов и из комментариев к классам онтологий;
4. Метод, формирующий ближайшее окружение ключевых слов, используя лингвистический тезаурус WordNet.

С помощью этих методов, можно сформировать поисковый образ процесса, представив его в виде XML-документа (см. листинг 1).

Проведена экспериментальная проверка предложенных моделей, алгоритмов и программных средств поиска и композиции процессов веб-сервисов по их семантическим описаниям.

Проводились эксперименты следующих типов:

1. На «точный» поиск процессов в соответствии с алгоритмами, представленными в главе 3. При этом эксперименты проводились с предварительно сформированным файлом XML-документов, представляющих поисковые образы процессов. Поисковые образы процессов, удовлетворяющих требованиям композиционных алгоритмов, искусственно вставлялись в файл XML-документов.

2. На «точный» поиск процессов в соответствии с алгоритмами, представленными в главе 3 без предварительно сформированных поисковых образов процессов.
3. На «неточный» поиск процессов в соответствии с алгоритмами, представленными в главе 3. Эксперименты проводились с предварительно сформированным файлом XML-документов, представляющих поисковые образы процессов и самыми разнообразными запросами. Найденные процессы упорядочивались по убыванию релевантности. Для заданного набора типов входов/выходов запроса проводилось два эксперимента с ключевыми словами и без. В выдачах отсекались все процессы с релевантностью ниже заданной, подсчитывалось количество процессов, попавших в отсечение, количество процессов соответствующих запросу (по мнению экспертов). Из проделанных экспериментов можно сделать вывод, количество процессов, попавших в отсечение, всегда увеличивается, однако при этом увеличивается и шум, иногда процессы с меньшей рассчитанной релевантностью соответствуют запросу, а с большей – нет (по мнению экспертов тестовой коллекции).

Разработанная система поиска и композиции процессов внедрена в компании «Банковский Производственный Центр» в ходе работ по внедрению и интеграции процессинговой системы SmartVista с ERP-системой SAP AG. Компания «Банковский Производственный Центр» (BPC Banking Technologies, www.bpcbt.com, сокращенно БПЦ), занимается реализацией сложных интеграционных проектов в сфере электронных платежей и розничного банковского бизнеса на базе перспективных и апробированных мировой практикой решений. Наряду с интенсивным ведением собственных разработок, для компании важно иметь средства для поиска реализованных решений в этой сфере бизнеса. Реализованные модели, алгоритмы и программные средства позволяют значительно ускорить и упростить процесс поиска и композиции веб-сервисов систем SmartVista и SAP AG. В связи со сложностью и распределенностью используемых в них веб-сервисах, оправданным является применение онтологий используемых понятий, способствующее интеграции систем.

В заключении отражены основные результаты, полученные в данной работе.

В приложениях содержатся копия акта о внедрении системы, состав реализованных программных компонентов системы поиска и композиции, характеристика репозитория семантических описаний. Приводятся примеры семантических описаний веб-сервисов из области «потребительское кредитование», а также концептуальная модель предметной области «потребительское кредитование», созданные в рамках внедрения в «Банковский Производственный Центр».

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Итогом диссертационной работы являются следующие научные и практические результаты.

1. Проведенный сравнительный анализ различных подходов к описанию семантики веб-сервисов показал, что рекомендации W3C разработаны с учетом всех существующих подходов и именно на них следует ориентироваться в исследовании семантического поиска.
2. Для более точного описания семантики веб-сервисов в модель процессов, поддерживаемых сервисом, наряду с предусловиями и эффектами предложено ввести отображение входов процесса на его выходы.
3. Анализ вопросов, связанных с проблемой семантического поиска веб-сервисов, показал, что в них не найдено место для предусловий, эффектов, что собственно и составляет сущность семантического описания. Существующие алгоритмы поиска и композиции ориентированы только на соответствие запроса и реальных процессов веб-сервисов по входам и выходам.
4. В диссертации впервые показано место предусловий, отображений и эффектов в процессе поиска и композиции веб-сервисов. В результате появилась возможность математически точно поставить задачи поиска и композиции процессов с использованием семантических описаний.
5. Разработаны интегрированная модель и алгоритмы поиска процессов веб-сервисов с использованием онтологий предметных областей и общепринятого лингвистического тезауруса WordNet. Это дало возможность осуществлять поиск процессов, использующих в описаниях различные онтологий предметной области, путем использования комментариев (наборов ключевых слов из тезауруса WordNet) к понятиям, являющимся типами входов и выходов искомого процесса (запроса) и реальных процессов.
6. Разработана расширенная модель семантического описания веб-сервиса за счет введения отображения входов на выходы, а также алгоритмы её обработки. Это позволило повысить точность поиска и производить проверку логического соответствия искомого и реальных процессов веб-сервисов.
7. Разработаны модель и алгоритмы использования предусловий, эффектов и отображений входов на выходы при поиске и композиции процессов веб-сервисов. Алгоритм сравнения предусловий, эффектов и отображений входов на выходы искомого и реальных процессов позволяет проверить их логическое соответствие, что приводит к повышению точности поиска и композиции процессов.
8. Разработана структура и механизмы взаимодействия компонентов программной системы поиска и композиции процессов веб-сервисов с

использованием семантических описаний, что дало возможность совместного использования средств программирования Java и Пролог в реализованной программной системе.

