

Федеральное государственное бюджетное учреждение науки
Институт проблем управления им. В.А. Трапезникова
РОССИЙСКОЙ АКАДЕМИИ НАУК

УДК 004.032

ББК 3.32

На правах рукописи

Душкин Дмитрий Николаевич

**Методы и алгоритмы выбора композиции веб-сервисов в
системах с сервисно-ориентированной архитектурой**

Специальность 05.13.15

«Вычислительные машины, комплексы и компьютерные сети»

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата технических наук

Москва – 2013

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте проблем управления им. В. А. Трапезникова Российской академии наук (ИПУ РАН).

Научный руководитель: Фархадов Маис Паша оглы
доктор технических наук, ИПУ РАН,
заведующий лабораторией №17.

Официальные оппоненты: Выхованец Валерий Святославович
доктор технических наук, доцент,
ИПУ РАН, ведущий научный сотрудник
лаборатории №31;
Шемончук Дмитрий Сергеевич
кандидат технических наук, доцент, МГТУ
МИРЭА, доцент кафедры «Технические и
информационные средства систем
управления».

Ведущая организация: Московский технический
университет связи и информатики

Защита состоится «28» октября 2013 г. в 11 часов на заседании диссертационного совета № 3, Д 002.226.03.

Диссертационный совет создан на базе ИПУ РАН. Адрес: Россия, 117997, Москва, ул. Профсоюзная, д. 65.

С диссертацией можно ознакомиться в библиотеке ИПУ РАН и в сети Интернет по адресу: <http://sky2high.net>.

Автореферат разослан «26» сентября 2013 г.

Ученый секретарь
диссертационного совета,
кандидат технических наук



Кулинич Александр Алексеевич

Общая характеристика работы

Актуальность. Повышение требований к современным информационным системам приводит к появлению новых архитектурных подходов к их созданию. Современные предприятия сталкиваются с проблемами, обусловленными ростом объема данных, необходимостью интеграции с партнёрами и новыми подразделениями, обновлением программной платформы предприятия с сохранением старых, но проверенных временем систем. Одним из современных подходов к решению описываемых проблем является использование *сервис-ориентированной архитектуры* (СОА).

К основным достоинствам СОА относятся: открытость за счёт использования стандартов, независимость от программно-аппаратной платформы, ориентированность на сетевое взаимодействие, использование машинно-читаемых описаний сервисов и др.

В настоящее время СОА широко применяется во многих крупных информационных системах таких предприятий, как Сбербанк, Аэрофлот, РЖД, eBay, Adobe, HP и многих других. При этом за счёт открытости подхода создано несколько платформ для организации СОА, реализующих все необходимые функции и распространяемых на условиях лицензии с открытым исходным кодом. Это позволяет реализовывать новые возможности малым и средним предприятиям.

Основные положения СОА заключаются в применении *композиции (объединения) нескольких автономных веб-сервисов* для решения определенных задач. Современные информационные системы часто используют в своей работе не только локальные сервисы, над которыми возможен контроль, но и внешние, которые с позиции потребителя услуг представлены в виде «чёрного ящика». К тому же, зачастую одну и ту же функцию предоставляют несколько веб-сервисов. Например, среди наиболее востребованных в потребительском секторе можно выделить функции картографии, предоставления информации по географическим объектам, размещения видео, графических и аудио-файлов и т.д.

В связи с этим, при проектировании систем с СОА возникают две связанные задачи: *рациональный выбор* композиции веб-сервисов для решения определенной задачи и формирование набора критериев сравнения веб-сервисов. Под *рациональным выбором* понимается действие лица, принимающего решения, с учетом всей имеющейся у него информации, которое проводит к наиболее *предпочтительным* для него результатам.

Степень разработанности проблемы. В течение недолгой истории развития технологии СОА предлагалось несколько подходов к решению описываемых задач. В нашей стране данный вопрос пока еще мало изучен. В работах Пырлиной И.В. представлены набор однородных критериев и методы выбора наилучшего проекта с СОА на базе методов порогового агрегирования и линейной свертки. В работах Климова В.В., Затесы А.В., Левашова Т.В., Смирнова А.В., Шилова Н.Г. разработаны методы и модели поиска и композиции веб-сервисов с учетом различных предусловий, семантических описаний и предварительно составленных онтологий. В работах Бабошина А.В. и Кашевника А.М. рассмотрен выбор веб-сервисов на основе вычисления интегральной оценки.

Большое количество исследований по сервис-ориентированным вычислениям выполнено в рамках европейского научно-исследовательского проекта SENSORIA, проводившегося в 2005-2010 гг. при поддержке 6-ой Европейской рамочной программы. Работы проекта SENSORIA обсуждаются на ежегодной международной конференции «Service Oriented Computing». Задача выбора композиции веб-сервисов рассматривается в работах G. Canfora, I. Cappello, I. Egambaram, A. Klein, M. Maximilien, D. Menasce. Исследования по теме выбора критериев сравнения проводятся в работах S. Gilmore, M. Tribastone, M. Wirsing.

Смежные темы, касающиеся теории принятия решения и теории важности критериев, детально исследованы в работах Васильева С.Н., Подиновского В.В., Ларичева О.И., Ногина В.Д., Уткина Л.В. и др. В диссертационной работе используются некоторые методы и алгоритмы, предложенные этими авторами.

Существует большое количество стандартов (более 80), описывающих различные аспекты работы систем с СОА, но, тем не менее, вопрос отбора и формирования значений критериев сравнения при проектировании систем с СОА, использующих внешние сервисы, остается открытым.

Несмотря на ценность полученных научных и практических результатов, необходимо дальнейшее исследование и развитие темы формирования и оценки значимых критериев сравнения веб-сервисов, повышения эффективности методов рационального выбора композиций веб-сервисов в рамках бизнес-процессов, а также разработка доступных и прозрачных систем организации рационального выбора композиции веб-сервисов.

Цель работы: теоретическое обоснование, разработка и развитие методов и моделей систем с сервис-ориентированной архитектурой для решения задачи многокритериального выбора композиций веб-сервисов с учётом

предпочтений лица, принимающего решения (проблема рационального выбора).

Для достижения поставленной цели необходимо решить следующие **задачи**:

1. Анализ особенностей проблемы рационального выбора, а также современных методов, моделей и программных средств её решения.
2. Разработка алгоритма рационального выбора композиции веб-сервисов и модели веб-сервиса, включающей измеряемые критерии сравнения.
3. Создание программного комплекса на основе использованных методов, моделей и алгоритмов.
4. Апробация и оценка эффективности разработанного программного комплекса на примере разработки приложения с сервисно-ориентированной архитектурой.

