

На правах рукописи

Проценко Дмитрий Сергеевич

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА
УПРАВЛЕНИЯ СЕТЕВЫМ СООБЩЕСТВОМ ПРАКТИКИ**

Специальность 05.13.18 – Математическое моделирование, численные методы
и комплексы программ

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата технических наук

Москва – 2009

Работа выполнена в Государственном университете – Высшей школе экономики (ГУ-ВШЭ).

Научный руководитель: доктор технических наук, профессор,
Мальцева Светлана Валентиновна

Официальные оппоненты: доктор технических наук, профессор,
Иванников Александр Дмитриевич

кандидат технических наук, доцент,
Восков Леонид Сергеевич

Ведущая организация: Институт системного анализа Российской академии наук

Защита состоится «___» декабря 2009 г. в ____ часов на заседании диссертационного совета Д 212.048.09 при Государственном университете – Высшей школе экономики по адресу: 105187, Москва, ул. Кирпичная, д. 33.

С диссертацией можно ознакомиться в библиотеке Государственного университета – Высшей школы экономики

Автореферат разослан «___» ноября 2009 г.

Учёный секретарь
диссертационного совета
д.т.н., доцент

В.А. Фомичев

Общая характеристика работы

Актуальность проблемы. Развитие Интернет, в частности, на основе концепции Web 2.0, предполагает развитие ресурсов сети, формируемых за счет пользовательского контента. Количество форм таких ресурсов постоянно увеличивается: блоги, форумы, интегрированные среды, которые объединяют блоги и форумы, коммерческие сообщества. Центральное место среди них занимают социальные сети. В последние несколько лет феномен развития социальных сетей привлек в их состав миллионы пользователей. По данным аналитического сервиса comScore World Metrix¹, с июня 2006 по июнь 2008 года у 7 крупнейших мировых социальных сетей резко увеличилось общее количество уникальных пользователей. Например, за 2 года самые крупные порталы MySpace продемонстрировал прирост количества пользователей на 75%, в результате чего его аудитория составила 117 млн. человек, а количество пользователей портала Facebook выросло на 423%, обогнав при этом MySpace, и составив аудиторию в 131 млн. человек. Сейчас насчитываются тысячи сетевых сообществ, которые сильно отличаются друг от друга: имеют различное программное обеспечение и технологии коммуникаций, тематику и направления деятельности, способы формирования аудитории и принятия новых участников. Распространение социальных сетей не обошло стороной и крупные компании. Стали появляться внутрикорпоративные социальные сети.

Сообщества практики (Community of Practice) представляют собой частный случай сетевого сообщества, объединяющего людей, заинтересованных в приобретении и развитии знаний в определенной области и использовании их на практике. Такие сообщества могут объединить практиков, компетентных в определенной предметной области по всему миру. Сообщества практики, как правило, представляют собой регламентированные по разным параметрам сетевые сообщества, имеющие свой подязык и правила общения, их деятельность, также как и деятельность любого сетевого коллектива, не требует необходимости присутствия всех участников в одном и том же месте.

¹ Международная аналитическая компания comScore. - [Электронный ресурс]. - Режим доступа: <http://www.comscore.com>

Одной из организационных форм научного сетевого сообщества является сеть мастерства (Network of Excellence). Объединение участников сети направлено на достижение долгосрочной и прогрессивной интеграции исследовательского потенциала участников, а также на получение научных результатов. Сети мастерства представлены в Европе как специфический инструмент, разработанный в 6-ой и 7-ой рамочных европейских Программах для создания Единого Европейского Пространства².

Сетевые сообщества создаются не только в Интернет, крупные компании также создают свои внутренние сетевые сообщества, как средство управления знаниями внутри компании. Примерами платформ для реализации сетевого сообщества внутри компании могут быть: Microsoft Share Point Portal Server, BEA WebLogic Portal, IBM WebSphere Portal, OracleAS Portal.

Важной особенностью сообществ практики является то, что основой и результатом их существования является формирование и развитие определенной предметной области (домен сообщества практики), представляющей собой область знаний, определяющую интересы членов сообщества, стимулирующую их общение и включающую основные проблемы, решением которых они занимаются. Это делает для сообществ практики неэффективными информационные модели, которые, как правило, являются основой создания программного обеспечения управления сообществами. Такие модели позволяют реализовать механизмы ручного управления контентом сообщества и формирования домена сообщества. Сервисы сообщества ограничены стандартным набором поисковых сервисов.

Для эффективного функционирования сообществ практики необходимы методы, позволяющие моделировать домен сообщества и основные процессы, связанные с его формированием и развитием. Созданные на их основе механизмы управления сообществом и реализующие их инструментальные средства должны учитывать семантические связи между объектами сообщества. Цен-

² Seventh Research Framework Programme (FP7). - [Электронный ресурс]. - Режим доступа: <http://ec.europa.eu/research/fp7>

тральной частью программной среды сообщества в этом случае является сервер отношений - программно-сервисная компонента, позволяющая определять отношения между объектами сообщества. При этом важной особенностью программной среды функционирования сетевого сообщества является интеграция ручных и автоматических методов управления, что делает необходимым присутствие сервисов информационного сопровождения принимаемых решений.

Исследования в области математического моделирования и создания программных средств для управления сетевыми сообществами проводились в последнее время многими исследователями и исследовательскими коллективами и отражены в обширной литературе. Следует отметить работы В. Alger, R. Hamman, M.S. Nilan, J. Preece, H. Rheingold, W.E. Steinmueller, B. Wellman, F. Bell, E. Zaitseva, N. Coates, C. Coppola, E. Neelly, R. Studer, V.R. Benjamins, D. Fensel, P. Spyns, D. Oberle, R. Volz, О.И. Боровиковой, Ю.А. Загоруйко.

Исследования в области моделирования социальных процессов отражены в работах С.А. Айвазяна, Ф.Т. Алескерова, Ю.М. Плотинского, Г.В. Градосельской, О.М. Роя, В.А. Давыденко, Г.Ф. Ромашкиной, С.Н. Чуканова, в том числе исследованиям в области моделирования социальных сетей посвящены работы R. Hamman, M.S. Nilan, H. Rheingold.

