

# Системы и интеллектуальное управление

УДК 004.822

## СЕМАНТИЧЕСКОЕ РАСПОЗНАВАНИЕ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ НА ОСНОВЕ ОНТОЛОГИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ЗНАНИЙ О ПРЕДМЕТНОЙ ОБЛАСТИ В ЗАДАЧАХ ИНТЕЛЛЕКТУАЛЬНОГО УПРАВЛЕНИЯ

В.И.Гриценко<sup>1</sup>, А.Я. Гладун<sup>1</sup>, Ю.В.Рогущина<sup>2</sup>

<sup>1</sup>*Международный научно-учебный центр информационных технологий и систем НАН Украины и МОН Украины*

<sup>2</sup>*Институт программных систем НАН Украины*

В работе проанализированы особенности семантического распознавания информационных объектов в сведениях, доступных через Web. В качестве примеров рассматривается обнаружение устройств в Internet of Things, обнаружение Web-сервисов и поддержка информационной службы экстренного вызова. Для решения проблемы семантического распознавания в данной работе предложен переход на новый качественный уровень при обработке информации — использование обработки на семантическом уровне.

**Ключевые слова:** онтология, информационный объект, Semantic Web, OWL, распознавание объектов.

У роботі проаналізовано особливості семантичного розпізнавання інформаційних об'єктів у відомостях, доступних через Web. У якості прикладів розглядається виявлення пристроїв в Internet of Things, виявлення Web-сервісів та підтримка інформаційної служби екстреного виклику. Для вирішення проблеми семантичного розпізнавання в цій роботі запропоновано перехід на новий якісний рівень при обробці інформації — використання обробки на семантичному рівні.

**Ключові слова:** онтологія, інформаційний об'єкт, Semantic Web, OWL, розпізнавання об'єктів.

### ВВЕДЕНИЕ

Сегодня в области информационных технологий (ИТ) происходят глобальные изменения в организации и использовании информационных ресурсов (ИР), непосредственно связанных с интеллектуализацией ИТ, и созданных на их основе прикладных информационных систем, которые функционируют в открытой и динамической информационной Web-среде.

В реальном мире нас окружают разнообразные материальные объекты, наличие информации о которых обеспечивает возможность взаимодействия с ними и использования для своих нужд. Однако совокупность сведений об объектах реального мира порождает другой мир — виртуальный, где существует только информация. Этому виртуальному миру мы даем разнообразные названия, одно из которых — информационное пространство.

© В.И.Гриценко, А.Я. Гладун, Ю.В.Рогущина, 2014

ISSN 0452-9910. Кибернетика и вычисл. техника. 2014. Вып. 178

Познание реального мира происходит через это информационное пространство, которое представляет собой некоторую информационную картину мира [1].

Человек формирует собственное представление о реальном мире, получая и осмысливая информацию о каждом реальном объекте, процессе или явлении. При этом у каждого существует своя информационная картина мира, которая зависит от множества факторов как субъективного, так и объективного порядка, на что влияет уровень образованности человека.

Моделирование является одним из способов познания реального мира и, прежде всего, оно связано с отбором необходимой информации и построением модели. Информационное пространство представляет собой модель реального мира, доступную сознанию человека. Однако любая информационная модель отражает реальный объект только в ограниченном аспекте — в соответствии с поставленной человеком целью.

Существует непрерывная связь между моделями и объектами реального мира. Если иметь дело только с моделями, то исчезнет такое понятие, как адекватность. Устранив объект, мы тем самым разорвем виртуальную связь, устанавливающую объектно-модельное отношение. Поэтому модели находятся в динамическом развитии, совершенствовании на основе реальных объектов и превращаются в некоторый самостоятельный объект, который представляет собой совокупность информации об этом объекте.

Для того чтобы эффективно обнаруживать в Web-ресурсах сведения, необходимые пользователю для решения стоящих перед ним проблем, представляется целесообразным вначале определить, какие именно информационные объекты интересуют пользователя, т.е. построить формальную модель, описывающую свойства множества соответствующих информационных объектов и позволяющую отличать их от других, а затем использовать знания об их структуре, представленные в этой модели, для того, чтобы распознавать конкретные значения их свойств.

**Цель данной работы** заключается в разработке концептуального подхода к построению формальной онтологической модели информационных объектов в виртуальном информационном Web-пространстве и создания методики использования этой модели для восприятия, распознавания, интерпретации и обработки этих объектов в задачах интеллектуального управления.

*Информационный объект* (ИО) — представление (модель) объекта предметной области (ПрО) в информационном пространстве, которое определяет структуру, атрибуты, ограничения целостности и, возможно, поведение этого объекта [2].

## **ТАКСОНОМИЯ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ**

Материальный объект — элемент действительности, взятый без связи со временем, который имеет характеристики, может вступать в отношения и является результатом выделения из некоторой совокупности.

Объект (с точки зрения программирования и, в частности, объектно-ориентированного подхода) представляет собой некоторую сущность в

виртуальном пространстве, которая имеет определенное состояние и поведение, т.е. заданные значения свойств (атрибутов), и для которой определены операции над ними (методы). Как правило, при исследовании объектов выделяется то, что каждый объект принадлежит одному или нескольким классам, которые в свою очередь обуславливают поведение объекта, т.е. являются его моделью.

