

18. Сергеев А. С. Параллельное программирование. Ростов-н/Д: Издательский центр ДГТУ, 2002. 77 с.

19. Ахо Альфред В., Джон Ульман Джеффри Д. Структуры данных и алгоритмы. М.: Издательский дом «Вильямс», 2003. 384 с.

20. Сергеев А. С. Разработка параллельного комбинированного биоинспирированного метода (гене-

тический алгоритм и алгоритм муравьиных колоний) для решения задач криптоанализа // Системный анализ в проектировании и управлении: сб. науч. тр. XXII Междунар. науч.-практ. конф. Ч. 1. СПб.: Изд-во Политехн. ун-та, 2018. С. 359–370.

Yu. O. Chernyshev, A. S. Sergeev, A. N. Ryasanov
Don State technical university

RESEARCH AND DEVELOPMENT OF THE PARALLEL COMBINED BIOINSPIRED ALGORITHM FOR THE SOLUTION OF CRYPTANALYSIS TASKS

The cryptanalysis task with the use of a new model of optimizing strategy – the combined bioinspired algorithm is considered. Application of the combined bioinspired algorithm based on hybridization by an investment (a genetic algorithm and an algorithm of ant colonies) for implementing the cryptanalysis of shift codes is described. The description of the combined algorithm is provided, and it is shown that the probability of obtaining an optimal variant of the decision while applying the hybrid algorithms of cryptanalysis can't be less than the probability of obtaining an optimal solution when using the classical bioinspired algorithms. The description of the main operations enabling parallel performance at the global level is provided, as well as the block diagram of a parallel algorithm and the data-logical graph-scheme, and the description of the following matrix is given. On the basis of definition of the sets of mutually independent operators and a critical way in the graph the problem of determining the minimum number of processors for the realization of a parallel combined algorithm is solved.

Cryptanalysis, the combined bioinspired algorithms, hybridization by an investment, a genetic algorithm, an algorithm of ant colonies, the data-logical graph-scheme, a following matrix, an independence matrix

УДК 004.9

И. А. Щербаков

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

А. В. Пономарев

Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН)

SemanTags: семантическая разметка объектов с использованием технологии краудсорсинга

Дополнение объектов (например, ресурсов сети Интернет) семантическими метками, характеризующими их содержание, позволяет существенно повысить качество выдачи поисковых машин и в целом способствует более эффективной работе с информацией. Однако полностью автоматическая семантическая разметка не всегда дает адекватный результат из-за существующих ограничений при алгоритмической обработке естественного языка, а ручная оказывается достаточно трудоемкой, особенно, если количество объектов велико. Частично снять проблему трудоемкости ручной разметки в ряде случаев позволяет технология краудсорсинга. В статье предлагается система (веб-сервис) SemanTags, позволяющая использовать технологию краудсорсинга для того, чтобы снабдить произвольные объекты (научные статьи, страницы в Интернете и пр.) метками (тегами), являющимися классами некоторой проблемно-ориентированной онтологии, записанной на языке OWL 2. Описана архитектура и особенности реализации системы, а также предложен новый механизм обеспечения качества, адаптированный для совместной разметки объектов классами онтологии.

Краудсорсинг, крауд-вычисления, онтологии, OWL 2, таксономия, семантическая разметка, семантический поиск

Семантический поиск информации, т. е. поиск, при котором происходит анализ не только

синтаксических конструкций, описывающих ресурсы (непосредственно термов языка), но и се-

мантики ресурсов (как минимум, разрешение неоднозначностей, вызванных языковой омонимией), обладает огромным потенциалом и способен давать значительно более точные результаты, чем простой поиск на основе анализа термов. Однако для обеспечения возможности такого поиска ресурсы необходимо снабдить семантической разметкой. В последние годы применение такой разметки (в том числе микроразметки) становится все более и более популярным. Основным ее достоинством является то, что она позволяет снабжать объекты (например, страницы Интернета) тегами и атрибутами, обладающими определенной семантикой, одинаково трактуемой различными участниками информационной среды. В частности, такая разметка активно используется поисковыми системами для более «осмысленного» поиска объектов.

Существуют два основных подхода в реализации семантической разметки: автоматическая и ручная. Автоматическая разметка предполагает понимание и анализ компьютером информации ресурса. Частично это выполнимо (в том числе и неявно производится современными поисковыми машинами при анализе текста и разрешении неоднозначностей), однако не всегда это оказывается возможным. Ручная разметка отличается тем, что человек выполняет анализ и структуризацию информационного содержания ресурса самостоятельно. Это позволяет полностью контролировать набор тегов и атрибутов разметки, но задача становится очень сложной при наличии большого количества объектов.