Научная новизна:

1. В результате анализа современного состояния проблемы рационального выбора композиции веб-сервисов сформулирован ряд основных задач, последовательное решение которых позволило качественно и достаточно полно решить обозначенную проблему.
2. Формализован критерий сравнения веб-сервисов «Чувствительность», позволяющий оценить поведение веб-сервисов при постепенном повышении нагрузки, разработана методика вычисления оценки этого критерия на базе методов и алгоритмов машинного обучения.
3. Разработан метод решения задачи рационального выбора композиции веб-сервисов на базе, во-первых, высокопроизводительного алгоритма Васильева С.Н. и Котлова Ю.В. получения множества Парето и, во-вторых, модифицированного автором метода Подиновского В.В. и Уткина Л.В. сужения множества Парето, позволяющего использовать в сравнении неоднородные критерии и информацию об относительной количественной важности критериев. Предложенный метод качественно отличается от предложенных ранее ввиду наличия понятной человеко-машинной процедуры уточнения предпочтений лица, принимающего решения, и отсутствия потерь информации об относительной важности критериев.

Теоретическая и практическая значимость исследования. Теоретическая значимость диссертационного исследования состоит в предложенных оригинальном формальном описании критерия чувствительности веб-сервиса и методике вычисления оценки значения критерия, а также в описании параметрической модели веб-сервисов и их композиции. Сформулированные в работе модели и методы могут быть полезны для развития области

системной инженерии, теории построения крупномасштабных информационных систем, а также в сфере искусственного интеллекта как пример использования методов машинного обучения.

Практическая значимость работы заключается в том, что впервые разработан и выложен в общем доступе на условиях свободной лицензии исходного кода (open-source) программный комплекс, позволяющий вести реестр веб-сервисов и их функций, проводить постоянный мониторинг значений характеристик веб-сервисов. Система отличается от подобных, во-первых, графическим интерфейсом, позволяющим гибко настраивать параметры предпочтений ЛППР за счёт введения понятий потери и компенсации значений по разнородным критериям, и, во-вторых, большим числом критериев сравнения веб-сервисов.

Разработанный программный комплекс полезен для предприятий, использующих сервисно-ориентированный подход при построении информационных систем, в силу того, что позволяет сократить время и, как следствие, финансовые затраты на этапах проектирования и разработки систем с сервисно-ориентированной архитектурой, а также существенно улучшить качество этих работ за счёт понятного обоснования принятых решений по конфигурации систем.

Методология и методы исследования. При проведении исследований были использованы методы математической статистики, машинного обучения, теории принятия решений, теории важности критериев, многокритериальный анализ.

Основные положения, выносимые на защиту:

1. Метод определения чувствительности веб-сервиса.
2. Параметрическая модель веб-сервиса, включающая ряд критериев, характеризующих производительность и надежность, и автоматизированные методы оценки их значений.
3. Интегрируемый программный комплекс, реализующий метод определения чувствительности, параметрическую модель веб-сервисов и их композиции, методику рационального выбора композиции веб-сервисов в рамках задаваемых пользователем условий задач.

Степень достоверности и апробация результатов. Основные положения и результаты диссертационной работы докладывались на следующих конференциях и выставках:

- Международная научно-практическая конференция «Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем» Радиоинфоком-2013, Россия, Москва, 2013 г.
- Международная научная конференция «Современные вероятностные методы анализа, проектирования и оптимизации информационно-телекоммуникационных сетей», Белоруссия, Минск, 2013 г.
- Международная научно-практическая конференция «Современные информационные технологии и ИТ-образование», Россия, Москва, 2012 г.
- Международная научно-практическая конференция «Современные технологии, материалы, оборудование и ускоренное восстановление квалифицированного кадрового потенциала — ключевые звенья в возрождении отечественного авиа- и ракетостроения». Россия, Казань, 2012 г.
- Десятая Всероссийская научно-техническая конференция «Приоритетные направления развития науки и технологии», Россия, Тула, 2011 г.
- Пятая Международная научно-практическая конференция «Интернет и общество» INSO 2011, Грузия, Кутаиси, 2011 г.
- Конференция «Технические и программные средства систем управления, контроля и измерения» УКИ'2010, Россия, Москва, 2011 г.
- Первая Международная конференция «Интеллектуальные технологии и средства реабилитации людей с ограниченными возможностями» ИТСР-2010, Россия, Москва, 2010 г.

Автор является участником гранта компании Google, предоставленном в 2012 году с целью поддержки и развития социально-значимого проекта «Сурдосервер». В 2011 году «Сурдосервер» на выставке «Иннопром 2011» награжден «как самый инновационно-социальный проект».

Связь с планом. Исследования по теме диссертации проводились в соответствии с плановой тематикой Федерального государственного бюджетного учреждения науки Института проблем управления им. В. А. Трапезникова Российской Академии наук. Автор является руководителем гранта РФФИ 12-07-31214 мол_а «Разработка методов, критериев и алгоритмов выбора оптимальной по предпочтениям конфигурации информационной системы с сервисно-ориентированной архитектурой» и соисполнителем работ по гранту РФФИ 12-08-00752-а.

Автор является участником молодежной научной школы под руководством Фархадова М.П. в ИПУ РАН.

Основное содержание работы

Во **Введении** рассматриваются актуальность темы диссертационного исследования, степень изученности и разработанности проблемы. Определяется основная цель работы и перечисляются задачи, решение которых необходимо для достижения поставленной цели. Описываются научная новизна работы, теоретическая и практическая значимость. Приводятся основные положения, выносимые на защиту.

В **Главе 1** выполнен обзор и анализ современного состояния и тенденций развития сервис-ориентированного архитектурного стиля. Приводится общая постановка задачи настоящей диссертационной работы. Описывается круг основных международных регламентирующих документов, касающихся тематики работы. На различных примерах доказываемая актуальность выбранной темы исследований. Приводится обзор и анализ отечественной и иностранной литературы по двум основным темам диссертационной работы.

Сервис-ориентированная архитектура (СОА) — это тип архитектуры распределенных систем, характеризуемый следующими принципами: ориентированность на обмен сообщениями; ориентированность на машинно-читаемое описание; невысокая степень детализации интерфейса доступа; ориентированность на сетевое взаимодействие; независимость от платформ.

Согласно определению, данному организацией W3C, *веб-сервис* — это абстрактное понятие, которое должно быть реализовано конкретным агентом. Агент — это конкретное аппаратное или программное средство, которое отправляет и принимает сообщения, сервис — это ресурс, характеризуемый абстрактным набором предоставляемых функций. Таким образом, определенный веб-сервис может быть реализован несколькими агентами, предоставляющими одинаковый набор функций.

Веб-сервисы имеют различные нефункциональные характеристики: доступность, время обработки запросов, время восстановления после сбоя и т.д. У систем, использующих веб-сервисы, также могут быть требования, описываемые в требованиях к QoS (сокр. Quality of Service, качество услуг) и/или в документе SLA (сокр. Service Level Agreement, соглашение об уровне оказываемых услуг). Таким образом перед инженерами, проектирующими систему с СОА, возникает *многокритериальная задача рационального выбора композиции веб-сервисов, реализующей определенную задачу*.