Информационный объект является расширением программного объекта и представляет собой некую сущность, содержащую в себе сведения в информационной системе о каком-либо реальном или виртуальном объекте (предмете, существе, событии, процессе и т.д.) — уникально идентифицированной материальной или нематериальной сущности реального мира, которая описывает структуру ИО, атрибуты, ограничение целостности и, возможно, поведение. Например, объектом может быть человек, публикация, организация, город, а информационным объектом — их описание.

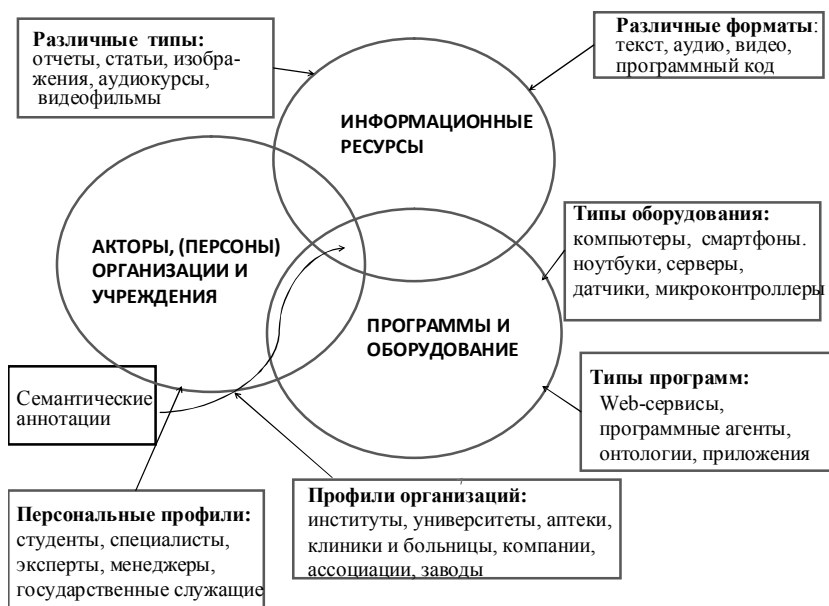


Рис. 1. Структурная схема информационных объектов, представленных в распределенной Web-среде

В концепции Semantic Web и информационном Web-пространстве информационными объектами являются и такие объекты нематериального мира, как онтологии, программные агенты, Web-сервисы, информационные ресурсы, метаданные, базы данных и т.д. В информационном пространстве, также как и в реальном мире, существует множество информационных объектов, что требует в свою очередь разработки моделей, алгоритмов и методов для решения задач поиска (обнаружения), распознавания, восприятия, интерпретации, компоновки и выбора ИО.

На рис. 1 представлена таксономия информационных объектов, где реальные объекты представлены их моделями в информационном пространстве. Одной из насущных проблем информационного Web-пространства является поиск информации и знаний и распределенная

обработка информации.

Научный труд, газетная статья, приказ — примеры ИО в виде текстовых документов. Рисунки, чертежи, схемы — это ИО в виде графических документов. Ведомость начисления заработной платы, таблица стоимости произведенных покупок в оптовом магазине, смета на выполнение работ и прочие виды документов в табличной форме, где производятся автоматические вычисления по формулам, связывающим ячейки таблицы, — это примеры ИО в виде электронных таблиц. Результат выборки из базы данных — это тоже ИО.

Довольно часто мы имеем дело с составными документами, в которых информация представлена в разных формах. Такие документы могут содержать и текст, и рисунки, и таблицы, и формулы, и многое другое. Научные статьи, отчеты, книги — это примеры составных документов, являющихся ИО сложной структуры. Другими примерами сложных ИО могут служить создаваемые на компьютере презентации и гипертекстовые документы.

На основе представленной структурной схемы информационных объектов, и используя соответствующие инструментальные средства, мы разработали TaxPRO — таксономию информационных объектов, существующих в распределенной Web-среде на формализованном онтологическом языке, что позволяет прикладным программам использовать этот объект при выполнении своих задач.

## **ИНФОРМАЦИОННЫЕ ОБЪЕКТЫ В WEB**

Информационную модель, которая не имеет непосредственной связи с объектом из материального мира, тоже можно считать объектом, но не материальным, а информационным [3]. Таким образом, информационный объект получается из информационной модели путем «отчуждения» информации от объекта-оригинала. Как правило, ИО возникают в результате сознательной или несознательной целенаправленной деятельности людей, но могут также порождаться компьютерами (программами), например, нелинейно взаимодействующими информационными потоками в компьютерных сетях и пр.

Сам объект предметной области может быть как материальным (например, человек, организация, место), так и виртуальным, не имеющим конкретного материального эквивалента (например, книга, программа), хотя и связанного с тем или иным материальным носителем информации. Кроме того, объект ПрО может представлять собой чисто абстрактное понятие, в принципе не связанное с конкретным носителем информации и отображаемое лишь в сознании человека (например, «время», «мораль»).

Объект ПрО является синонимом экземпляра объекта – это уникально идентифицированная материальная или нематериальная сущность реального мира. Объекты имеют состояние и, возможно, явным образом специфицированное поведение. Примерами объектов могут служить личности, определенные предметы или их части, явления, события или процессы, которые с ними связаны. В распределенном информационном

Web-пространстве — это объекты нематериального мира: онтологии, программные агенты, Web-сервисы, метаданные, информационные ресурсы (ИР), которые могут быть как активными, так и пассивными.