Важной частью моделирования деятельности сообщества является представление знаний о предметной области. Исследованиям в области посвящены работы Д.А. Поспелова, Т.А. Гавриловой, В.Ф. Хорошевского, Г.С. Плесневича, Дж. Элти, М. Кумбса, Л.С. Болотовой, А.А. Смолянинова, В.А. Геловани, А.А. Башлыкова, В.Б. Бриткова, Е.Д. Вязилова, П. Джексона, Б.Я. Советова, Д. Марселлуса, В.А. Фомичева, С.О. Кузнецова, В.В. Корнеева, В.В. Девяткова, Ж.-Л. Лорьера, А.И. Змитровича. Учитывая, что в концепции Web 2.0 основой семантической метаинформации являются онтологии, следует отметить также работы специалистов в области онтологического моделирования: N. Guarino, R. Mizoguchi, H. Takeda, M. Uschold, M. Gruninger, T.R. Gruber, C. Vogel, Б.В. Доброва, Н.В. Лукашевич, О.И. Боровиковой, Ю.А. Загоруйко, А.С. Клещева, И.Л. Артемьевой, А.С. Нариньяни, Т.А. Гавриловой.

Важным аспектом создания программного инструментария сетевого сообщества является его реализация как Интернет-ресурса. Методы создания программных средств для порталных решений отражены в работах: N. Guarino, H. Takeda, M. Uschold, M. Gruninger, О.И. Боровиковой, Ю.А. Загорулько, А.Д. Иванникова, М.В. Булгакова, В.З. Ямпольского, А.Ф. Тузовского.

Проведенные исследования позволяют осуществить переход от информационных моделей, как основы для управления функционированием сетевых сообществ практики к новым моделям на основе методов представления знаний, что позволит создать более эффективный инструментарий управления сообществами.

Основной проблемой в этом направлении является создание моделей, практически применимых при разработке программного обеспечения сообществ практики, учитывающих особенности среды функционирования, требования сложившейся практики управления сообществами, особенности источников информации, на основании которой формируется домен сообщества.

Целью работы является совершенствование процессов функционирования сетевых сообществ практики на основе создания математических моделей сообщества и инструментальных программных средств для обеспечения его функционирования.

Достижение этой цели предполагает необходимость решения следующих задач:

- исследование принципов функционирования сетевых сообществ практики;
- исследование методов представления и интеграции знаний для идентификации предметной области (домена) сетевого сообщества;
- создание математической модели сетевого сообщества практики;
- разработка методов и алгоритмов для обеспечения функционирования сетевого сообщества практики.

Объектом исследования являются модели, методы и программные средства, обеспечивающие функционирование сетевого сообщества практики.

Предметом исследования являются способы и методы моделирования сетевого сообщества практики и его предметной области.

Методы исследования. Для решения поставленных задач в диссертационной работе рассматривались существующие модели и методы представления знаний, применены методы теории множеств, системный подход, методы объектно-ориентированного проектирования и программирования.

Научная новизна: В работе предложены:

- интегрированная модель сетевого сообщества практики на основе комбинации инфологической и онтологической моделей;
- метод формирования групп объектов сетевого сообщества на основе анализа сходства частных онтологий объектов;
- методика создания информационного и программного обеспечения для поддержки функционирования сетевого сообщества на основе разработанных моделей и методов.

Практическая ценность работы состоит в:

- математическом, программном и методическом обеспечении для создания и поддержки функционирования сетевого сообщества практики, разработанном на основе предложенных моделей и методов;
- реализации набора основных операций над объектами сообщества в виде программно-сервисного компонента — сервера отношений, позволяющих интегрировать его в программное обеспечение сообществ произвольной структуры.

Апробация работы. Основные положения диссертационной работы докладывались и обсуждались на научно-технических конференциях студентов, аспирантов и молодых специалистов МИЭМ (Москва, 2004 – 2007 гг.), на IV, V, VII Межвузовских ежегодных научно-практических конференциях «Информационные технологии XXI века (Москва, 2003, 2005 г.), на XII Международной студенческой школе-семинаре (Судак, 2004 г.), на XIV Всероссийской научно-методической конференции "Телематика'2007" (Санкт-Петербург, 2007 г.).

Описываемые в диссертации методы моделирования сетевого сообщества практики разработаны и апробированы автором в проекте сети мастерства ГУ-ВШЭ «Модели и методы управления сложными социально – экономическими системами на основе современных информационных технологий». Результаты диссертационной работы были использованы в Институте развития бизнес-информатики ГУ-ВШЭ при выполнении работ по обеспечению логической систематизации и семантической согласованности данных по проектам, реализуемым в предметной сфере, и при разработке новой концепции порталов ВИНТИ и Межгосударственного Координационного Совета стран СНГ по научно-технической информации.

Публикации. По теме диссертационной работы опубликовано 13 научных работ, включая 2 публикации в журналах "Автоматизация и современные технологии" и "Известия высших учебных заведений. Проблемы полиграфии и издательского дела", входящих в Перечень ВАК РФ.

Получено свидетельство Федеральной службы по интеллектуальной собственности, патентам и товарным знакам (Роспатент) о государственной регистрации программы для ЭВМ «Сервер отношений сетевых сообществ практики» (Relationship Server of Network Communities of Practice) № 2009610664 от 28 января 2009 г.

Структура диссертационной работы. Диссертационная работа состоит из введения, 4 глав, заключения, списка литературных источников из 120 наименований и 2 приложений. Работа содержит 4 таблицы и 34 рисунка.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационной работы, сформулированы цели и задачи работы, определена практическая значимость, приведены сведения об апробации и внедрении работы.

В первой главе рассматривается понятие сетевого сообщества, принципы функционирования сетевых сообществ практики, приведена классификация сообществ. На основании проведенного анализа устанавливаются факторы, от которых зависит существование и развитие сетевого сообщества.

Отмечаются проблемы, возникающие при создании и функционировании сообществ практики. Важной особенностью сообществ практики является то, что результатом их существования является формирование и развитие определенной предметной области (домен сообщества практики), представляющей собой область знаний, определяющую интересы членов сообщества, стимулирующую их общение и включающую основные проблемы, решением которых они занимаются.