Для решения задачи рационального выбора композиции, во-первых, *необходимо иметь соответствующую модель веб-сервиса, отражающую*

значимые функциональные и нефункциональные характеристики, и методы нахождения оценок выбранных характеристик, а, во-вторых, метод осуществления рационального выбора композиции веб-сервисов.

В **Главе 2** вводится понятие чувствительности как критерия сравнения веб-сервисов по производительности и надежности при повышении нагрузки. Приводится постановка и решение задачи определения класса чувствительности в терминах задачи машинного обучения. Особое внимание уделено методам оценки обобщающей способности полученного алгоритма классификации. Приводятся результаты вычислительного эксперимента, где в качестве входных данных использовались результаты статистического исследования действующих веб-сервисов.

Автором разработан подход, позволяющий описать веб-сервис в *долгосрочной перспективе*, что важно в ситуациях, когда необходимо быть уверенным в обеспечиваемом уровне производительности при увеличении нагрузки. Такой критерий может быть полезен как при рациональном выборе архитектуры системы с SOA, так и при составлении документов, регламентирующих соглашение об уровне предоставления услуги.

В действительности отдельный сервис при единичных запросах может демонстрировать хорошие показатели времени обработки и надежности, но при небольшом повышении нагрузки количество необработанных запросов и время обработки могут значительно повыситься. С другой стороны, при качественном проектировании сервиса, использовании балансировщиков нагрузки, облачной архитектуры и пр. технологий, увеличение нагрузки не должно существенно влиять на время обработки запросов.

Чувствительность веб-сервиса — критерий, на основе которого может быть оценена возможность обеспечения определенного уровня производительности веб-сервиса при возрастающей нагрузке.

Класс чувствительности — качественная характеристика веб-сервиса, определяемая совокупностью диапазонов значений *признаков*.

Признаки выбираются таким образом, чтобы охарактеризовать измененные времени обслуживания запросов и числа ошибок при увеличении нагрузки.

Ввод разделения веб-сервисов на классы мотивирован большим числом критериев, характеризующих чувствительность и вариациями их возможных значений. Поскольку формирование экспертом набора продукционных правил выделения отдельных классов является трудоёмкой задачей, целесообразно использовать методы *машинного обучения*, которые хорошо подходят для решения задач в случае неприменимости точных алгоритмов.

Формальная постановка задачи определения класса чувствительности. Дано X — множество описаний веб-сервисов, Y — конечное множество классов чувствительности. Существует неизвестная целевая зависимость — отображение $y^*: X \rightarrow Y$, значения которого известны только на объектах конечной обучающей выборки X^m и соответствующим им классам y^m . Требуется построить алгоритм $a: X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$. Говоря менее формально, необходимо, используя предварительно обученный алгоритм классификации, присвоить класс чувствительности новому веб-сервису.

Конечная обучающая выборка X^m и соответствующие классы y^m , вычисляются следующим образом:

1. Составляется случайная выборка веб-сервисов объемом m .
 2. Проводится тестирование выборки по определенному плану.
 3. Вводится описание ряда признаков, характеризующих чувствительность веб-сервисов, и способов их вычисления.
 4. На основе результатов тестов вычисляется матрица «объект-признак» X^m .
 5. При помощи визуализации матрицы X^m проводится эвристическое выделение схожих по значению показателей подмножеств веб-сервисов.
 6. Формируются классы чувствительности на основе упорядоченных по предпочтению подмножеств веб-сервисов.
 7. Составляется множество классов веб-сервисов y^m , соответствующих X^m .
- Пункт 7 при достаточно большом объеме выборки является крайне трудозатратным, поэтому целесообразно автоматизировать этот процесс, проведя кластеризацию данных с числом кластеров, равным числу классов чувствительности. Также результаты кластеризации могут служить индикатором правильности эвристического разделения на классы: если полученные кластеры достаточно обособлены друг от друга, то классы выбраны верно. Кластеризация проводится методом k -средних.

Получив матрицу «объект-признак» X^m и множество соответствующих классов y^m , возможно обучить и оценить алгоритм классификации «с учителем». На Рисунке 1 представлено схематическое описание предлагаемого метода.

Вычисление критериев обучающей выборки. Пусть λ_{max} — число запросов в секунду, отправляемое в последней итерации, S — шаг теста (число, на которое увеличивается количество запросов в секунду в последующей итерации). Общее число итераций в тесте N_{iter} :

$$N_{iter} = \left\lceil \frac{\lambda_{max}}{S} \right\rceil. \quad (1)$$

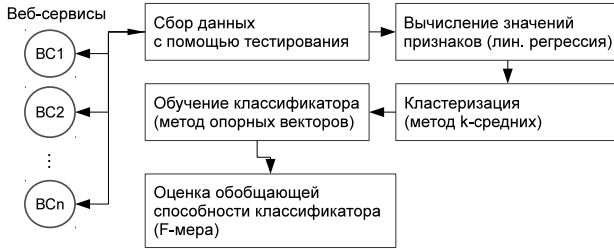


Рисунок 1: Схематическая диаграмма предлагаемого комбинированного метода определения чувствительности

Примем за $\bar{r}_{j,k}^{(i)}$ — усреднённое по трем тестам значение времени обработки запроса, где $i = 0, \dots, n$ — номер веб-сервиса, n — объем выборки, $j = 0, \dots, N_{iter}$ — номер итерации, $k = 0, \dots, j \cdot S$ — номер запроса в рамках j -ой итерации.

Необходимо вычислить матрицу «объект-признак» $\widetilde{X}^m \in \mathbb{R}^{m \times n}$, где m — число объектов, n — количество признаков, которая формируется на основе результатов тестов веб-сервисов. Ряды матрицы соответствуют объектам (веб-сервисам), столбцы — признакам. Обозначим $\mathbf{x}^{(i)}$ i -ый ряд матрицы X . Опишем каждый веб-сервис следующим образом:

$$\mathbf{x}^{(i)} = [\mathbf{t}^{(i)}, \mathbf{d}^{(i)}, \mathbf{e}^{(i)}], \quad (2)$$

где $\mathbf{t}^{(i)}$ — вектор, описывающий изменение среднего времени обработки запросов i -ым веб-сервисом при изменении нагрузки, $\mathbf{d}^{(i)}$ — вектор, описывающий изменение стандартного отклонения, $\mathbf{e}^{(i)}$ — вектор, описывающий изменение числа ошибок.

По каждому веб-сервису сначала вычислим среднее время обработки запросов, стандартное отклонение и число необработанных запросов по итерациям. Далее с целью устранения ограничения на однозначный план теста (шаг теста и максимальное число запросов в секунду) и более экономного описания характеристики чувствительности веб-сервиса произведем аппроксимацию значений признаков по итерациям полиномом первой степени с помощью алгоритма линейной регрессии. Значения коэффициентов полученной регрессионной модели, а также суммарную квадратичную ошибку будем использовать в качестве признаков, характеризующих чувствительность веб-сервиса.