## ПРОБЛЕМА РАСПОЗНАВАНИЯ В СЕМАНТИЧЕСКОЙ СРЕДЕ

Как известно, проблема распознавания образов (изображений) является одной из наиболее фундаментальных проблем информатики и теории интеллектуальных систем, которая имеет огромное практическое значение. Распознавание образов — это научная проблема, целью которой является классификация объектов по нескольким категориям или классам с помощью выделения существенных признаков, характеризующих эти данные, из общей массы несущественных данных [4]. В традиционном распознавании объекты называются образами. Связано это с тем, что информация о многих объектах и явлениях в настоящее время регистрируется и хранится в виде цифровых изображений [5].

Проблема распознавания, как отдельная научная проблема и как научное направление, возникла и развивается с конца 50-х годов прошлого века и в настоящее время получает все большую значимость и все более широкое распространение. Трудно назвать такую сферу деятельности, где не используется задача распознавания. Проблема распознавания в глобальном масштабе (в материальном и нематериальном мире) определяется как отнесение исходных данных к определенному классу с помощью выделения существенных признаков, характеризующих эти данные, из общей массы несущественных данных.

В информатике распознавание образов — раздел информатики и смежных дисциплин, развивающий основы и методы классификации и идентификации предметов, объектов, явлений, процессов, сигналов, ситуаций и т. п., которые характеризуются конечным набором некоторых свойств и признаков.

Классическая постановка задачи распознавания образов [6] может быть сформулирована следующим образом:

- задано конечное множество классов  $P$ ,  $P = \{p_j\}, j = \overline{1, m}$ ;
- задано конечное множество объектов  $O$ ,  $O = \{o_i\}, i = \overline{1, n}$ , о каждом из которых известно, что он может быть однозначно отнесен к одному из классов из  $P$ :  $\forall o_i \in O \exists p_j \in P : o_i \in p_j$ ;
- каждый объект  $o_i = \langle a_{i_1}, \dots, a_{i_k} \rangle$  из  $O$  обладает набором из  $k$  свойств  $A_q, q = \overline{1, k}$  таких, что  $a_{i_q} \in A_q$ ;
- задано классифицированное подмножество объектов (обучающая выборка)  $O_{Cl} \subseteq O$  такое, что  $\forall o_i \in O_{Cl}$  известно, к какому именно классу  $p_j \in P : o_i \in p_j$  относится этот объект.

Необходимо определить, к каким классам относятся все объекты из  $O$ .

Для объектов реального мира типовая постановка задачи классификации

предполагает выделение определенного конкретного объекта среди подобных ему (например, узнать среди множества школьников своего ребенка).

В наиболее общем случае задача распознавания может быть разделена на две подзадачи, которые выполняются итеративно до получения приемлемого результата:

**Шаг 1.** Построение гипотезы о том, к каким классам могут относиться объекты, и как именно характеризуются объекты внутри класса;

**Шаг 2.** Отнесение конкретного объекта к одному из этих классов.

Если на 2-м этапе распознавание оказывается unsuccessful, то необходимо вернуться к шагу 1 и сформировать другое множество классов (расширить начальное множество или изменить критерии принадлежности к классу). Например, услышав звук шагов в квартире, человек может предположить, что их издает другой человек — так как ему известно о наличии в помещении других людей. Однако если затем он слышит, например, мяуканье, то он может предположить, что источником звуков является либо кошка, либо работающий телевизор.

В *информационном пространстве Web* задача распознавания объектов непосредственно связана с информационным поиском и той целью, для которой пользователь осуществляет этот поиск. В отличие от реального мира инициатором распознавания является сам пользователь, так как он получает сведения из информационного пространства лишь в ответ на какие-то свои действия.

Задачу распознавания в таком случае можно разделить на несколько этапов (рис.2):



Рис. 2. Этапы распознавания ИО в Web

**Этап 1.** Проанализировав возникшую перед пользователем цель, необходимо определить, к какому классу относится необходимый ему информационный объект.

Для этого есть несколько возможных путей:

- выбрать явным образом класс искомого ИО из таксономии ИО (например, «Web-сервис», «сотрудник», «статья», «организация»;
- задать свойства и ограничения для искомого ИО с целью автоматического обнаружения этого класса в таксономии ИО (например, «возраст», «фамилия», «профессия», «URL-адрес», «количество страниц» — обнаружение будет успешным, если заданный набор свойств не противоречив и может принадлежать одному ИО;
- задать примеры ИО, относящихся к этому классу, приведя указания на экземпляры ИО этого класса (например, «Келли Лада де Мандрака» позволит идентифицировать класс «собака», а «ISBN 978-3-659-56520-5» — класс «публикация»).

Кроме описания того класса  $p_{ИО} \in P$ , к которому относится искомый ИО, следует задать связи этого класса с другими семантически связанными классами ИО:

- подклассы  $p_{подкласс}, p_{подкласс} \in P: \forall o \in p_{подкласс} \Rightarrow o \in p_{ИО}$ ;
- надклассы  $p_{надкласс}, p_{надкласс} \in P: \forall o \in p_{ИО} \Rightarrow o \in p_{надкласс}$ ;
- классы, к которым относятся свойства объектов  $o_i = \langle a_{i_1}, \dots, a_{i_k} \rangle$  из  $O$   $A_q, q = \overline{1, k}$ , если эти свойства также являются классами  $A_q \in P$ .