Частично снять проблему трудоемкости ручной разметки в ряде случаев позволяет технология краудсорсинга. Она представляет собой привлечение широкого круга лиц к решению проблем производственной деятельности на добровольных началах с применением инфокоммуникационных технологий. Краудсорсинг в настоящее время широко используется в самых разных областях для решения задач, которые оказываются относительно сложны для автоматизации, но просты для человека [1]–[3], а задачи семантического анализа и аннотирования являются именно такими.

Однако при всех достоинствах этой технологии ее применение связано с рядом трудностей. Прежде всего встает вопрос о компетентности участников и качестве получаемых с их помощью результатов [4], [5]. Под качеством принято понимать соответствие результата выполнения некоторой задачи (сбора или обработки информации)

с помощью краудсорсинга и «истинного» результата этой задачи, который на практике часто является результатом, полученным от компетентного «эксперта» [5]. Данная проблема вытекает из общей склонности человека к ошибкам и неправильных алгоритмов оценки результатов, и для ее решения системы краудсорсинга применяют различные механизмы обеспечения качества (общий обзор подобных механизмов приведен, например, в [5]).

В существующих механизмах обеспечения качества важную роль играет сопоставление результатов, полученных от разных участников, которое, как правило, производится посредством их строгого бинарного сравнения. Это универсальный подход, полностью оправданный в большинстве случаев, однако при семантической разметке между концептами, указанными различными участниками, может присутствовать семантическая связь, которую возможно учесть механизмом обеспечения качества. Учет этой информации позволит более точно сопоставлять результаты и получать более адекватные оценки индивидуальных характеристик участников. В частности, факторы, которые необходимо учесть при согласовании семантических меток, это соответствие и уточнение. Например, метка «Лошадь» является уточнением по отношению к метке «Млекопитающее», однако не противоречит ей.

Одним из наиболее распространенных способов представления семантики в информационных системах являются онтологии, а наиболее широко применяемым языком для описания онтологий является язык OWL 2 (спецификация которого стандартизована консорциумом W3C) [6]. Вследствие этого именно язык OWL 2 был выбран для представления семантических отношений в разработанной системе SemanTags.

Таким образом, в данной статье предлагается система (веб-сервис) SemanTags, позволяющая использовать технологию краудсорсинга для того, чтобы снабдить произвольные объекты (научные статьи, страницы в Интернете и пр.) метками (тегами), являющимися классами некоторой проблемно-ориентированной онтологии, записанной на языке OWL 2. Описана архитектура и особенности реализации системы, а также предложен новый механизм обеспечения качества, адаптированный для совместной разметки объектов классами онтологии.

Краткий обзор современных исследований. Управление качеством результатов, получаемых с помощью краудсорсинга, является актуальной

проблемой, привлекающей пристальное внимание исследователей в этой области. Общая перспектива существующих методов обеспечения качества очерчена, например, в обзорных статьях [4] и [5]. Так, в статье [4] приведена следующая классификация методов обеспечения качества:

1) методы согласования (consensus), основанные на использовании избыточности собираемых результатов для разрешения противоречий в них;

2) методы проектирования потока работ – своего рода развитие методов согласования, позволяющее, в частности, рассматривать системы с разнородными заданиями и динамически формируемым потоком работ;

3) методы централизованного назначения работ, предполагающие, что для каждого участника известна ожидаемая эффективность (качество результатов, время выполнения), и распределяющие задания таким образом, чтобы максимизировать общую эффективность;

4) теоретико-игровые методы, позволяющие оценить ту или иную схему функционирования системы крауд-вычислений, исходя из представления о поведении участника как рационального агента, стремящегося к максимизации собственной пользы;

5) методы, основанные на учете свойств заданий, позволяющие использовать свойства заданий или их компьютерной обработки для повышения итогового качества результатов крауд-вычислений;

6) методы, основанные на анализе действий пользователя и воздействии на него.