9. Проведена экспериментальная проверка предложенных моделей, алгоритмов и программных средств поиска и композиции процессов веб-сервисов по их семантическим описаниям. Проверка подтвердила достоверность полученных результатов.
10. Созданная программная система внедрена в компании «Банковский Производственный Центр». Использование системы способствует повышению эффективности интеграции процессинговой системы SmartVista с ERP-системой SAP AG. Реализованные модели, алгоритмы и программные средства позволяют значительно ускорить и упростить процесс поиска и композиции веб-сервисов систем SmartVista и SAP AG. В связи со сложностью и распределенностью используемых в них веб-сервисах, оправданным является применение онтологий используемых понятий, способствующее интеграции систем.
11. Разработанная программная система внедрена в Министерство образования и науки РФ в рамках государственного контракта №П2187 от 9 ноября 2009 года. Использование системы позволило сделать более точным процесс поиска и композиции веб-сервисов, привело к ускорению разработок и снижению трудоёмкости создания сервис-ориентированных приложений.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в журналах, входящих в перечень ВАК:

1. Климов В.В. Система поиска и интеграции веб-сервисов с использованием семантических описаний // Информационные технологии в проектировании и производстве.-2010.-№1.-С. 73 - 79
2. Климов В.В. Система описания и выполнения композиций семантических веб-сервисов / В.В. Климов, М.Е. Ульянов, П.А. Шапкин и др. // Информационные технологии в проектировании и производстве.-2010.-№4.-С. 64 - 70
3. Климов В.В. Угрозы нарушения информационной безопасности систем поиска и композиции веб-сервисов / В.В. Климов, В.П. Климов // Безопасность информационных технологий.-2010.-№3.-С. 42 - 46
4. Климов В.В. Подход к построению интерфейса композиции веб-сервисов на основе семантических описаний / В.В. Климов, П.А. Шапкин, В.П.Климов и др. // Вестник Воронежского государственного технического университета.- 2010.- №12.- С. 152 - 157
5. Климов В.В. Композиция семантических веб-сервисов / Н.Г. Волченков, В.В. Климов, Б.А. Щукин и др. // Информационно-измерительные и

управляющие системы. Интеллектуальные системы и технологии.-2011.-№ 6,т.9
- С.35-42

Работы, опубликованные в сборниках трудов конференций:

6. Климов В.В. Система поиска и интеграции семантических веб-сервисов // Сборник трудов III Международной научно-практической конференции «Информационные технологии в образовании, науке и производстве», Серпухов, 29 июня - 3 июля 2009 г. Часть 2. стр.169-172

7. Климов В.В. Система поиска и интеграции веб-сервисов, основанная на семантических описаниях // Приоритетные направления инновационного развития.-2009.-№1.- С. 33.

8. Климов В.В. Разработка распределенной многоуровневой системы динамической композиции веб-сервисов / В.В. Климов, Н.Л. Савинов // В кн.: Современные технологии в задачах управления, автоматизации и обработки информации. Труды XIX международного научно-технического семинара. Сентябрь 2010 г., Алушта. – М.:МЭИ, 2010, с. 260.

9. Климов В.В. Разработка математических методов сопоставления эффектов и предусловий семантических веб-сервисов / А.В. Орлов, В.В. Климов, Б. А. Щукин // Современные технологии в задачах управления, автоматизации и обработки информации. Тезисы докладов XX международного научно-технического семинара. Сентябрь 2011 г., Алушта. – Пенза: Изд-во ПГУ, 2011, с.44-45.

10. Климов В.В. Разработка системы семантического поиска веб-сервисов на основании их OWL-S описаний и базовой онтологии предметной области / О.В. Кондратенко, В.В. Климов, Б.А. Щукин // Современные технологии в задачах управления, автоматизации и обработки информации. Тезисы докладов XX международного научно-технического семинара. Сентябрь 2011 г., Алушта. – Пенза: Изд-во ПГУ, 2011, с. 50-51.

11. Климов В.В. Разработка системы автоматического планирования композиции веб-сервисов на основании их семантических описаний и базовой онтологии предметной области / А.А. Хоботов, В.В. Климов, Б.А. Щукин // Современные технологии в задачах управления, автоматизации и обработки информации. Тезисы докладов XX международного научно-технического семинара. Сентябрь 2011 г., Алушта. – Пенза: Изд-во ПГУ, 2011, с. 79-80.

12. Климов В.В. Задача связи сервис-ориентированного приложения и эмулятора мобильного устройства / Климов В.В., Хейн А.А. // VIII Курчатовская молодёжная научная школа. Сборник работ. М.:РНЦ «Курчатowski институт», 2010. С.94.