**Этап 2.** Сформировать массив информации, в котором надо найти сведения об ИО.

Для этого можно:

- явно указать адрес хранилища структурированных сведений выбранного класса (например, адрес репозитория онтологий или Web-сервисов);
- указать критерий поиска в Web (поисковый запрос, переадресуемый внешней информационно-поисковой системе);
- задать перечень ИР, релевантных решаемой задаче, в которых неявным образом (неструктурированно) содержатся сведения о распознаваемом ИО.

**Этап 3.** Сформировать ограничения на ИО выбранного класса, удовлетворяющие потребностям пользователя. Для этого надо — явно или неявно — связать свойства ИО с приемлемыми или неприемлемыми их значениями. Чем более точно задан на этапе 1 класс искомого ИО, тем более детализированно могут быть заданы эти условия. Например, если на этапе 1 было определено, что ИО относится к классу «человек», то в условиях можно задавать его фамилию и возраст, но если ИО был отнесен к классу «сотрудник», являющемуся подклассом класса «человек», то для него можно определять значения таких свойств, как «профессия», «место работы» и «стаж».

Например, если пользователю для каких-то целей надо найти научных сотрудников, проживающих на территории Украины и имеющих не менее 100 научных публикаций, то при распознавании нужно будет распознавать их

место работы, место жительства и должность, и не будут учитываться такие свойства людей, как возраст или наличие домашних животных.

**Этап 4.** Осуществить распознавание сведений об искомым ИО, которое сводится к семантической разметке набора ИР, сформированного на этапе 2, понятиями (свойствами), выделенными на этапе 3.

При этом могут применяться методы лингвистического анализа, распознавания речи и изображений (для анализа мультимедийных ИР), а также средства менеджмента знаний, направленные на автоматизированное использование знаний о структуре искомым ИО, представленных в интероперабельной форме, допускающих автоматизированную обработку.

Кроме того, для интероперабельного представления знаний соответствующих ПрО целесообразно использовать онтологии, используемые в качестве их формальных моделей. Понятия онтологий позволяют семантически аннотировать свойства классов и их значения. Например, достаточно часто при обнаружении ИО, относящихся к классу «Web-сервис», именно онтологический подход позволяет семантически аннотировать их входы и выходы.

#### **ОНТОЛОГИИ КАК ИСТОЧНИК СВЕДЕНИЙ О СТРУКТУРЕ ИО**

Достаточно очевидно, что задание иерархии классов той или иной предметной области, а также свойств этих классов, является сложной и трудоемкой задачей, требующей как экспертных знаний об этой ПрО, так и навыков, связанных с менеджментом знаний. Поэтому целесообразно использовать такие формализованные и интероперабельные модели ПрО, как онтологии, в которых эти знания представлены явным образом и пригодны для автоматизированной обработки различными программными средствами. На сегодня онтологии фактически стали стандартом для представления знаний в Web.

При этом следует учитывать, что сегодня Web обеспечивает доступ к большому количеству онтологий как верхнего уровня, так и различных достаточно узких ПрО. В частности, ряд языков, стандартов, методов и программных средств создания и развития онтологий был разработан в рамках проекта Semantic Web, а на их основе уже создано достаточно много прикладных онтологий.

Научить компьютер (программу) распознавать смысл ИО (его семантику), интерпретировать эти объекты и на этой основе создать Web нового поколения, Semantic Web, — это проблема, над которой активно работают многочисленные исследователи в мире, и это направление в ИТ является сегодня актуальным и получает все более широкое развитие.

Проект Semantic Web направлен на то, чтобы дать информации "четкое значение, позволяя компьютерам и людям работать в тесном и более эффективном и плодотворном сотрудничестве" [7]. При этом для решения представленной проблемы предлагается переход на новый качественный уровень при обработке информации — использование обработки на семантическом уровне.

Целью проекта Semantic Web является преобразование всей



совокупности информационных ресурсов Web в единую базу знаний, пользоваться которой могут как люди, так и программы. Для этого необходимо снабдить каждый ресурс описанием его семантики и предоставить средства для автоматизированной обработки этих описаний и представления знаний о них [8].

Автором этой концепции является Т. Бернерс-Ли, который ранее задумал и разработал Web, а теперь возглавляет Консорциум W3. Целью Бернерса-Ли было сформировать информационное пространство, к которому каждый имеет непосредственный и интуитивный доступ, и не только для просмотра, но и для создания информации. Компьютеры становятся способными анализировать все данные в Web — контент, связи и транзакции между людьми и компьютерами. Эта концепция была принята и продвигается Консорциумом W3 — лидером в развитии технологий для Web (многие из основополагающих технологий, таких, как XML и RDF [9], были разработаны именно W3C).

Для её внедрения предполагается создание сети документов, содержащих метаданные о ресурсах Web, тогда как сами ресурсы предназначены для восприятия человеком, метаданные используются машинами (поисковыми роботами и другими интеллектуальными агентами).

## **ПРЕДСТАВЛЕНИЕ ИНФОРМАЦИОННЫХ ОБЪЕКТОВ В SEMANTIC WEB**

В Semantic Web информации придают четкое значение (смысл), что улучшает доступ компьютеров и людей к информации и обеспечивает их совместную работу в тесном сотрудничестве. Для того, чтобы Web достигла своего полного потенциала, она должна трансформироваться в Semantic Web, обеспечивая универсально доступную платформу, позволяющую при этом распределенным данным быть обработанными как программами, так и людьми.