Предлагаемый в данной статье механизм относится к методам согласования в том смысле, что он не использует дополнительной информации, кроме меток, получаемых от различных участников, что обеспечивает универсальность данного метода. Вместе с тем большинство существующих методов согласования основаны на строгом сравнении результатов, получаемых от отдельных участников, игнорируя взаимосвязи между метками. Лишь в небольшом количестве работ делается попытка адаптации методов согласования к ситуации, когда между метками могут присутствовать отношения. Разработке таких методов, в частности, посвящена [7], авторы которой предлагают расширение алгоритма DS и исследуют разные модели представления отношений между метками (включая байесовскую сеть как компактное представление совместного распределения). В [8] предлагается вероятностная модель разметки для иерархической классифика-

ции, т. е., для ситуации, когда метками, которые участники присваивают объектам, являются классы, организованные в иерархию (классификация книг, товаров).

Однако результаты, полученные в данных работах, оказываются не вполне применимы к рассматриваемой задаче: набор методов согласования, предложенных в [7], основывается на алгоритме DS, обладающем высокой вычислительной сложностью, что ограничивает его применимость при разметке с использованием больших онтологий. Метод же, предлагаемый в [8], предназначен для сценария последовательного уточнения метки участником, в то время как в SemanTags участник может сразу указать множество меток и последовательного уточнения не производится.

Существует также ряд работ (например, LexiTags [9]), где онтология используется для помощи участнику при выборе меток. И хотя этот подход к обеспечению качества может быть эффективным, он относится к группе б (методы, основанные на анализе действий пользователя и воздействии на него) в приведенной ранее классификации и может быть использован только как вспомогательный, не отменяя необходимости согласования меток.

Таким образом, в настоящий момент не существует метода согласования множеств меток, который мог бы быть использован в создаваемой системе с учетом подхода к осуществлению разметки и типового размера онтологий, применяемых при разметке.

Модель и метод обеспечения качества. В качестве основного механизма обеспечения качества для системы SemanTags предлагается механизм согласования, т. е. механизм, позволяющий из набора меток, полученных от участников разметки, получить согласованный набор меток (в идеале – близкий к истинному). Использование в качестве допустимого набора меток онтологии, записанной на языке OWL 2, означает, что механизм обеспечения качества при определении согласованного множества меток объекта должен учитывать отношения между классами (метками), представимые с помощью языка OWL 2.

В результате анализа различных видов отношений, описанных в OWL 2, были выделены два отношения, на обработке которых решено было сосредоточиться в первую очередь: SubClassOf и EquivalentClasses. Согласно спецификации языка OWL 2, отношение SubClassOf («является подклассом») между классами C_1 и C_2 отражает тот

факт, что любой экземпляр класса C_1 также является экземпляром класса C_2 (обратное не обязательно верно). Отношение же `EquivalentClasses` («эквивалентные классы») между классами C_1 и C_2 означает, что любой экземпляр класса C_1 является экземпляром класса C_2 и любой экземпляр класса C_2 также является экземпляром класса C_1 . Таким образом, отношения `SubClassOf` и `EquivalentClasses` используются для описания базовой таксономии понятий, являющейся основой (своеобразным «скелетом») любой онтологии.

В ходе разметки объекта i участник u предлагает набор меток $l_{i,u}$ характеризующих, по его мнению, данный объект. Каждый элемент множества $l_{i,u}$ принадлежит C – множеству классов онтологии. Для снижения влияния индивидуальных особенностей участников (знаний, квалификации, мотивации) на результат разметки каждый объект размечается несколькими участниками. Пусть c_i – множество участников, предложивших свои метки для объекта i . Результатом механизма обеспечения качества разметки F должно стать множество l_i – согласованный набор меток. Символически это можно записать как $l_i = F(l_{i,u}, \xi)$, где ξ обозначает всю дополнительную информацию (кроме исходных меток), которая может использоваться механизмом согласования. Типовым подходом к определению согласованного результата в крауд-вычислениях является получение некоторой оценки возможности того, что при имеющихся свидетельствах определенный результат является правильным результатом задачи. Такая оценка возможности может основываться как на теоретико-вероятностных представлениях, так и на других формальных моделях. При следовании этому подходу основная задача механизма обеспечения качества заключается в получении таких оценок возможности, а в итоговый список согласованных меток попадают, например, те, у которых эта оценка возможности превышает определенный порог. Пусть $Q(i, c | l_{i,u}, \xi)$ – оценка возможности того, что метка c соответствует объекту i при имеющихся свидетельствах $l_{i,u}$ и ξ .