Проведенный анализ текущего состояния развития сетевых сообществ и технологических приемов, направленных на управление ими, показал, что программная инфраструктура сообщества и заложенные в ее основу модели должны обеспечивать решение следующих основных задач: формирование и развитие домена сетевого сообщества; возможность семантической интеграции участников сообщества на основе единого понятийного аппарата; соответствие тематики и интересов участников сообщества; эффективное управление сообществом. Показано, что решение этой задачи возможно на основе интеграции методов информационного и онтологического моделирования.

Определено понятие объекта сообщества. Таким объектом может быть любой структурный компонент сообщества или протекающих в нем процессов, с которым связана определяющая его информация, на основании которой происходит пополнение или изменение домена сообщества. Показано, что формирование и развитие домена сообщества происходит на основе анализа информации, связанной с объектами сообщества и их взаимодействием.

Проведен анализ функциональной инфраструктуры сообщества практики. Показано, что особенностью функционирования сообщества является отсутствие четко определенных процессов. Основу функционирования составляет совместное использование объектами сообщества определенного набора сервисов. В качестве сервисного ядра программного инструментария сообщества выделена компонента сервера отношений - программно-сервисная компонента, позволяющая определять отношения между объектами сообщества.

Определен перечень задач сервера отношений в контексте решения задач семантической интеграции объектов сетевого сообщества. Основные функции сервера отношений:

- установление отношений между объектами внутри сети;
- установление отношений между внутренними и внешними объектами;
- формирование групп из множества внутренних объектов в соответствии с задачами внутренних процессов сообщества;
- формирование групп из множества внутренних объектов для задач, связанных с взаимодействием с внешней информационной средой.

Определены основные задачи управления функционированием сообщества и возможность их реализации на основе использования сервера отношений.

Во второй главе проведен анализ сетевого сообщества практики как сложной системы и предложена математическая модель сообщества.

Показано, что современное сетевое сообщество практики представляет собой сложную систему, состоящую из технической, программной, информационной, семиотической, лингвистической, семантической и административной подсистем, совместное функционирование которых проявляется в виде процессов создания и потребления информации и знаний в определенной предметной области. При описании сетевого сообщества выделены три основных уровня абстрагирования их описания: информационный, семантический и функциональный, на каждом из которых используются разные модели сообщества.

На информационном уровне рассматривается только совокупность информации, представляющая информационное пространство сетевого сообщества и определяющая занимаемую им область в глобальном информационном пространстве. Эта информация используется в процессах формирования контента сайта сетевого сообщества, его администрирования и информационного обеспечения формирования семантической подсистемы.

На семантическом уровне рассматривается совокупность знаний, которые порождается функционированием сообщества практики. На этом уровне происходит абстрагирование от физической или абстрактной сущности объекта со-

общества и замена его семантической моделью на основе содержания связанного с ним информационного элемента и общих для всех компонент сообщества формальных правил представления ее смысла.

На функциональном уровне рассмотрению подлежат описания алгоритмов операций по обработке информации и знаний, реализующие основные сервисы сообщества. На информационном уровне используется инфологическая модель, позволяющая отразить иерархию информационных объектов сообщества: информационный объект, информационный кластер, информационная область. На семантическом уровне каждому из информационных объектов ставится в соответствие семантический объект: домен объекта, домен группы, домен сообщества. Для представления объектов на семантическом уровне используется онтологическая модель.

На функциональном уровне для каждого из выделенных уровней иерархии определяется сервисный объект: сервис объекта, сервис группы, сервис сообщества. Основные процессы управления сообществом направлены на повышение получаемого эффекта от индивидуальных и коллективных источников информации и знаний для накопления и облегчения доступа к знаниям, повторного или многократного их использования. Использование объектного подхода к формированию инфологической модели сетевого сообщества практики на всех уровнях иерархии позволяет представить ее в форме теоретико-множественной модели вида

$$M_1 = \langle K, A, R, W, Z_{KA}, Z_{RW} \rangle, \quad (1)$$

где:

— $K = \{k_i \mid i = 1, \dots, N\}$ — множество классов объектов сетевого сообщества, где k_i — i -ый класс объекта сетевого сообщества;

— $A = \{a_i \mid i = 1, \dots, N\}$ — множество атрибутов классов $A = \bigcup_{i=1}^N A_i$, где $A_i = \{a_{il_q} \mid l_q = 1, \dots, N_i\}$ — множество атрибутов i -го класса;

— $R = \{r_j \mid j = 1, \dots, M\}$ — множество связей между классами объектов, где r_j - j -ое отношение между компонентами (k_1, k_2) множества K ;

- $W = \{w_j \mid j = 1, \dots, M\}$ — множество атрибутов связей $W = \bigcup_{j=1}^M W_j$, где $W_j = \{w_{jk_p} \mid k_p = 1, \dots, M_j\}$ — множество атрибутов i -го класса ;
- $Z_{KA} : A \rightarrow K$ — отображение множества атрибутов классов на множество классов;
- $Z_{RW} : W \rightarrow R$ — отображение множества атрибутов связей на множество связей.

В этой модели объекты сетевого сообщества рассматриваются как экземпляры классов, а связи между ними как экземпляры связей между классами.

Инфологическая модель является наиболее распространенной в практике основой для разработки базы данных при создании программного обеспечения администрирования сообщества. Ее недостатком для моделирования сообщества практики является то, что она ориентирована, в основном, на поддержку процессов информационного обмена между объектами сообщества, поэтому на ее основе можно построить концептуальные модели задач и процессов.

Для моделирования предметной области в контексте основной цели сообщества практики использован инструментарий онтологического моделирования. На семантическом уровне формируются модели предметных областей разных уровней иерархии на основе онтологий. Важным требованием к модели является обеспечение возможности ее интеграции с инфологической моделью.