Средства подхода Semantic Web — это набор методов и технологий, предназначенных для представления информации в пригодном для машинной обработки виде. Эти средства позволяют представлять информацию в виде семантической сети, специфицированной с помощью онтологий. Благодаря стандартизованному представлению программа-клиент может непосредственно извлекать факты и делать из них логические заключения, используя протокол HTTP и идентификаторы ресурсов URI. Использование средств Semantic Web становится в последние годы актуальным для разработки интеллектуальных приложений.

С точки зрения распознавания ИО важно, что Semantic Web является источником семантической метаинформации о распределенных ИО и ИР, аннотируя их контент и обеспечивая возможность автоматизированного извлечения из них сведений об искомым ИО.

При этом задачи традиционного распознавания образов и семантического распознавания становятся, по сути, очень близкими. Можно сказать, что задача распознавания образов является подклассом более общей и глобальной задачи распознавания материальных и нематериальных объектов, с которыми взаимодействует человек в своей деятельности.

В семантическом распознавании проблема восприятия, распознавания и интерпретации объектов — комплексная проблема, которая разбивается на отдельные подзадачи [10]. Например, программные агенты должны воспринимать окружающую среду для успешного выполнения поставленных перед ними задач, агенты могут распознавать других агентов в мультиагентной системе для совместного выполнения задачи. Задача поиска информации и знаний — это поиск релевантных объектов для задач функционирования интеллектуальной информационной системы (ИИС), также как и проблема содержательной интерпретации объектов (онтология дает содержательное описание объекта).

Например, пользователю необходимо найти набор исполнителей для выполнения определенной НИР или проекта. Онтология ПрО позволяет узнать, что искомым объектом является сущность, относящаяся к классу «человек» и обладающая такими параметрами, как образование, стаж, навыки, наличие публикаций, ученой степени, дипломов и опыта работы и т.д. Но только наличие соответствующей онтологии позволяет структурировать сведения, извлекаемые из ИР, с точки зрения конкретной проблемы, так как при решении другой проблемы информация, извлекаемая из тех же ИР о тех же самых людях, может быть совсем иной (например, при обнаружении эффективных способов лечения пациентов с похожими симптомами или при поиске владельцев собак одной породы).

Традиционное распознавание образов, распознавание речи и распознавание текста являются лишь частными случаями значительно более общей проблемы. Распознавание предполагает обнаружение в каком-либо информационном ресурсе (ИР) сведений о том или ином информационном объекте (ИО), интересующем пользователя [11]. Например, распознавание лиц — это обнаружение в графическом ИР элементов, характеризующих внешние данные ИО типа «человек».

Основопологающим понятием Semantic Web является онтология. Одним из определений понятия «онтология» является понятие «концептуализация», под которым понимается строгое описание системы понятий, объектов и других сущностей и отношений между ними. Можно сказать, что концептуализация — это абстрактное, упрощенное видение мира, который мы стремимся представить для каких-то целей. Таким образом, концептуализация расчленяет какую-либо область знаний, которая существует в целостном виде, выделяет из этой области отдельные объекты, а потом формулирует отношения, присущие для данной области. Основная часть формально представленного знания базируется на концептуализации; каждая база знаний или система, которая базируется на знаниях, явным образом или неявно связывается с какой-то концептуализацией.

Созданные на основе семантических технологий ИТ отличаются от традиционных тем, что используют явно выраженные (в виде онтологии) знания о предметной области, где онтология выступает как средство формализации знаний. ИО имеют заложенное в них человеком или программой поведение и взаимодействуют как с информационным, так и с физическим пространством. В последнем случае взаимодействие осуществляется через физических носителей (людей).

Онтологию можно рассматривать как основу для представления структуры ИО, т.е. класс онтологии, а различные ИР — как источник для создания экземпляров этого объекта [13]. Такой подход позволяет интегрировать информацию, поступающую из различных источников, и сформировать необходимые пользователю знания. При этом задача подразделяется на несколько подзадач:

- *формирование (или поиск) онтологии*, отражающей структуру ИО (или множества объектов), знания о которых необходимы пользователю для решения стоящей перед ним проблемы;
- *поиск ИР*, явно или неявно содержащих сведения об этих ИО;
- *извлечение знаний об ИО* из ИР;
- *представление извлеченных знаний в форме, понятной и удобной пользователю.*

Следует отметить, что во многих случаях этот процесс является итеративным и при повторном решении проблемы информацию об ИО необходимо обновлять, извлекая необходимые сведения из тех ИР, которые доступны пользователю (например, через Web или корпоративную сеть).

## РАСПОЗНАВАНИЕ WEB-СЕРВИСОВ

Рассмотрим в качестве примера распознавание сведений о таком ИО, как Web-сервис. Концепция Web-сервисов возникла в конце 90-х гг. XX века, и их архитектура базируется на протоколах SOAP, UDDI и WSDL. Стандарт SOAP (Simple Object Access Protocol) — протокол для посылки сообщений по HTTP; WSDL (Web Services Description Language) — язык для описания программных интерфейсов Web-сервисов; UDDI (Universal Description, Discovery and Integration) — стандарт индексации Web-сервисов.

Используя OWL-S (Web Ontology Language for Services), Web-сервис может сообщать о своих функциональных возможностях потенциальным пользователям. Описание OWL-S для сервиса состоит из профиля, модели и обоснования сервиса (рис. 3) [14].