Вычисление значений критериев. Примем за $t_j^{(i)}$ среднее время обработки запросов i -ым веб-сервисом в рамках j -ой итерации:

$$t_j^{(i)} = \frac{1}{j \cdot S} \sum_{k=0}^{j \cdot S} \bar{r}_{j,k}^{(i)}. \quad (3)$$

Воспользуемся регрессионной моделью вида:

$$h_t(j \cdot S)^{(i)} = \theta_{t,0}^{(i)} + \theta_{t,1}^{(i)} \cdot (j \cdot S). \quad (4)$$

Параметр $\theta_{t,0}^{(i)}$ является показателем ординаты точки пересечения прямой с осью ординат, $\theta_{t,1}^{(i)}$ — тангенс угла наклона прямой. Необходимо подобрать такие параметры $\theta_t^{(i)}$, чтобы регрессионная модель наиболее точно описывала исходную зависимость среднего времени обслуживания запросов от числа запросов в секунду. Подбор параметров производится с помощью метода наименьших квадратов. Суть метода в минимизации значения стоимостной функции, равной разности между фактическим значением $t_j^{(i)}$ и вычисленным регрессионной моделью значением $h_t(x)$:

$$\min_{\theta_0, \theta_1} J_t(\theta_t^{(i)}) = \sum_{j=1}^{N_{iter}} (t_j^{(i)} - h_t(j \cdot S)^{(i)})^2. \quad (5)$$

Преобразуем тангенс угла наклона аппроксимирующей прямой, $\theta_{t,1}^{(i)}$, в градусы:

$$\alpha_t^{(i)} = \arctan(\theta_{t,1}^{(i)}) \frac{180}{\pi}. \quad (6)$$

С целью вычисления относительных величин дополним стоимостную функцию постоянным делителем, равным максимальному числу запросов. В итоге, в качестве признаков, характеризующих изменение среднего времени обслуживания запросов при увеличении нагрузки, примем:

$$\mathbf{t}^{(i)} = \left[\alpha_t^{(i)}, \theta_{t,0}^{(i)}, \frac{J_t(\theta_t^{(i)})^{(i)}}{\lambda_{max}} \right]. \quad (7)$$

На Рисунке 2 кругами показаны усредненные по трем тестам данные веб-сервиса картографии «Яндекс.Карты», вертикальные линии — среднее стандартное отклонение времен обработки запросов, линией — аппроксимирующий полином. Видно, что в данном случае полином хорошо описывает тенденцию роста среднего времени обслуживания запросов

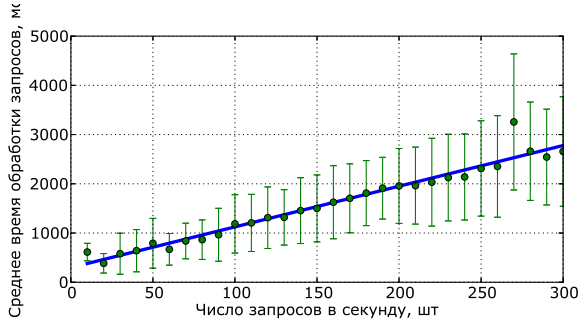


Рисунок 2: Аппроксимация среднего времени обслуживания запросов от числа запросов в секунду. Точки — исходные данные, линия — аппроксимация.

при росте числа запросов в секунду, поэтому значение стоимостной функции $J(\theta_t^{(i)})$ будет мало. Значения признаков для примера на Рисунке 2: $\alpha_t = 83.1, \theta_{t,1} = 298.1, J_t(\theta_t)/\lambda_{max} = 80.9$.

Аналогично проведем вычисления для стандартного отклонения времени обработки запросов:

$$d_j^{(i)} = \sqrt{\frac{1}{j \cdot S - 1} \sum_{k=0}^{j \cdot S} (\bar{r}_{j,k}^{(i)} - t_j^{(i)})^2}, \quad (8)$$

$$\mathbf{d}^{(i)} = \left[\alpha_d^{(i)}, \theta_{d,0}^{(i)}, \frac{J_d(\theta_d^{(i)})^{(i)}}{\lambda_{max}} \right].$$

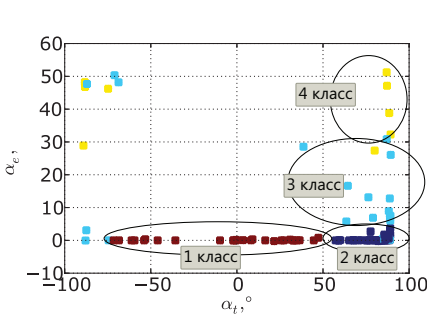
И для числа ошибок:

$$e_j^{(i)} = \sum_{k=0}^{j \cdot S} s_k, s_k = \begin{cases} 1 & \text{если запрос } \bar{r}_{j,k}^{(i)} \text{ не обработан} \\ 0 & \text{если запрос } \bar{r}_{j,k}^{(i)} \text{ обработан} \end{cases}, \quad (9)$$

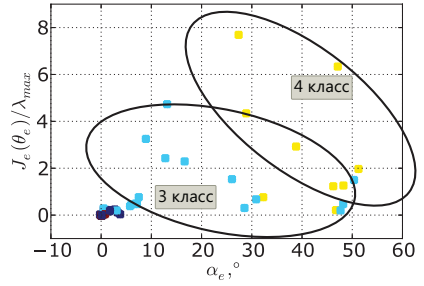
$$\mathbf{e}^{(i)} = \left[\alpha_e^{(i)}, \theta_{e,0}^{(i)}, \frac{J_e(\theta_e^{(i)})^{(i)}}{\lambda_{max}} \right].$$

В итоге каждый веб-сервис описывается следующим набором признаков:

$$\mathbf{x}^{(i)} = \left[\alpha_t^{(i)}, \theta_{t,0}^{(i)}, \frac{J_t(\theta_t^{(i)})^{(i)}}{\lambda_{max}}, \alpha_d^{(i)}, \theta_{d,0}^{(i)}, \frac{J_d(\theta_d^{(i)})^{(i)}}{\lambda_{max}}, \alpha_e^{(i)}, \theta_{e,0}^{(i)}, \frac{J_e(\theta_e^{(i)})^{(i)}}{\lambda_{max}} \right].$$



(a) Значения показателей α_t и α_e



(b) Значения показателей α_e и $J_e(\theta_e)/\lambda_{max}$

Рисунок 3: Значения показателей веб-сервисов

Зададим функцию нормализации z_j :

$$z_j : x_j^{(i)} \mapsto \frac{x_j^{(i)} - E_j[x_j^{(i)}]}{\mu_j(x_j^{(i)})}, \quad (10)$$

где E_j и μ_j — среднее и стандартное отклонение по j -ому признаку соответственно. Примем за $X^m = z(\widetilde{X^m})$ нормализованную по признакам матрицу «объект-значение».