Предлагаемые модель и метод являются адаптацией модели Бьюкенена, предназначенной для рассуждений в условиях неопределенности. Так, каждому участнику ставится в соответствие параметр b_u – коэффициент доверия, принимающий значения в диапазоне $[0; 1]$. Каждое свидетельство,

получаемое от участника (элементы множества $l_{i,u}$), снабжается соответствующим коэффициентом уверенности (степенью уверенности), равным коэффициенту доверия участника, от которого это свидетельство получено. Механизм согласования:

а) вычисляет коэффициенты уверенности для меток-классов, не сообщенных пользователем напрямую;

б) уточняет коэффициенты доверия участников на основе сопоставления множеств меток.

Оценка возможности метки для объекта основывается на свойствах отношения `SubClassOf`. Поскольку, как уже указывалось ранее, `SubClassOf` (C_1, C_2) означает, что если объект c является экземпляром класса C_1 , он также является и экземпляром класса C_2 , то уверенность в принадлежности объекта классу C_1 также может быть перенесена и на принадлежность классу C_2 . Таким образом, при наличии свидетельства о принадлежности объекта i классу c со степенью уверенности μ данная степень уверенности переносится также на принадлежность этого объекта всем классам (меткам), находящимся выше c в таксономии, образованной отношением `SubClassOf` (см. рис. 1, а; штрихами выделены две метки, полученные от разных участников).

Особым случаем является определение степени уверенности в принадлежности объекта классу, в пользу которого получены свидетельства от разных участников (в том числе с разными коэффициентами доверия). Пусть μ_1, \dots, μ_n – степени уверенности в принадлежности объекта i классу c , полученные от разных участников (возможно, в ходе описанного ранее «распространения» степени уверенности). Функция, используемая при сочетании результатов f_m , должна удовлетворять следующим ограничениям:

1) $f_m(\mu_1, \dots, \mu_n) \geq \max(\mu_1, \dots, \mu_n)$. Действительно, при наличии нескольких свидетельств в пользу факта итоговая степень уверенности не должна быть ниже, чем наименьшая из степеней уверенности свидетельств;

2) $f_m(\mu_1, \dots, \mu_n) = f_m(\mu_{n1}, \dots, \mu_{nn})$, где $n1, \dots, nn$ задают произвольную перестановку индексов $1, \dots, n$, т. е. значение функции не должно зависеть от перестановки аргументов.

В частности, в предлагаемой системе в качестве f_m была выбрана функция $1 - \prod (1 - \mu_i)$ (легко убедиться в том, что она удовлетворяет сформулированным ограничениям).

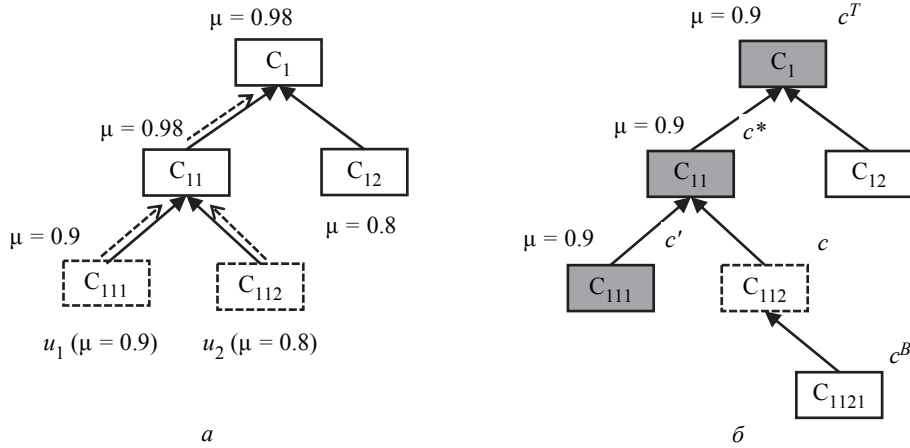


Рис. 1

Степени уверенности меток для объекта, в свою очередь, являются основой для пересчета коэффициентов доверия участников. В частности, для этого используются различные меры близости классов онтологии, определенные на порождаемом онтологией ориентированном графе. Эвристический алгоритм пересчета коэффициента уверенности участника u задается следующим образом:

1. Описанным ранее способом вычисляются согласованные степени уверенности для всех возможных меток, используя данные, полученные от всех участников, кроме u . В частности, это означает, что, если объект был размечен только одним пользователем, то это никак не влияет на его коэффициент доверия.