Онтологическая модель сообщества на всех уровнях иерархии представлена парой

$$M_2 = \langle \mathbf{D}, \mathbf{C} \rangle, \quad (2)$$

где:

- $\mathbf{D} = \{\mathbf{d}_k \mid k = 1, \dots, K\}$ — множество понятий предметной области, где \mathbf{d}_k — k -ое понятие;
- $\mathbf{C} = \{\mathbf{c}_l \mid l = 1, \dots, L\}$ — множество связей между понятиями, где \mathbf{c}_l — l -ая связь между понятиями.

Связь между инфологической и онтологической моделью сообщества осуществляется за счет внесения в онтологическую модель знаний о структуре объектов сообщества, задаваемой инфологической моделью, а также понятий и связей между ними, выделенных на основе анализа информации объектов со-

общества, задаваемой значениями атрибутов классов объектов, и информации о связях между объектами, задаваемой значениями атрибутов связей между классами. Это позволяет совместить онтологию задач, задаваемую инфологической моделью, с онтологией предметной области.

Для отображения в онтологии структуры инфологической модели выделим: из множества A подмножество $A_K \subset A$ — наименований классов; из множества W подмножество $W_R \subset W$ — наименований связей между классами. Дополнив модель (1) множеством параметров атрибутов классов, B , и множеством параметров атрибутов связей между классами T , выделим из множества B подмножество $B_A \subset B$ — наименований атрибутов классов и из множества T подмножество $T_R \subset T$ — наименований атрибутов связей между классами.

Тогда подмножество понятий онтологической модели, $D_1 \subset D$, описывающее структуру инфологической модели, $D_1 = A_K \cup W_R \cup B_A \cup T_R$.

Подмножество связей между понятиями онтологической модели, $C_1 \subset C$, описывающее структуру связей инфологической модели, включает E_{1K} — множество связей между понятиями-наименованиями классов и понятиями-атрибутами классов, множество связей между понятиями-наименованиями связей и понятиями-наименованиями классов. В инфологической модели объекты сообщества представлены как экземпляры классов. В тоже время инфологическая модель, как правило, наряду с классами, описывающими реальные объекты сообщества, содержит классы вспомогательных объектов, необходимых для описания процессов внутри сообщества, которые не несут информации о предметной области. Выделим в модели (1) множество информационных объектов, обладающих знаниями о предметной области («контейнеры знаний»), в явном виде:

— $O = \{o_{i_o} \mid i_o = 1, \dots, N_o\}$ — множество информационных объектов сообщества, где o_{i_o} - i_o -ый объект,

— $K_I = \{k_{i_o} \mid i_o = 1, \dots, N_o\}$ - множество классов информационных объектов сообщества, где k_{i_o} - класс, к которому принадлежит i_o -ый объект; $K_I \subset K$.

Для множества K_I введем множество механизмов классов: $X_I = \{x_{i_o} \mid i_o = 1, \dots, N_o\}$, где x_{i_o} — механизм i_o -го класса, задающий функцию

$f_{i_o} : o \rightarrow \langle D, C \rangle$ отображения объекта o , принадлежащего к классу k_{i_o} на множества онтологической модели предметной области сообщества D и C , $X_o \subset X$, где X — множество всех механизмов классов модели M_1 . Механизм преобразует инфологическое представление объекта в онтологическую модель объекта $M_o = \langle \mathbf{D}_o, \mathbf{C}_o \rangle$, где $D_o \subset D$, $C_o \subset C$.

Для создания программного обеспечения, поддерживающего функционирование сообщества, важно, чтобы представление сообщества на разных уровнях абстрагирования и иерархии было методологически единым. Учитывая, что формирование понятийного пространства сообщества, так же, как и модерирование объектов сообщества предполагает ручные операции, связанные с отбором объектов и уточнением связей между ними, целесообразно понятийное пространство модели представить также, как и информационное, в формате, который может быть программно реализован в реляционной базе данных.

Дополним множество классов K модели M_1 множеством понятий D онтологии M_2 , рассматривая их как отдельные классы, для которых введем множество A_D , расширяющее выразительные возможности онтологии. Пусть $K' = K \cup D$. Аналогично дополним множество связей между классами R , множеством связей между понятиями онтологии, C , для которых введем множество W_C . Пусть $R' = R \cup C$; расширенное множество атрибутов классов $A' = A \cup A_D$; расширенное множество атрибутов связей $W' = W \cup W_C$. Введем новые отображения: $Z_{K'A'} : A' \rightarrow K'$ — отображение множества атрибутов классов на множество классов; $Z_{R'W'} : W' \rightarrow R'$ — отображение множества атрибутов связей на множество связей.

Тогда модель сообщества практики примет вид

$$M_3 = \langle K', A', R', W', Z_{K'A'}, Z_{R'W'}, O, K_I, X_I \rangle \quad (3)$$

На каждом из уровней иерархии объекты сообщества представляются аналогичной моделью, компоненты которой являются подмножествами множеств модели M_3 . Это позволяет развивать модель сообщества «снизу-вверх», проводя мониторинг изменений информационных объектов и доменов объектов, что позволяет избежать разрыва между доменом сообщества и совокупно-

стью доменов его объектов, что является одной из проблем сетевых сообществ.

В третьей главе описываются задачи функционирования сетевого сообщества практики и предлагаются методы их решения. В частности, решаются такие задачи, как: определение сходства онтологий, определение кластеров объектов сети, включение нового объекта в состав сообщества, модернизация домена сети.

При решении этих задач используется тот факт, что объекты сетевого сообщества практики описаны в нашей работе с использованием онтологических моделей. Приводится подробное описание разработанных методов вычисления близости онтологий объектов, позволяющую количественно определить схожесть двух объектов сетевого сообщества. Предлагается использовать разработанные методы при проектировании программно-сервисной компоненты – сервера отношений. На основании экспериментальных данных предлагается использовать меру сходства Жаккарда (экспериментальные данные приведены на рис. 1.) как наиболее точно определяющую сходство объектов. Она располагается ближе всех мер сходства к прямой, зависящей от процента одинаковых понятий в сравниваемых объектах, полученной при помощи экспертов.