Эвристическое разделение на классы чувствительности. После проведения теста по заданному плану и вычисления матрицы «объект-признак» X^m можно визуализировать полученные данные (Рисунок 3а и Рисунок 3б). Введем эвристическое разделение веб-сервисов по классам чувствительности: от предпочтительной низкой чувствительности к высокой.

Первый класс — низкая чувствительность — характеризуется небольшим изменением среднего времени обслуживания запросов и стандартного отклонения, а также отсутствием необработанных запросов.

Второй класс характеризуется более быстрым повышением среднего времени обслуживания запросов, стандартного отклонения и наличием небольшого числа необработанных запросов.

Третий класс характеризуется быстрым повышением среднего времени обслуживания запросов и стандартного отклонения, а также большим количеством необработанных запросов.

Четвертый класс — высокая чувствительность — характеризуется в первую очередь наибольшим числом необработанных запросов. При неболь-

шом повышении нагрузки быстро растет число необработанных запросов, среднее время обработки запросов и стандартное отклонение.

Стоит отметить, что классы расположены в порядке предпочтения, при этом число ошибок является более важным критерием, чем время обработки запросов. Если предпочтения обратные, то 2 и 3 классы меняются местами.

Автоматизированное разделение на классы чувствительности (кластеризация). С целью проверки обоснованности эвристического разделения веб-сервисов, а также для последующей автоматизации процесса определения класса чувствительности используется кластеризация данных с помощью алгоритма k -средних.

На Рисунке 3а и Рисунке 3б обозначены эвристически выделенные классы чувствительности. Принимая во внимание эвристическое описание классов чувствительности, можно графически обозначить классы на рисунке.

Далее представлено **решение задачи определения класса чувствительности произвольного веб-сервиса** с помощью методов машинного обучения.

В качестве алгоритма классификации используется *метод опорных векторов*. Для решения задачи классификации необходимо выбрать подходящую оценку обобщающей способности алгоритма и выбрать наилучшую комбинацию ядра алгоритма классификации и его параметров по показателям полноты и точности с помощью процедур скользящего контроля (кросс-валидации) и сеточного поиска (англ. *grid search*).

В результате проведения процедур скользящего контроля и сеточного поиска лучший набор параметров наблюдался для ядра радиально базисной функции (функции Гаусса) с параметрами $C = 1000, \gamma = 0.001$. Для контрольного подмножества F -мера = 0.955 ± 0.025 , что говорит о достаточно высокой обобщающей способности классификатора.

Полученные результаты можно представить визуально. Для этого уменьшив размерность пространства признаков с трех до двух с помощью метода главных компонент, настроим параметры классификатора, используя полученное пространство признаков и построим границы принятия решений по классам (Рисунок 4).

В **Главе 3** раскрываются мотивация и постановка задачи рационального выбора композиции веб-сервисов на примере мобильного приложения поиска ближайших больниц. С использованием терминов теорий принятия решений и важности критериев приводится формальная постановка задачи. Подробно описывается реализация каждого пункта разработанного алгоритма решения задачи. Приводится параметрическая модель веб-сервиса.

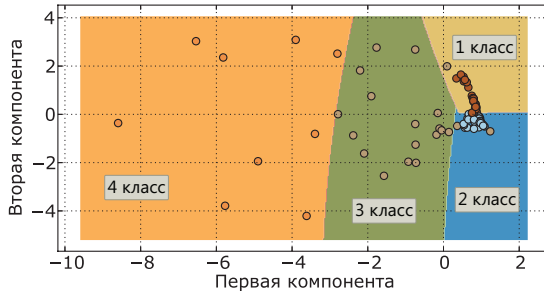


Рисунок 4: Границы принятия решения классификатора по классам

На Рисунке 5 представлен пример композиции сервисов в рамках задачи получения списка ближайших больниц. Каждая из используемых функций может быть реализована множеством веб-сервисов. Например, поставщиками функций геопозиционирования и картографии могут быть сервисы Google Maps, Яндекс Карты, Bing Maps (сервис Microsoft), Nokia Maps и другие.

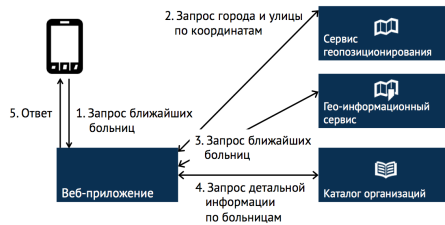


Рисунок 5: Пример композиции сервисов в рамках задачи получения списка ближайших больниц

Формальная постановка задачи рационального выбора композиции веб-сервисов. Введем следующие обозначения:

- Задача T состоит из ряда подзадач t^i , $i = \overline{1, N_t}$, где N_t — число подзадач задачи T .
- S^i , $i = \overline{1, N_s}$ — i -ый сервис в реестре веб-сервисов, где N_s — число веб-сервисов в реестре.
- $S^i.f^j$, $j = \overline{1, N_{fi}}$ — j -ая функция, реализуемая i -ым веб-сервисом, N_{fi} — число функций i -го сервиса.
- $Implement(t^k) = \{S^i.f^j \mid S^i.f^j := t^k\}$ — множество возможных реализаций подзадачи t^k посредством функций сервисов $S^i.f^j$.
- $Implement(T) = \{Implement(t^k) \mid k = \overline{1, N_t}\}$ — множество возможных реализаций задачи T , или возможные композиции сервисов в рамках задачи T .

- $K(S^i.f^j) = (K_1(S^i.f^j), \dots, K_{N_k}(S^i.f^j))$ — векторная оценка j -ой функции i -го веб-сервиса по K_n , $n = \overline{1, N_k}$ критериям.

Для примера определим некоторые критерии выбора функций веб-сервисов:

- $ServiceTime(S^i.f^j)$ — время обработки запроса;
- $Availability(S^i.f^j)$ — доступность функции;
- $Cost(S^i.f^j)$ — стоимость обработки запроса.

Введем нотацию выбора вариантов по предпочтениям:

- $\Omega = \{ServiceTime \succ Availability; Availability = Cost\}$ — обозначение предпочтения критериев выбора лица, принимающего решения (ЛПР). В примере указывается, что время обработки важнее, чем доступность, доступность равноважна стоимости. Можно сделать вывод, что время обработки также важнее стоимости.
- P_Ω — отношение предпочтения между двумя вариантами (в нашем случае, функциями веб-сервисов). Например, запись $S^1.f^1 P_\Omega S^2.f^1$ означает, что функция f^1 сервиса S^1 предпочтительнее функции f^1 сервиса S^2 с учетом информации о предпочтении Ω .
- Функция $S^i.f^{j*}$ реализующая задачу t^k такая, для которой не существует функции $S^i.f^j$, лучшей по отношению P_Ω , называется **недоминируемой**, или **оптимальной по Эджворту-Парето**.
- $Implement(t^k)_\Omega$ — множество оптимальных по предпочтениям реализаций k -ой задачи.
- $Implement(T)_\Omega$ — множество оптимальных по предпочтениям Ω композиций сервисов в рамках задачи T .

$$Implement(T)_\Omega = \{Implement(t^k)_\Omega, k = \overline{1, N_t}\} \quad (11)$$

Таким образом, задачу выбора можно сформулировать следующим образом: необходимо произвести выбор оптимальной по предпочтениям Ω композиции веб-сервисов $Implement(T)_\Omega$ в рамках определенной задачи T .