2. Для каждой метки $c \in l_{i,u}$ находятся метки (классы онтологии) c^* , c^T , c^B и c' (рис. 1, б – штрихами выделена анализируемая метка, заливкой – метки с ненулевой степенью уверенности). Для нахождения этих меток рассматривается ациклический ориентированный граф онтологии OG , в котором вершинами являются классы, а дугами – отношения SubClassOf, c^* – ближайшая достижимая из c вершина OG , у которой определена степень уверенности; c^T – наиболее удаленная вершина OG , достижимая из c ; c^B – наиболее удаленная от c вершина OG , из которой c достижима, а c' – наиболее удаленная от c^* вершина с определенной степенью уверенности, из которой достижима c^* .

3. Коэффициент доверия пользователя определяется как среднее арифметическое соответствия его меток меткам других участников на одних и тех же объектах:

$$\mu_u = \frac{\sum_{i \in n_u} \sum_{i \in l_{i,u}} (\alpha_1 f_1(c, l_{-c}) + \alpha_2 f_2(c, l_{-c}) + \alpha_3 f_3(c, l_{-c}))}{\sum_{i \in n_u} |l_{i,u}|},$$

где n_u – множество объектов, размеченных участником u ; l_{-c} – согласованная разметка без учета участника u , а функции f_i выражают три эмпирических фактора, влияющих на оценку разметки, выполненную пользователем u . Коэффициенты α_i задают относительную важность этих факторов (их сумма должна быть равна 1). В реализованной системе $\alpha_1 = \alpha_2 = 0.25$, $\alpha_3 = 0.5$.

Функции для вычисления факторов используют найденные на шаге 3 метки и задаются следующим образом:

$$f_1(c, l_{-c}) = \frac{d(c^*, c^T)}{d(c^B, c^T)} \mu(c^*),$$

$$f_2(c, l_{-c}) = \frac{\max[d(c', c^*), d(c, c^*)]}{d(c^B, c^T)} \mu(c'),$$

$$f_3(c, l_{-c}) = \frac{d(c, c^T)}{d(c^B, c^T)},$$

где $d(c_1, c_2)$ означает расстояние между вершинами в OG . Сами факторы обеспечивают своеобразный перенос степени уверенности согласованных меток на коэффициент доверия данного участника в зависимости от близости предложенной им метки к согласованным. Так, первый фактор характеризует то, насколько специфичным является класс c^* (относительно которого мнение данного участника и согласованное мнение остальных участников совпадают) в общей иерархии, второй фактор обеспечивает перенос уверенности от самой специфичной согласованной метки (с учетом расстояния до нее), наконец, третий фактор просто показывает, насколько общей является метка (более конкретные метки являются предпочтительными).

Пересчет степеней уверенности и коэффициентов доверия происходит итерациями. На основе начальных коэффициентов доверия рассчитываются степени уверенности, затем, используя новые значения степеней уверенности, оцениваются коэффициенты доверия, и так до момента, когда при очередном пересчете максимальное изменение значения коэффициентов доверия не будет по абсолютному значению меньше заданного порога.

Архитектура системы. Система SemanTags разрабатывается как веб-приложение. Это связано с тем, что она должна быть доступна для очень широкого круга лиц, потенциально заинтересованных в разметке объектов. В целом реализация системы на базе веб-технологий обеспечивает следующие достоинства:

- кроссплатформенность;
- общедоступность;
- качество восприятия информации зависит только от выбранного пользователем устройства;
- отсутствие локальной установки на пользовательское устройство;
- мобильность и простота обновления системы;
- безопасность.

Представление вариантов использования. Основными ролями (актерами) в системе SemanTags являются: Гость, Участник, Администратор (эксперт) и Модератор. Гость – это основной потребитель главной функции системы, обеспечиваемой с помощью семантической разметки, – «интеллектуального» поиска объектов. А именно,

Гость имеет возможность просмотреть информацию об объектах, выполнить поиск и войти в систему (как Участник, Администратор или Модератор). Участник – провайдер описаний объектов. Помимо просмотра или поиска объектов, он может добавлять набор тегов к выбранному объекту. Администратор, в свою очередь, обладает всеми возможностями Участника, а также имеет возможность заблокировать Участника, чью работу он считает низкокачественной, просмотреть все наборы тегов, добавленные Участниками к определенному объекту, добавить наборы тегов от лица эксперта, а также задавать или изменять онтологию семантических отношений. Важную роль играют и Модераторы, которые обладают возможностью добавлять объекты в систему. Графически функциональность системы SemanTags отображена на схеме вариантов использования, приведенной на рис. 2.