Сравнение мер сходства

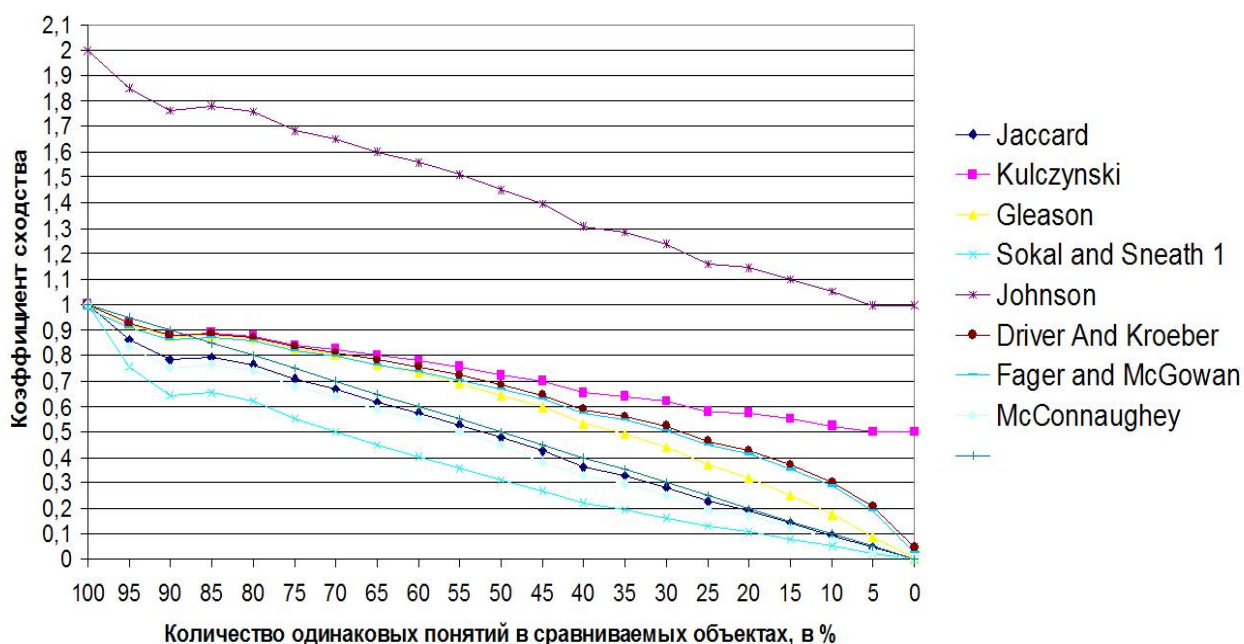


Рис. 1. Результаты эксперимента по выбору меры сходства

Методика установления отношений между объектами сетевого сообщества. Задача установления отношений между объектами сетевого сообщества решается как задача установления степени сходства между локальными онтологиями объектов. Исходными данными являются заранее построенные онтологии объектов сетевого сообщества. Методика установления отношений выглядит следующим образом:

Для всех пар объектов, представленных онтологиями Ω_i и Ω_{i+1} , подлежащих сравнению, выполняются следующие действия:

1. Определяются множества общих элементов, при этом учитываются как понятия, так и связи между понятиями:

$$a_n = D_{\Omega_i} \cap D_{\Omega_{i+1}} \text{ и } a_c = C_{\Omega_i} \cap C_{\Omega_{i+1}}.$$

2. Вычисляются степени общности двух множеств по формуле:

$$S_{Jac} = \frac{a}{p_1 + p_2 - a}.$$

Общая онтология сетевого сообщества практики строится на основании доступной информации о пользователях включая их профили и отражает результаты их деятельности, выраженные в научных работах, проектах, всевозможных публикациях и др. Общая онтология Ω может быть сформирована с применением двух подходов: автоматическом, когда происходит анализ всей информации о пользователях сообщества без вмешательства эксперта, и полуматематический, когда эксперт в режиме диалога с ЭВМ принимает решения о вариантах построения онтологии. Отметим, что второй подход реализуется как автоматизированный и зачастую эксперту необходимо выбирать из предложенных вариантов. Предполагается, что при построении общей онтологии происходит анализ множества документов, в которых заключается информация о пользователях и связанная с их деятельностью. При анализе документов следует обеспечить минимальную дублируемость информационных элементов и минимизировать связи между информационными элементами. Анализ документов предполагает разбор каждого предложения, с выделением значимых информа-

ционных элементов и связей между ними. В результате документ разбивается на n групп, где n – количество предложений в документе.

Метод формирования групп объектов. Задача разбиения объектов сетевого сообщества практики, которые представлены в виде документов, на группы объектов схожих по тематике, представляет собой задачу кластеризации документов в пространстве весовых коэффициентов терминов. Кластеризация документов – процесс выявления семантически похожих групп документов, при этом, в отличие от классификации, никакие характеристики групп не задаются заранее.

Учитывая специфику представления данных в модели предметной области сетевого сообщества, становится возможным получать более точную кластеризацию. Для этого предлагается использовать дополнительную кластеризацию по связям между понятиями объектов. Результаты кластеризации наборов объектов по понятиям и дополнительной кластеризации по связям приведены на рис. 2.