На Рисунке 6 представлена блок-схема алгоритма, решающего поставленную задачу.

Далее будут рассмотрены некоторые этапы представленного алгоритма.

Этап **формирования множества реализаций подзадач различными веб-сервисами** состоит в предварительном связывании подзадач с функциями веб-сервисов, которые могут реализовать эту подзадачу в рамках композиции.

Параметрическая модель веб-сервиса состоит из *критериев сравнения веб-сервисов*, значения которых можно получить автоматизированным способом: чувствительность, среднее и стандартное отклонение времени обслуживания запроса, среднее время между отказами (средняя наработка

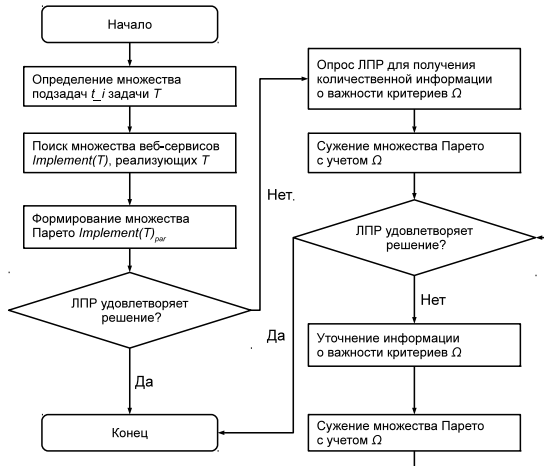


Рисунок 6: Блок-схема алгоритма решения задачи рационального выбора композиции веб-сервисов

на отказ), среднее время на восстановление, доступность, безотказность, стоимость и бесплатный лимит запросов в день.

Для вычисления **времени обслуживания запроса** предложен оригинальный метод, который снижает разброс значений, наличие которого обусловлено следующими факторами: различие между средами передачи информации на пути прохождения сигнала, физическое расстояние между клиентом и сервером, наличие множества различного сетевого оборудования на пути прохождения сигнала.

Кратко метод вычисления **времени обслуживания запроса** можно представить следующим образом:

1. Производится необходимое количество последовательных запросов x_i определенной функции веб-сервиса. Число запросов можно, например, определить с помощью метода доверительных интервалов.
2. Вычисляется среднее время обработки запросов:

$$\mu_x = \frac{1}{m} \sum_{i=1}^m (Waiting(x_i) - Connecting(x_i)), \quad (12)$$

где $Waiting(x_i)$ — время отправки запроса, ожидание ответа и получение первого пакета с заголовком HTTP-ответа, $Connecting(x_i)$ — время установления TCP-соединения с сервером.

3. Вычисляется стандартное отклонение времени обработки запроса:

$$\sigma_x = \sqrt{\frac{1}{m} \sum_{i=1}^m (Waiting(x_i) - Connecting(x_i) - \mu_x)^2} \quad (13)$$

Среднее время между отказами вычисляется по инкрементной формуле, которая лучше подходит при условии работы с большим количеством данных:

$$MTBF_n = \frac{(n-1) \cdot MTBF_{n-1} + w_n}{n}, \quad (14)$$

где w_n — текущее время безотказной работы, n — порядковый номер текущего промежутка безотказной работы веб-сервиса.

Аналогичным образом вычисляется **среднее время на восстановление** (MTTR). **Доступность** (готовность, англ. Availability) и **безотказность** (англ. Reliability) вычисляются по традиционным формулам.

Алгоритм формирования множества Парето. Для малого числа вариантов и критериев достаточно быстро работает простой метод полного перебора, когда все варианты попарно сравниваются и множество Парето формируют из недоминируемых вариантов. Однако при увеличении размерности задачи экспоненциальная сложность алгоритма полного перебора приводит к невозможности завершения формирования множества Парето за разумное время. С целью устранения этого недостатка в работе используется алгоритм, предложенный Васильевым С.Н. и Котловым Ю.В. [Васильев С.Н., Котлов Ю.В. Методы и алгоритмы многокритериальной оптимизации на основе нестрогих ранжировок альтернатив по частным критериям и опыт компьютерной реализации // Проблемы управления и информатики. — 2006. — Т. 1-2. — С. 28–38].

Алгоритм сужения множества Парето на основе информации об носительной важности критериев. Множество Парето обычно содержит большое число вариантов, поэтому с целью сужения полученного множества используются процедуры выявления предпочтений ЛПР. В настоящей работе используется модифицированный подход теории важности критериев, описываемый в работах Подиновского В.В. [Подиновский В.В. Введение в теорию важности критериев в многокритериальных задачах принятия решений. — Москва: Физматлит, 2007] и Уткина Л.В. [Уткин Л.В. Анализ риска и принятие решений при неполной информации. — СПб.: Наука, 2007]. Суть предлагаемой автором модификации в *стандартизации значений критериев*.

Введем основные определения. Пусть x — функция, предоставляемая веб-сервисом. Критерий $K_i(x)$ важнее критерия $K_j(x)$ с заданными положительными параметрами v_i и v_j , если для любой векторной оценки $K(x) = (K_1(x), \dots, K_m(x))$ имеет место соотношение

$$K'(x) \succ K(x), \quad (15)$$

где $K'(x) = (K'_1(x), \dots, K'_m(x))$, причем

$$\begin{aligned} K'_i(x) &= K_i(x) + v_i, \\ K'_j(x) &= K_j(x) - v_j, \\ K'_k(x) &= K_k(x), k = \overline{1, m}, k \neq i, k \neq j \end{aligned} \quad (16)$$

Если критерий K_i важнее критерия K_j с заданными положительными параметрами v_i и v_j , то можно сузить множество Парето заменой векторного критерия $K(x)$ векторным критерием $K^*(x) = (K_1^*(x), \dots, K_m^*(x))$, компоненты которого вычисляются по формулам:

$$\begin{aligned} K_j^*(x) &= v_j K_i(x) + v_i K_j(x), \\ K_k^*(x) &= K_k(x), k = \overline{1, m}, k \neq j, \end{aligned} \quad (17)$$

причем критерии должны быть стандартизованы:

$$K_i(x) = \begin{cases} K_i(x) & \text{если } K_i \text{ имеет положительный тренд} \\ -K_i(x) & \text{если } K_i \text{ имеет отрицательный тренд} \end{cases}. \quad (18)$$

Таким образом векторный критерий K^* получается из K заменой менее важного критерия K_j на линейную комбинацию K_i и K_j с положительными коэффициентами v_j и v_i . Перед вычислением K^* все значения критериев проходят процедуру стандартизации: критерии для которых большее значение является менее предпочтительным (т.е. они имеют отрицательный тренд) домножаются на -1 , остальные критерии сохраняют свои значения.