Модель хранения. Схема базы данных системы приведена на рис. 3. Таблица Тег содержит классы онтологии, которые могут быть использованы для разметки. Таблица заполняется при конфигурировании системы онтологией. Таблица Объект содержит общую информацию о размечаемых объектах. Следует отметить, что сами размечаемые объекты не хранятся в базе данных. Атрибут «Параметры объекта» таблицы Объект хранит Json-строку со всеми свойствами размечаемого объекта (например, гиперссылку по кото-

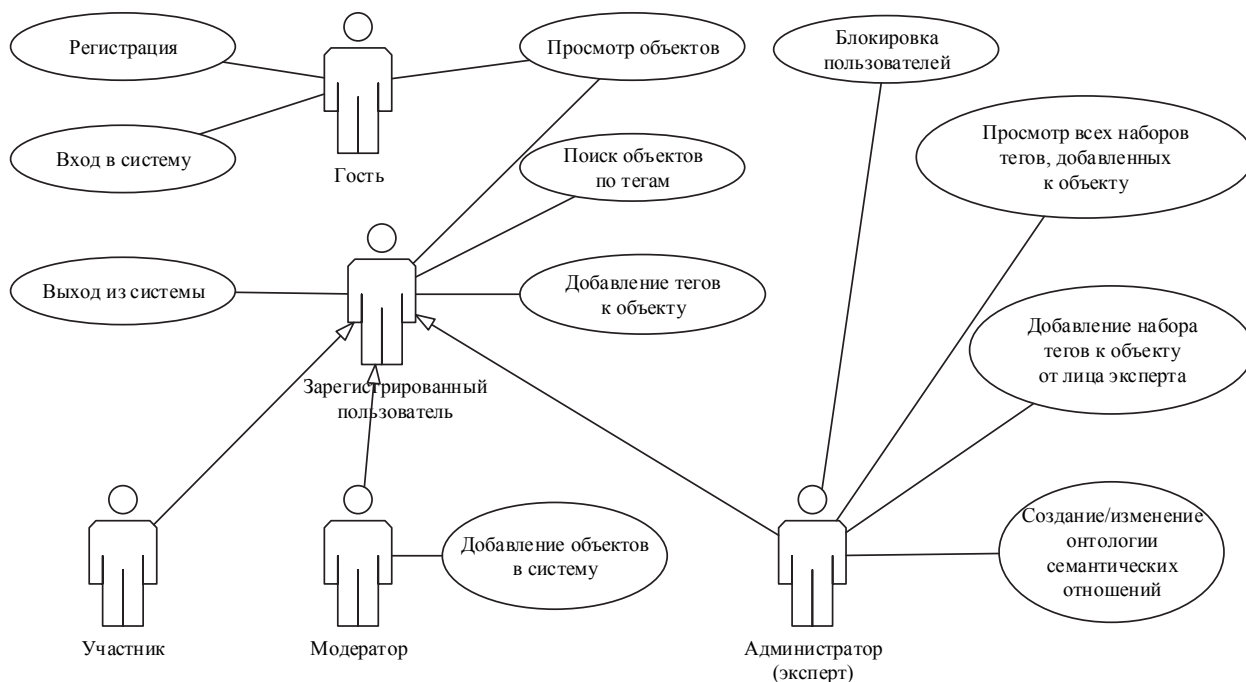


Рис. 2



Рис. 3

рой объект может быть загружен). Интерпретация Json-строки происходит при формировании страниц разметки объекта и просмотра информации о нем. Это, в частности, позволяет добиться универсальности и возможности настройки практически на любой тип объектов. Таблица Пользователь содержит информацию о пользователе, включая его роль (Участник, Администратор или Модератор) и коэффициент доверия (если данный пользователь может выступать в роли участника разметки). Таблицы «Тег, добавленный к объекту» и «Тег объекта» хранят метки, полученные от определенного пользователя и производные метки соответственно.

Структура приложения. В настоящее время существует несколько типовых архитектур веб-приложений. Одной из самых популярных и надежных считается типовая архитектура MVC (Model-View-Controller). Она обеспечивает функциональное разделение и модульность разрабатываемой системы.

В соответствии с этой типовой архитектурой, в разрабатываемой системе выделяется 4 группы объектов: Entity (Сущность), DAO (Доступ к данным), Service (Сервис), Controller (Управление). Общая структура приложения SemanTags представлена на рис. 4.

На уровне Entity задаются классы объектов базы данных с соответствующими процедурами – set(), get(), toString() для работы с данными.