Кластеризация объектов сетевого сообщества

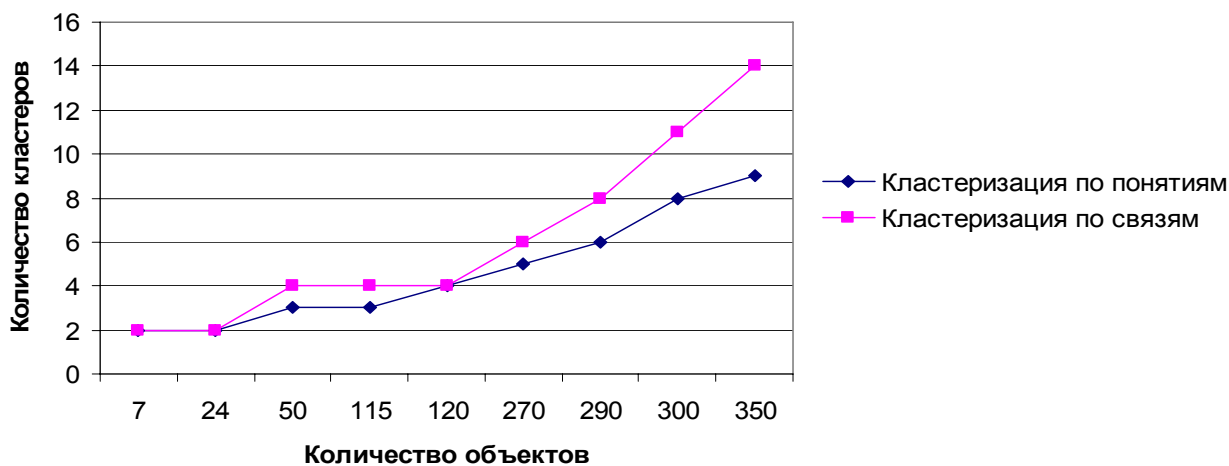


Рис. 2. Кластеризация объектов сетевого сообщества по понятиям и связям

Эксперименты показывают, что использование в качестве параметров кластеризации связей между понятиями, позволяет получить кластеры, в которых среднее значение сходства объектов выше по сравнению с кластеризацией только с использованием понятий, при этом границы кластеров хорошо разли-

чимы. Это происходит в тех случаях, когда при дополнительной кластеризации образуются дополнительные кластеры.

Методика классификации объектов сетевого сообщества практики.

Исходными данными являются заранее построенные онтологии объектов сетевого сообщества практики. Как правило, при классификации используются онтологии кластеров Ω_k сетевого сообщества и онтологии внешних Ω' и внутренних объектов Ω_o . Компоненты онтологий Ω_k , Ω_o принадлежат общей онтологии сетевого сообщества Ω . Методика представима в виде последовательности следующих шагов:

1. Ввод объектов, представленных онтологиями Ω_k или Ω_o , подлежащих классификации и указание порогового значения Rn .

2. Задание групп, существующих в сообществе, к которым необходимо отнести объекты. Каждая группа представлена онтологией Ω_k .

3. Выбор пары двух множеств из пунктов 1 и 2.

4. Определение множества общих элементов, учитываются понятия и связи между понятиями:

$$a_n = D_{\Omega_k} \cap D_{\Omega_o} \text{ и } a_c = D_{\Omega_k} \cap D_{\Omega_o}.$$

5. Вычисление степени общности двух множеств по формуле

$$S_{Jac} = \frac{a}{p_1 + p_2 - a}.$$

6 Сравнение полученного значения S_{Jac} с Rn .

7. Если полученное значение S_{Jac} больше Rn , то отнесение объекта к текущему кластеру, в противном случае выбирается следующий кластер и переход к шагу 4.

8. Проверка просмотрены ли все пары, если нет, то на шаг 3.

Модернизация домена сети. В процессе функционирования сетевого сообщества практики неизменно происходит его развитие за счет творческого процесса его участников. В сообществе появляются новые результаты научной и практической деятельности его участников, какие то результаты устаревают и теряют свою актуальность. Подобный процесс развития сетевого сообщества

неизбежно приводит к изменению его предметной области. Домен сетевого сообщества всегда должен находиться в актуальном состоянии, и отражать процессы, происходящие в сетевом коллективе. Обязательной является возможность учитывать вектор развитие сетевого сообщества. Это позволит отслеживать развитие тематической направленности работы сообщества со временем и, возможно, будет способствовать правильному выбору направлений научных тематик.

Возможны два направления модернизации домена сетевого сообщества практики:

1. Без изменения структуры. Суть подобной модернизации заключается в изменении весов понятий и связей общей онтологии домена сети. Подобный подход может быть применим, когда происходят дискуссии, рецензии, комментарии материалов внутри сообщества без добавления новых работ.

2. С изменением структуры. Стоит отметить, что при подобном подходе может в общей онтологии домена сети происходить два вида изменений, связанных как с добавлением или удалением понятий, так и добавлением или удалением связей между понятиями. Ситуация, когда необходимо изменять онтологию сетевого сообщества, чаще всего происходит при добавлении нового члена сети или новых материалов.

Добавление новых объектов в сетевое сообщество практики и связанное с этим изменение общей онтологии домена сети происходит с учетом анализа новых объектов на предмет соответствия области деятельности сообщества. Предварительно происходит оценивание нового объекта с использованием меры близости.

Первым этапом является построение онтологии нового объекта на основе имеющейся информации о нем.

На втором этапе происходит сравнение онтологии домена сети с онтологией нового объекта. Осуществляется поиск общих информационных элементов и связей между ними для сравнения. Существует экспериментально полу-

ченный пороговый коэффициент Rn , ниже которого включение нового объекта в сообщество является не целесообразным.

Третий этап происходит, если сравнение двух онтологий дало значение больше порогового коэффициента Rn . На этом этапе происходит изменение структуры общей онтологии с учетом онтологии нового объекта – изменяются веса понятий и связей общей онтологии, и как правило, добавляются новые понятия и связи.

Заключительным этапом является отнесение нового объекта к одной из групп, существующей внутри сетевого сообщества. На этом этапе происходит просмотр общих понятий онтологий, выявленных на втором этапе, и подсчет их количества по принадлежности к каждой из существующей групп. Далее происходит отбор групп, где число совпавших терминов получилось максимальным. Таким образом, мы выявляем заранее кластеры, к которым может быть близок новый объект. Далее происходит сравнение онтологии нового объекта с онтологией кластера. Если сравнение двух онтологий дали значение больше порогового коэффициента Rn_2 , то происходит отнесение нового объекта к этому кластеру. В противном случае, происходит сравнение со следующим кластером, где присутствует наибольшее количество общих понятий. Надо отметить, что возможна ситуация, когда новый объект не был отнесен ни к какой группе, но решение о принятии объекта в сеть принято. Становится понятным, что сформированные ранее группы должны быть изменены с учетом нового объекта. Тогда администратор может принять решения о новой кластеризации объектов сетевого сообщества практики, с целью обеспечения актуального состояния кластеров.