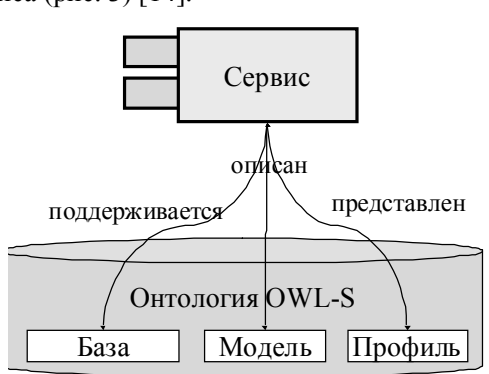


Рис. 3. Структура OWL-S описания Web-сервиса

Чтобы использовать Web-сервис, нужно интерпретировать компьютерам описание сервиса и средств доступа к нему, т.е. решить проблему распознавания нужного пользователю сервиса из множества других сервисов,

содержащихся в репозитории. Информацию о том, какие функции предоставляет данный Web-сервис, дает нам его WSDL-описание, а UDDI-реестры используются для поиска существующих Web-сервисов. Самой большой проблемой при обнаружении сервисов является их распределенный характер.

На основе языка онтологий OWL разработан язык для сервисов OWL-S. Цель разработки OWL-S состоит в том, чтобы сделать возможным использование логического вывода для Web-сервисов, планирование компоновки Web-сервисов, автоматическое применение сервисов программными агентами. OWL-S обеспечивает декларативные описания свойств Web-сервиса и возможности, которые могут использоваться для автоматического обнаружения сервиса.

На рис. 4 показано, как OWL-S используется для распознавания Web-сервисов.

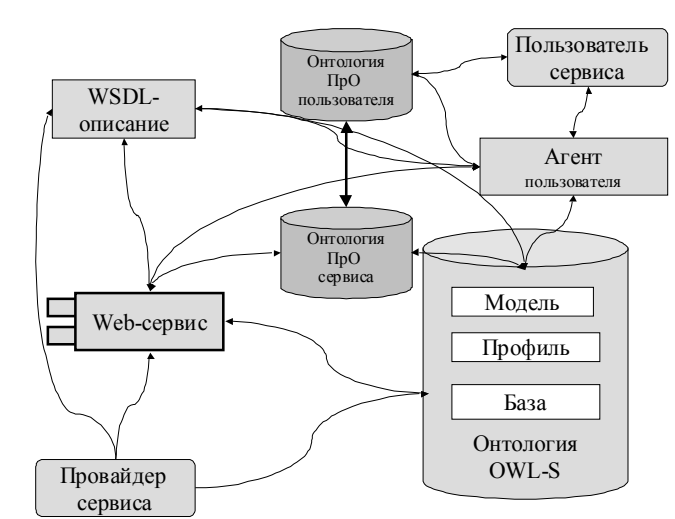


Рис. 4. Процесс распознавания Web-сервиса

Для описания функциональных возможностей Web-сервиса можно использовать семантические аннотации в соответствии со спецификацией OWL-S. Запрос ссылается на онтологию домена на языке OWL, которая может аннотировать запрос на основе общей семантики. Онтологический инжиниринг играет ведущую роль в добавлении семантики к описанию Web-сервисов. Описание OWL-S содержит:

- *ServiceProfile*, включающий в себя общую информацию, такую как имя, описание, входы и выходы;
- *ServiceModel*, описывающий, как действует функция Web-сервиса;
- *ServiceGrounding*, показывающий, как связаться с Web-сервисом.

Среди трех видов информации, предоставляемой в соответствии с описанием OWL-S, для аннотирования Web-сервиса наиболее интересны понятия профиля OWL-S (входы и выходы) и *ServiceGrounding*. Поскольку оба эти понятия являются полезными для поиска и композиции Web-сервисов, то понятие *ServiceGrounding* базируется на следующих принципах соответствий между OWL-S и WSDL:

- атомарный процесс OWL-S соответствует операции в WSDL;
- вход или выход атомарного процесса соответствуют сообщению в WSDL;
- типы сообщений атомарного процесса соответствуют абстрактным типам в WSDL.

Описание запроса выглядит следующим образом. Запрос рассматривается как сервис Semantic Web в OWL-S. Любой запрос, представленный клиентом, анализируют с помощью парсера (синтаксического анализатора), который извлекает только те параметры, которые необходимы для композиции (входы и выходы).

## **ПРАКТИЧЕСКИЕ ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ РАСПОЗНАВАНИЯ ИО**

Реальными примерами использования семантического подхода к распознаванию ИО явился ряд международных проектов, подготовленных нами в рамках европейской программы научных исследований и технологического развития Horizon–2020 (2014–2020). Один из проектов главной целью имел реализацию семантической интероперабельности в гетерогенной среде Интернета вещей [15]. Как известно, концепция Интернета вещей (Internet of Things, IoT) предполагает объединение вещей (физических объектов), используя Интернет-протоколы для решения задач в разных сферах деятельности человека (экономия энергоемких ресурсов, городской и другие виды транспорта, промышленность, безопасность человека и многие другие). Концепция IoT предусматривает технологии взаимодействия вещей друг с другом или с внешней средой, что может перестроить экономические и общественные процессы, при этом исключая из части действий и операций необходимость участия человека. Задача семантической интероперабельности предполагает онтологическое представление знаний об устройстве — семантическим описанием (онтология — модель знаний), а также наделение каждого устройства определенным интеллектом, который должен помочь ему как распознать контекст своего окружения (местонахождение, собственное состояние и положение объекта, который контролирует эту вещь), так и наладить диалог с окружающими вещами и управляющим устройством. Таким образом, имея онтологическое представление каждой вещи в виде виртуального ИО, пользователь или программа смогут найти, распознать и использовать релевантное устройство для решения своей задачи. Кроме того, гетерогенные устройства смогут правильно понимать друг друга при обмене информации на основе онтологического представления о них знаний.