Для работы с базой данных напрямую выделена группа DAO (Data Access Object). В ней определяются классы, осуществляющие отправку запросов к базе данных и получения данных в ответ.

На уровне Service определены классы, которые используют методы DAO для выполнения комбинированных или более сложных действий к базе данных, а также работу с онтологией семантических отношений.

Наконец, на уровне Controller находятся классы, обрабатывающие URL-запросы от интерфейсов пользователя, и иницирующие те или иные действия с базой данных или переходы между страницами.

В статье предложена концепция системы SemanTags, предназначенной для решения задачи семантической разметки больших коллекций объектов классами онтологии, записанной на языке OWL 2. В целях обеспечения максимальной доступности система реализована как веб-приложение. При реализации серверной части использован язык Java с фреймворком Spring, а при реализации пользовательского интерфейса использована библиотека Bootstrap, позволяющая комфортно работать с веб-приложением на широком спектре устройств, обеспечивающих доступ в Интернет.

Предложен также оригинальный метод согласования результатов семантической разметки, основанный на анализе семантических отношений между классами онтологии, учитывающий неточность и неполноту результатов, получаемых

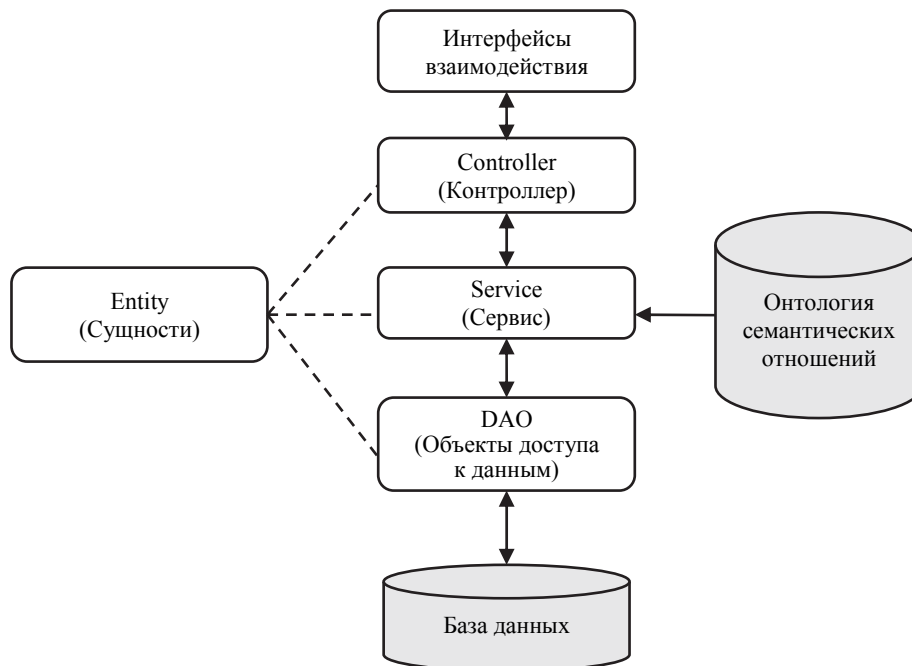


Рис. 4

от отдельных участников. Разработанный метод позволяет оценить как ожидаемое качество разметки, получаемой от отдельных участников, так и степень достоверности собранных меток.

Разработанная система может применяться для структуризации объектов определенной области в целях обеспечения семантического поиска. Это может быть библиотека научных статей, сборник литературы, разделенный по жанрам, или же база данных для обучения нейросетей. Область применения ограничена лишь наличием

заранее спроектированной онтологии семантических отношений, записанной на языке OWL 2. В частности, данная система будет использована для организации библиотеки исследований в области крауд-вычислений, в которой научные публикации, содержащие описания систем, использующих крауд-вычисления, будут снабжены метками из онтологии CROSS-ODF [10].

Работа выполнена при финансовой поддержке РФФИ (проекты 16-37-60107, 19-07-01120).