Далее все новые векторные оценки $K^*(x)$ проходят процедуру получения множества Парето. Оставшиеся варианты и являются результатом выбора по предпочтениям.

В **Главе 4** представлены структура и реализация системы, помогающей осуществить рациональный выбор композиции веб-сервисов, за счёт использования методов и алгоритмов, описываемых в предыдущих главах. Приводится архитектура разработанного программного комплекса, его основных компонентов и деталей их реализации. Демонстрируется эксплуатация системы на примере разработки мобильного приложения поиска ближайших больниц.

На Рисунке 7 представлена архитектура разработанного программного комплекса. Основные функции системы:

1. Осуществление рационального выбора композиции веб-сервисов в рамках определенной задачи.
2. Ведение реестра веб-сервисов и их функций.
3. Мониторинг значений критериев сравнения веб-сервисов.

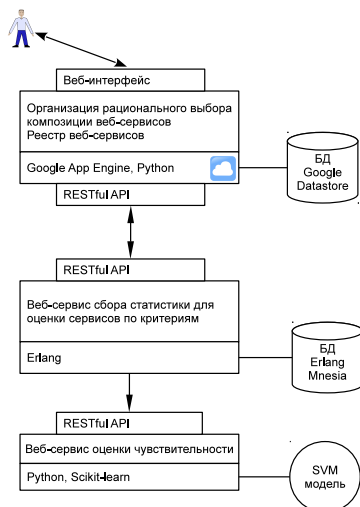


Рисунок 7: Архитектура системы поддержки принятия решений

Программный комплекс реализован как система с сервисно-ориентированной архитектурой, состоящая из трёх основных веб-сервисов.

Веб-сервис рационального выбора композиции и реестр сервисов.

Пользовательский веб-интерфейс компонента позволяет пользователю осуществить следующие операции (Рисунок 8):

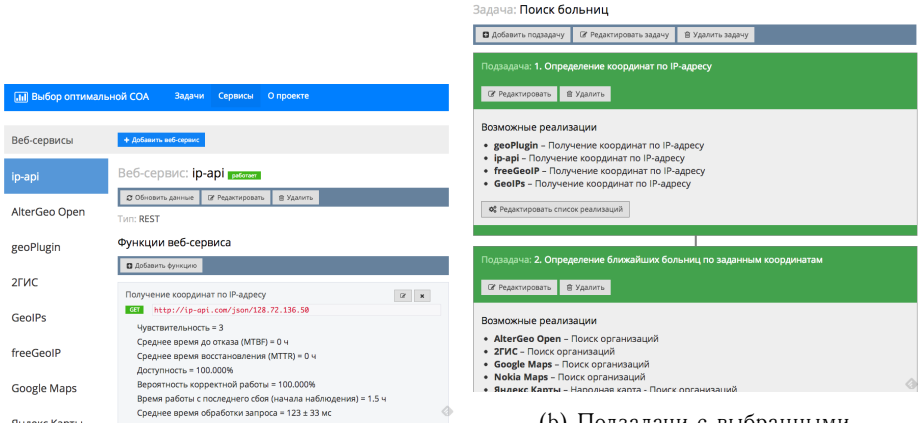
1. Добавление, редактирование и удаление веб-сервисов и их функций в реестре веб-сервисов.
2. Создание моделей задач, состоящих из подзадач.
3. Связывание подзадач с функциями веб-сервисов, которые могут реализовать эту подзадачу.
4. Осуществление человеко-машинной процедуры рационального выбора композиции веб-сервисов.

Веб-приложение доступно по адресу <http://ws-decisioner.appspot.com>.

Веб-сервис сбора статистических данных для получения оценок сервисов по критериям. Компонент периодически обращается к реестру для синхронизации списка рассматриваемых веб-сервисов. Собранные ста-

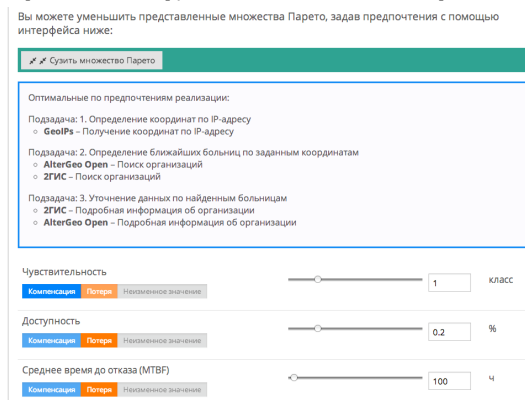
статистические данные используются для вычисления оценок по следующим критериям: чувствительность, среднее и стандартное отклонение времени обработки запросов, среднее время между отказами, среднее время восстановления, доступность, безотказность, время работы, время простоя.

Веб-сервис оценки чувствительности. Данный веб-сервис реализует метод получения оценки чувствительности, описываемый во 2-ой главе диссертационной работы. В качестве API веб-сервис предоставляет одну функцию вычисления оценки чувствительности.



(а) Список веб-сервисов и их функций

(б) Подзадачи с выбранными реализациями



(с) Результат сужения множества Парето

Рисунок 8: Использование системы при проектировании мобильного приложения

Заключение

Итоги диссертационного исследования. В результате проведенных автором исследований разработаны теоретические и практические положения, совокупность которых можно считать новым вкладом в область проектирования и разработки систем с сервисно-ориентированной архитектурой.

1. Выполнен анализ современного состояния проблемы рационального выбора композиции веб-сервисов в системах с сервисно-ориентированной архитектурой. В результате анализа сформулирован ряд основных задач, последовательное решение которых позволит качественно и достаточно полно решить задачу рационального выбора. В диссертационной работе рассмотрены те задачи, известные решения которых либо отсутствуют, либо обладают определенными недостатками.
2. Формализован критерий сравнения веб-сервисов «Чувствительность», позволяющий оценить поведение веб-сервисов при постепенном повышении нагрузки. С помощью вычислительного эксперимента и кластерного анализа полученных данных обосновано выделение 4 классов чувствительности. Сформирована теоретическая и алгоритмическая основа программного комплекса для определения класса чувствительности произвольного веб-сервиса с помощью методов и алгоритмов машинного обучения.
3. Проведен анализ критериев сравнения веб-сервисов, оценку значений которых можно получить автоматизированным способом. Для некоторых критериев приведены итерационные формулы вычисления их оценки, учитывающие специфику вычислений в условиях хранения большого количества данных о моментах измерения времен обслуживания запросов в общей базе данных.
4. Разработана методика решения задачи рационального выбора композиции веб-сервисов, состоящая из следующих основных положений:
 - 4.1. Задача рационального выбора рассматривается как ряд отдельных задач выбора веб-сервисов, наилучших по совокупности выбранных критериев.
 - 4.2. Для получения множества Парето используется эффективный алгоритм, предложенный Васильевым С.Н. и Котловым Ю.В.
 - 4.3. Для сужения множества Парето используется модифицированный вариант метода рационального выбора, учитывающий количественные предпочтения лица, принимающего решения.