В четвертой главе описывается программное обеспечение для создания и поддержки функционирования сетевого сообщества практики, приводятся методические рекомендации по его применению для создания и поддержки функционирования сетевых сообществ практики. Предложена архитектура и программная реализация сервера отношений.

Общая архитектура портала сетевого сообщества, использующего сервер отношений (рис. 3) состоит из 4-х основных компонент: сервера приложений, обеспечивающего функционирование портала сетевого сообщества практики; сервера онтологий, позволяющего работать с онтологиями сетевого сообщества, формализованных на естественном языке; Windows-приложения, осуществляющего семантический анализ объектов сетевого сообщества и построение онтологий; сервера отношений, осуществляющего поддержание функционирования сетевого сообщества практики.

Логически архитектуру портала сетевого сообщества, использующего сервер отношений можно разделить на три взаимосвязанных уровня: уровень интерфейса, уровень бизнес-логики и уровень данных.

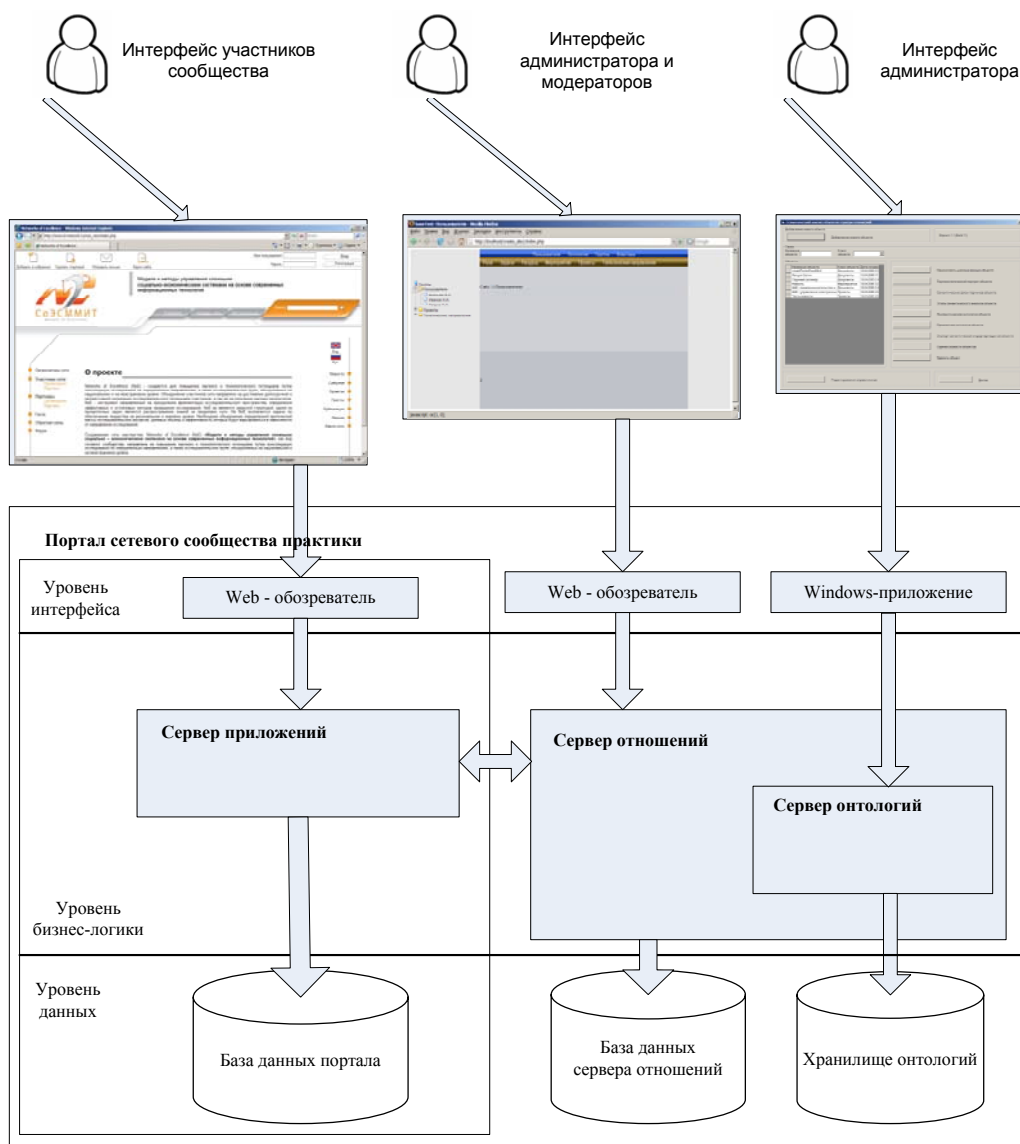


Рис. 3. Общая архитектура портала сетевого сообщества, использующего сервер отношений

Дается описание этапов проектирования и программной реализации сервера онтологий и сервера отношений сетевого сообщества практики. Поясняются основные программные интерфейсы. Приведены данные экспериментального исследования разработанных моделей и методов.

Основные результаты и выводы

Разработанные в диссертации методы, находясь в русле доминирующих сегодня направлений исследовательской активности в сфере сетевых сообществ и онтологического моделирования, обеспечивают совершенствование теоретической базы моделирования сетевых сообществ практики. Разработанный по результатам исследований сервер отношений, реализующий предложенные методы формирования и обработки семантических данных объектов, может служить основой для организации взаимодействия членов сетевого сообщества в различных предметных областях. В ходе исследований, проведенных в настоящей диссертационной работе, получены следующие основные результаты:

1. На основе проведенного исследования принципов функционирования сетевых сообществ практики сформулированы особенности современного сетевого сообщества практики, представляющего собой сложную систему, состоящую из технической, программной, информационной, семиотической, лингвистической, семантической и административной подсистем, совместное функционирование которых проявляется в виде процессов создания и потребления информации и знаний в определенной предметной области.

2. На основе проведенного анализа сообщества практики как сложной системы определены уровни абстрагирования: информационный, семантический и функциональный, на каждом из которых сетевое сообщество практики можно представить семейством моделей, описывающих его поведение и структуру в рамках независимых между уровнями законов, принципов и понятий.