Еще одним проектом, требующим решения семантического распознавания ИО, является проект геоинформационной системы — системы вызова экстренных оперативных служб по номеру «112» для оказания экстренной помощи населению при угрозах для жизни и здоровья, для уменьшения материального ущерба при несчастных случаях, авариях, пожарах, нарушениях общественного порядка и при других происшествиях и чрезвычайных ситуациях.

Одной из главных проблем проекта является оперативное распознавание человека, обратившегося в службу за помощью, в условиях ограниченных

информационных сведений о нем и ресурсов по времени. В данном случае ИО является сам пострадавший человек (т.е. некоторая модель, описывающая характеристики этого человека: адрес, поликлиника, к которой он прикреплен, информация как о пациенте и др.). Стоит задача распознать потребности этого человека и предоставить ему определенные ресурсы для ликвидации последствий. Семантический подход позволяет определить недостающие знания о человеке на основе базы знаний (онтологий), агентно-ориентированных приложений, что обеспечивает быструю реакцию на аварийный вызов. Кроме того, решается та же задача интероперабельности устройств, с помощью которых был осуществлен вызов: смартфон, планшет, ноутбук, телефон и пр.

## Выводы

Для решения проблемы семантического распознавания в данной работе предлагается переход на новый качественный уровень при обработке информации — использование обработки на семантическом уровне. ИТ, созданные на основе семантических технологий, отличаются от традиционных тем, что используют явно выраженные (в виде онтологии) знания о предметной области, где онтология выступает как средство формализации знаний.

В интеллектуальных Web-приложениях есть своя специфика — они функционируют в открытом информационном пространстве, поэтому они нуждаются в постоянном обновлении знаний и их подкачке из внешней среды. В распределенном информационном Web-пространстве ИО, отчужденные от реальных физических объектов, определяют структуру, атрибуты, ограничения целостности и поведение физического объекта. Онтологическое представление знаний о рассматриваемых объектах позволяет улучшить задачу поиска, распознавания и интерпретации ИО (и соответствующих им физических объектов) при создании интеллектуальных информационных систем и приложений в распределенной и гетерогенной Web-среде. В работе рассмотрены конкретные примеры проектов, в которых решающей составляющей является задача распознавания ИО в интеллектуальных информационных системах различного целевого назначения.

1. Гриценко В.И. Введение в архитектуру информационного пространства / Гриценко В.И., Вовк М.И., Котова А.Б. — К.: Наукова думка, 2003. — 167с.
2. Глоссарий по информационному обществу / Под общ. ред. Ю.Е. Хохлова. — М.: Институт развития информационного общества, 2009. — 162 с.
3. Newhagen J. Interactivity, Dynamic Symbol Processing, and the Emergence of Content in Human Communication. *The Information Society* (20) (Taylor and Francis Inc, 2004), 395p.
4. Ту Д. Принципы распознавания образов / Д. Ту, Р. Гонсалес. — М. 1978. — 325с.
5. Лифшиц Ю. *Современные задачи теоретической информатики. Лекции по статистическим методам распознавания образов, распознаванию лиц, классификации текстов*— <http://yury.name/modern/>
6. Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания и классификации // Проблемы кибернетики. — М.: Наука, 1978, вып. 33. — С. 5–68.

7. Berners-Lee T., Hendler J., Lassila O. The Semantic Web. *Scientific American Magazine*. Retrieved March 26, 2008.
8. *W3C Semantic Web Activity*. — <http://www.w3.org/2001/sw/Activity/>.
9. *Resource Description Framework (RDF) Model and Syntax Specification. W3C Proposed Recommendation*. January 1999. <http://www.w3.org/TR/PR-rdf-syntax>.
10. Kirasic D., Basch D. Ontology-Based Design Pattern Recognition Knowledge-Based Intelligent Information and Engineering Systems. *Lecture Notes in Computer Science*, V. 5177, 2008. — P. 384–393.
11. Gladun A., Rogushina J., Andrushevich A. and Kurbatski A. User-oriented recognition of intelligent information objects in distributed dynamic informational Web-space. *Proceedings of the 12-th Inter. conf. on Pattern Recognition and Information Processing (PRIP'14)*, 2014, Minsk, Belarus, 2014, P.70–75.
12. Gladun A., Rogushina J. Use of Semantic Web Technologies and Multilinguistic Thesauri for Knowledge-Based Access to Biomedical Resources. *International Journal of Intelligent Systems and Applications*, 2012, №1, P.11–20.
13. *OWL-S Home Page*. — <http://www.daml.org/services/>
14. Рогушина Ю.В. Онтологический подход к поиску Web-сервисов в распределенной среде Интернет / Рогушина Ю.В., Гладун А.Я. //Информатика, Минск, № 4, 2006. — С.116–127.
15. Gladun A., Rogushina J., Andrushevich A., Kurbatski A. Use of the Ontological Approach to Semantic Search in the Environment of the Internet of Things. *Proceedings of the 4-th Inter. conf. Open Semantic Technologies for Intelligent Systems OSTIS-2014*, Minsk, Belarus, 2014, P.353–356.