СПИСОК ЛИТЕРАТУРЫ

1. Harvey D. Observing Dark Worlds: A crowdsourcing experiment for dark matter mapping // *Astronomy and Computing*. 2014. Vol. 5. P. 35–44.
2. Urban sensing: Using smartphones for transportation mode classification / Shin D., D. Aliaga, B. Tunçer, S. M. Arisona, S. Kim, D. Zünd, G. Schmitt // *Computers, Environment and Urban Systems*. 2015. Vol. 53. P. 76–86.
3. Meisel Z. F., Von Holtz L. A. H., Merchant R. M. Crowdsourcing healthcare costs: Opportunities and challenges for patient centered price transparency // *Healthcare*. 2015. Vol. 4, № 1. P. 3–5.
4. Daniel F. Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques and Assurance Actions // *ACM Computing Surveys*. 2018. Vol. 51, № 1. P. 1–40.
5. Пономарев А. В. Методы обеспечения качества в системах крауд-вычислений: аналитический обзор // *Тр. СПИИРАН*. СПб: СПИИРАН, 2017. Т. 54, № 5. С. 152–184.
6. OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax. URL: <https://www.w3.org/TR/owl2-syntax/> (дата обращения 14.03.2019).
7. Separate or joint? Estimation of multiple labels from crowdsourced annotations / L. Duan, S. Oyama, H. Sato, M. Kurihara // *Expert Systems with Applications*. 2014. Vol. 41, № 13. P. 5723–5732.
8. Otani N., Baba Y., Kashima H. Quality control of crowdsourced classification using hierarchical class structures // *Expert Systems with Applications*. 2016. Vol. 58. P. 155–163.
9. Veres C. Crowdsourced semantics with semantic tagging: «Don't just tag it, LexiTag it!» // *Proc. of the 1st Intern. Conf. on Crowdsourcing the Semantic Web (CrowdSem'13)*. CEUR-WS, 2013. Vol. 1030. P. 1–15.
10. Пономарев А. В. Онтология для описания приложений, использующих элементы крауд-вычислений // *Кибернетика и программирование*. 2018. № 3. С. 25–37.

I. A. Shcherbakov
Saint Petersburg Electrotechnical University «LETI»

A. V. Ponomarev
Saint Petersburg Institute of Informatics and automation
of the Russian Academy of Sciences (SPIIRAS)

SEMANTAGS: SEMANTIC TAGGING OF DOCUMENTS BASED ON CROWDSOURCING

Enriching objects (e.g., Internet resources) with semantic labels (or, tags) characterizing the content of these objects allows to significantly improve the search results quality and, in general, contributes to more efficient processing of information. However, fully automatic semantic tagging does not always give an adequate result due to existing constraints in the algorithmic processing of natural language, while manual tagging is rather laborious, especially if the number of objects is large. The latter can be partly alleviated in some cases by using the crowdsourcing technology. The article proposes Seman-Tags system (web-service), which allows to leverage crowdsourcing to tag arbitrary objects (scientific articles, Internet pages, etc.) with classes of some problem-oriented ontology written in OWL 2. The paper describes the architecture of the system, as well as the new quality assurance mechanism, adapted for social tagging of objects by classes of an ontology.

Crowdsourcing, crowd computing, ontologies, OWL 2, taxonomy, semantic tagging, semantic search

УДК 681.518.5

А. А. Алексеев, А. М. Спиваковский
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Применение гибридного вейвлет-спектрального метода для выявления диагностических признаков

Рассматриваются актуальные информационно-вычислительные аспекты в решении задач диагностики сложных объектов и процессов управления. Для оперативного формирования пространства диагностических признаков минимальной размерности предлагается гибридный подход на основе быстрых ортогональных вейвлет-спектральных преобразований, адаптируемых к виду характеристик, описывающих состояния диагностируемого объекта. Данный подход обеспечивает высокую скорость сходимости ряда для анализируемого класса сигналов и характеризуется высокой степенью избирательности преобразования в приспособленном базисе. Достоинством предлагаемого гибридного подхода является однозначность вычисления коэффициентов разложения при любой размерности исходных данных и любом усечении ряда разложения, что позволяет увеличивать точность восстановления аппроксимируемой функции путем добавления компонентов из отсеченной части ряда.

Вейвлет-спектральные преобразования, адаптивные базисы, диагностика объектов управления, диагностические признаки

Нормальное функционирование технической системы управления обычно сопровождается периодическим, экспоненциальным, пилообразным и т. д. характером наблюдаемых колебаний и устойчивым во времени поведением их характеристик. Напротив, аномальное функционирование, обусловленное проявлением каких-либо дефектов объекта, сопровождается возникновением новых составляющих в наблюдаемых сигналах,

таких как всплески, линейные тренды, прямоугольные или треугольные импульсы, а также шумовые реализации аддитивного или мультипликативного вида [1]. Поэтому для выявления информации, позволяющей оценить текущее состояние диагностируемого объекта, необходимо наличие диагностических признаков как для нормального течения процесса, так и для аномального его поведения [2].