3. На основе проведенного исследования методов представления и интеграции знаний для идентификации предметной области сетевого сообщества,

сформулированы особенности формирования знаний о предметной области сетевого сообщества практики, показана взаимосвязь моделей предметной области сообщества с моделями задач и процессов сообщества.

4. На основе проведенного анализа методов моделирования информационной архитектуры и знаний о предметной области сообщества предложена оригинальная интегрированная математическая модель сетевого сообщества практики на основе теоретико-множественного представления инфологической модели сообщества и онтологической модели предметной области сообщества. Преимуществом модели по сравнению с существующими является возможность использования единой модели представления сообщества на разных уровнях абстрагирования, что обеспечивает использование единой информационной архитектуры при разработке программного обеспечения для поддержки функционирования сообщества.

5. На основе анализа задач управления функционированием сетевого сообщества выделены основные задачи, решение которых направлено на развитие домена сообщества. Предложены методы их решения, в том числе, разработан метод формирования групп объектов сетевого сообщества на основе анализа сходства частных онтологий объектов, обеспечивающий более высокую точность определения сходства по сравнению с существующими методами.

6. На основе проведенного анализа функциональной инфраструктуры сообщества практики определено сервисное ядро программного инструментария сообщества, выделена компонента сервера отношений. Разработана методика создания сервера отношений и предложен программный инструментарий реализации набора основных операций над объектами сообщества в виде сервера отношений, позволяющих интегрировать его в программное обеспечение сообществ произвольной структуры.

Публикации по теме диссертации

Публикации в журналах, входящих в Перечень ВАК

1. С.В. Мальцева, Д.С. Проценко. Серверы отношений сетевых сообществ практики на основе онтологических моделей // Автоматизация и современные технологии. – 2008. – № 3. – С. 26-29. (вклад автора 0,16 п.л.).

2. В.В. Никитин, С.В. Мальцева, Д.С. Проценко. Интегрированная система разработки образовательных стандартов на основе технологий управления знаниями // Журнал "Известия высших учебных заведений. Проблемы полиграфии и издательского дела". – 2009. – № 1. – С. 51–58. (вклад автора 0,1 п.л.).

Другие публикации и тезисы докладов

3. С.В. Мальцева, Д.С. Проценко. Программный комплекс для создания типовых сервисов и компонент в составе Интернет-ресурсов // Информационные технологии XXI века. Материалы IV Межвузовской ежегодной научно-практической конференции. – Москва, 2003. – С. 24–25. (вклад автора 0,05 п.л.).

4. Д.С. Проценко, С.В. Мальцева. Программные средства информационного сопровождения WEB-ресурсов Интернет // Информационные технологии XXI века. V Межвузовская научно-практическая конференция: Материалы конференции. – М., ИИТ МГУС, 2003. – 379с. – С. 64–68. (вклад автора 0,1 п.л.).

5. Д.С. Проценко. Автоматизация сопровождения WEB-ресурсов Интернет // Научно-техническая конференция студентов, аспирантов и молодых специалистов МИЭМ. Тезисы докладов. – М.: МИЭМ, 2004. – 617 с. – С. 269–270.

6. Д.С. Проценко. Управление информационным содержанием WEB – ресурсов Интернет // «Новые информационные технологии». Тезисы докладов XII Международной студенческой школы-семинара. – М.: МГИЭМ, 2004. – 421с. – С. 324–325.

7. Д.С. Проценко. Онтологическое моделирование Интернет-ресурсов // Научно-техническая конференция студентов, аспирантов и молодых специалистов МИЭМ. Тезисы докладов. – М.: МИЭМ, 2005. – 429с. – С.67–68.

8. Д.С. Проценко. Использование онтологий для проектирования Интернет-ресурсов // Информационные технологии XXI века. Материалы VII Межвузовской научно-практической конференции. – М., 2005. – 302 с. – С. 84–87.

9. Д.С. Проценко. Проектирование и расширение онтологий программно-информационных продуктов // Научно-техническая конференция студентов, аспирантов и молодых специалистов МИЭМ. Тезисы докладов. – М.: МИЭМ, 2006. – 440с. – С. 128–129.

10. Д.С. Проценко. Методы построения онтологий предметных областей программно-информационных продуктов // Искусственный интеллект в XXI веке. Решения в условиях неопределенности: сборник статей IV Международной научно-технической конференции. – Пенза, 2006. – С. 48–49.

11. Д.С. Проценко. Проектирование онтологий предметных областей программно-информационных продуктов // Научно-техническая конференция студентов, аспирантов и молодых специалистов МИЭМ. Тезисы докладов. – М.: МИЭМ, 2007. – 469с. – С. 142–143.

12. С.В. Мальцева, Д.С. Проценко. Серверы отношений сетевых сообществ // Труды XIV Всероссийской научно-методической конференции "Телематика'2007". Том 2. Секции D, E. – Санкт-Петербург, 2007. – С. 340–341. (вклад автора 0,06 п.л.).

13. С.В. Мальцева, Д.С. Проценко, В.А. Фомичев, и др. Отчет о научно-исследовательской работе по теме: «Разработка вариантов и правил обеспечения логической систематизации и семантической согласованности данных по проектам, реализуемым в предметной сфере». – Регистрационный номер научного отчета в ВНИИЦ – 01.2.00951569. – 2009. (вклад автора 0,5 п.л.).

Свидетельства об официальной регистрации программ, разработанных на основе результатов диссертации:

14. Программа «Сервер отношений сетевых сообществ практики» (Relationship Server of Network Communities of Practice) / Д.С. Проценко // Свидетельство о регистрации в Роспатенте программы для ЭВМ № 2009610664 от 28 января 2009 г.

Лицензия ЛР № 020832 от 15 октября 1993 г.
Подписано в печать 18 ноября 2009 г. Формат 60x84/16
Бумага офсетная. Печать офсетная.
Усл. печ. л. 1,0.
Тираж 100 экз. Заказ №__
Типография издательства ГУ-ВШЭ
125319, г. Москва, Кочновский пр-д, д. 3