UDC: 004.822

## SEMANTIC RECOGNITION OF INFORMATION OBJECT BASED ON ONTOLOGICAL REPRESENTATION OF DOMAIN KNOWLEDGE IN PROBLEMS OF INTELLECTUAL MANAGEMENT

Gritsenko V.I.<sup>1</sup>, Gladun A.Ya.<sup>1</sup>, Rogushina Yu.V.<sup>2</sup>

<sup>1</sup>*International Research and Training Center for Information Technologies and Systems of National Academy of Sciences of Ukraine and Ministry of Education and Science of Ukraine*

<sup>2</sup>*Institute of Software Systems of National Academy of Sciences of Ukraine*

**Introduction:** In this paper analysis of the features of semantic recognition of information objects from the resources available by the Web is made. To solve the problem of semantic recognition we propose a transition to a qualitatively new level of information processing — the use of the processing at the semantic level. As examples, the detection of devices in the Internet of Things, the discovery of Web-services and support of emergency information services are consider.

In order to effectively detect information necessary for the user problem solving from the Webresource it seems appropriate to determine at first which information objects are interested to users, ie, construct a formal model that describes the properties of a set of relevant information objects and to distinguish them from the others, and then use the knowledge about the structure presented in this model for identification of the specific values of their properties.

**The purpose** of this work is to develop a conceptual approach to the construction of formal ontological model of information objects in the virtual information space of the Web and to create a technique of this model using for perception, recognition, interpretation and processing of these objects for the tasks

of intelligent control. Information object is a representation that models an object from subject domain in the information space, which defines the structure, attributes, constraints, and perhaps behavior of the object.

**Results:** To solve the problem of semantic recognition we propose a transition to a qualitatively new level of information processing – at the semantic level. Modern IT that are based on semantic technologies, different from the traditional ones by the explicit (ontology-based) knowledge representation of the subject area, where the ontology serves as a means of knowledge formalizing .

The specifics of the new intelligent Web-based applications are an operatiin in an open information space: they need permanent updating of knowledge and its pumping from the external environment. In the distributed information space of the Web information objects alienated from the real physical objects define the structure, attributes, constraints and behavior of these physical objects.

**Conclusions:** The ontological representation of knowledge about the considered objects can improve the problem of finding, recognizing and interpreting the IOs (and corresponding physical objects) for creating intelligent information systems and applications in a distributed and heterogeneous Web-based environment.

**Keywords:** ontology, an information object, Semantic Web, OWL, object recognition.

1. Gritsenko V.I. Vovk M.I., Kotova A.B. *Introduction to the architecture of the information space*. Kiev, Naukova Dumka, 2003, 167 p. (in Russian).
2. Newhagen J. Interactivity, Dynamic Symbol Processing, and the Emergence of Content in Human Communication. *The Information Society (20)* (Taylor and Francis Inc, 2004), 395 p. (in Russian).
3. *Glossary on Information Society* / under total. Ed. Y.E. Khokhlova. Moscow: Institute of the Information Society, 2009. — 162 p.
4. Tu D., Gonzalez R. *Principles of Pattern Recognition*. Moscow: 1978. 325 p. (in Russian).
5. Lifshitz Y. *Modern problems of theoretical computer science. Lectures on statistical methods for pattern recognition, face recognition, text classification* — <http://yury.name/modern/> (in Russian).
6. Zhuravlev Y.I. On the algebraic approach to the solution of problems of recognition and classification. *Problems of Cybernetics*. Moskow: Nauka, 1978, vol. 33. P. 5–68 (in Russian).
7. Berners-Lee T., Hendler J., Lassila O. The Semantic Web. *Scientific American Magazine*. Retrieved March 26, 2008.
8. *W3C Semantic Web Activity*. — <http://www.w3.org/2001/sw/Activity/>.
9. *Resource Description Framework (RDF) Model and Syntax Specification. W3C Proposed Recommendation*. January 1999. — <http://www.w3.org/TR/PR-rdf-syntax>.
10. Kirasic D., Basch D. Ontology-Based Design Pattern Recognition Knowledge-Based Intelligent Information and Engineering Systems. *Lecture Notes in Computer Science*, vol. 5177, 2008. P. 384–393.
11. Gladun A., Rogushina J., Andrushevich A. and Kurbatski A. User-oriented recognition of intelligent information objects in distributed dynamic informational Web-space. *Proc. of the 12-th Inter. conf. on Pattern Recognition and Information Processing (PRIP'14)*, 2014, Minsk, Belarus, 2014, P.70–75.
12. Gladun A., Rogushina J. Use of Semantic Web Technologies and Multilinguistic Thesauri for Knowledge-Based Access to Biomedical Resources. *International Journal of Intelligent*



- Systems and Applications*, 2012, no 1. P.11–20. OWL-S Home Page. – <http://www.daml.org/services/>
13. *OWL-S Home Page*. — <http://www.daml.org/services/>
14. Rogushina J., Gladun A. Ontological approach to search of Web-services in the distributed environment of the Internet. *Informatics*, Minsk, no 4, 2006. P.116–127 (in Russian).
15. Gladun A., Rogushina J., Andrushevich A., Kurbatski A. Use of the Ontological Approach to Semantic Search in the Environment of the Internet of Things. *Proceedings of the 4-th Inter. conf. Open Semantic Technologies for Intelligent Systems OSTIS-2014*, Minsk, Belarus, 2014. P.353–356.

Получено 24.07.2014