- 4.4. Метод рационального выбора дополнен итерационной человеко-машинной процедурой уточнения предпочтений лица, принимающего решения.
5. На основе описываемых в диссертационной работе моделей, методов и алгоритмов разработан программный комплекс на базе технологий Python, Erlang и JavaScript. В качестве технологической площадки размещения системы (хостинга) используется облачное решение вида «платформа как услуга» (Platform as a Service, PaaS) Google App Engine, что позволяет без существенных финансовых затрат предоставлять публичный доступ к системе при условии невысокого числа активных пользователей.

Рекомендации и перспективы дальнейшей разработки темы диссертации. Сервисно-ориентированная архитектура — одно из последних решений многих проблем современных средних и крупных информационных систем, которое поддерживают и развивают практически все современные поставщики программного обеспечения. Несмотря на небольшой возраст технологии (первые стандарты появились в 2003 году), SOA нашла применение как в корпоративном, так и государственном секторах. Министерства обороны многих стран рекомендуют и уже успешно используют SOA как основу для построения сетцентрических систем, обеспечивающих модернизацию военной отрасли.

В связи с этим появляются новые задачи, затрагивающие нерассмотренные аспекты использования SOA, такие как системы с ненадежными сервисами, системы с необходимостью динамической реконфигурации с целью сохранения полнофункционального состояния и др. Представленные в настоящей работе анализ и разработанные модели и методы могут быть использованы в качестве основы для решения подобных задач и задач более высокого порядка.

Список работ, опубликованных автором по теме диссертации

Публикации в рецензируемых журналах ВАК:

1. Душкин Д.Н. Автоматизированное определение классов чувствительности веб-сервисов // Управление большими системами. Выпуск 45. М.: ИПУ РАН, 2013.
2. Душкин Д.Н. Анализ чувствительности веб-сервисов в задаче выбора оптимальной конфигурации систем с сервисно-ориентированной архитектурой

турой // Управление большими системами. Выпуск 40. М.: ИПУ РАН, 2012. С. 164-182.

3. Фархадов М.П., Душкин Д.Н. Сетецентрические технологии: эволюция, текущее положение и области дальнейших исследований // Автоматизация и современные технологии. — 2012. — N 1. — С. 21-29.

Публикации в материалах научных мероприятий, научных сборниках и периодических научных изданиях:

4. Душкин Д.Н. Решение задачи рационального выбора композиции веб-сервисов в системах с сервис-ориентированной архитектурой / Информационные технологии и информационная безопасность в науке, технике и образовании «ИНФОТЕХ – 2013» // Материалы международной научно-практической конференции «ИНФОТЕХ – 2013». Севастополь: СевНТУ, 2013. – ISBN: 978-617-612-037-7 – С. 75-76
5. Душкин Д.Н. О задаче выбора оптимальной конфигурации системы с сервисно-ориентированной архитектурой // Массовое обслуживание: потоки, системы, сети = Queues: flows, systems, networks: материалы международной научной конференции «Современные вероятностные методы анализа, проектирования и оптимизации информационно-телекоммуникационных сетей». Под ред. А.Н. Дудина [и др.], Минск: Изд. центр БГУ, 2013. С. 24–31.
6. Душкин Д.Н., Душкин Н.Д. О проблеме выбора оптимальной по предпочтениям композиции сервисов // Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем. Сборник научных трудов Первой Международной научно-практической конференции «Радиоинфоком – 2013» – М.: МГТУ МИРЭА, 2013 – ISBN: 978-5-7339-0959-2 – ч. 2 – С. 62-66
7. Васильев С.Н., Смирнова Н.В., Суконнова А.А., Душкин Д.Н., Абраменков А.Н. Методы интеллектуализации обучающих систем // Современные информационные технологии и ИТ-образование. Сборник избранных трудов VII Международной научно-практической конференции. Под ред. проф. В.А. Сухомлина. — М.: ИНТУИТ.РУ, 2012. 1050 с. - ISBN 978-5-9556-0140-3. С. 503-519
8. Душкин Д.Н., Смирнова Н.В. Применение инструментов символьных вычислений для проверки решений задач из курсов для конструкторов и технологов в интеллектуальной обучающей веб-системе «Волга» / Сборник докладов международной научно-практической конференции «Современные технологии, материалы, оборудование и ускоренное восстановление квалифицированного кадрового потенциала — ключевые звенья

в возрождении отечественного авиа- и ракетостроения». Казань: Изд-во «Вертолет», 2012. Т. IV. С. 558-564.

9. Душкин Д.Н., Душкин Н.Д. Применение сетецентрического подхода в системах мониторинга природных катастроф // Десятая Всероссийская научно-техническая конференция «Приоритетные направления развития науки и технологий». Тула. 2011. С. 187-189 URL: <http://www.semikonf.ru/archive/?detailID=190>
10. Душкин Д.Н., Каргин А.Н., Фархадов М.П. Новые вехи в развитии систем массового обслуживания // Материалы V международной научно-практической конференции «Интернет и общество» INSO 2011. Грузия, Кутаиси: Кутаисский государственный университет имени Акакия Церетели. 2011. URL: <http://inso.ge/stats2011-2rus.html>
11. Душкин Д.Н., Мясоедова М.А., Фархадов М.П. Современная система управления порталом для создания сервисов в интернет нового поколения // Труды конференции «Технические и программные средства систем управления, контроля и измерения» (УКИ'2010). М.: Учреждение Российской академии наук Институт проблем управления им. В.А. Трапезникова РАН. 2010. С. 276
12. Мясоедова М.А., Душкин Д.Н., Мясоедова З.П., Петухова Н.В., Фархадов М.П. Разработка интернет портала «Сурдосервер» с ресурсами русского жестового языка // Четвертый междисциплинарный семинар «Анализ разговорной русской речи». СПб.: СПИИ РАН, 2010. С. 57-61

Личный вклад. В совместной работе [3] вклад автора заключается в написании разделов «История развития сетецентрического подхода» и «Текущее состояние области сетецентрических технологий». В работе [6] автору принадлежит постановка задачи и её моделирование. В работах [7, 8] автором предложены архитектурная и техническая реализации некоторых компонентов интеллектуальной обучающей системы. В [9] автором предложены методы реализации поставленной задачи. В [10] автором написан раздел о сетецентрических технологиях. В работе [11] автором подготовлены разделы «Типы систем управления содержимым сайта», «Программная архитектура веб-приложения на основе Zend Framework», «Архитектура GreenyCMS». В работе [12] автором выполнены архитектурное проектирование и техническая реализация информационной системы «Сурдосервер».

Зак. 93. Тир. 100. ИПУ РАН.