

УДК 004.934

Н.В. Бабкова¹, О.В. Канищева², З.А. Кочуева³¹ НТУ "ХПИ", м. Харьков, Украина, nadienna@gmail.com;² НТУ "ХПИ", м. Харьков, Украина, olva-kanisheva@rambler.ru;³ НТУ "ХПИ", м. Харьков, Украина, kochueva@kochuev.com

МОДЕЛИРОВАНИЕ СЕМАНТИЧЕСКИХ ОТНОШЕНИЙ В ОНТОЛОГИИ

В статье рассмотрен подход к моделированию семантических отношений в онтологии, проанализированы теоретические аспекты использования онтологий в семантической паутине. В работе показаны типы связей, присущие онтологиям, дано описание онтологических отношений средствами алгебры конечных предикатов. Приведен пример родо-видового дерева для лексико-семантического класса существительных.

ЕСТЕСТВЕННЫЙ ЯЗЫК, ОНТОЛОГИИ, АВТОМАТИЗИРОВАННЫЙ СЕМАНТИЧЕСКИЙ АНАЛИЗ, СЕМАНТИЧЕСКИЕ ОТНОШЕНИЯ, АЛГЕБРА КОНЕЧНЫХ ПРЕДИКАТОВ

Введение

Развитие наукоемких областей человеческой деятельности в современном обществе сопровождается возрастанием роли компьютерных технологий. Резко возросший поток информации обусловил необходимость поиска новых способов ее хранения, представления, формализации, систематизации и автоматической обработки. Возрос интерес к всеобъемлющим базам знаний, которые можно использовать для различных практических целей. Большой интерес вызывают системы, способные без участия человека извлечь сведения из текста. Все это обусловило появление Semantic Web (SW), как расширения World Wide Web. В Semantic Web гипертекстовые страницы снабжены дополнительной разметкой, в которую включены сведения о семантике элементов страницы. Неотъемлемым компонентом Semantic Web является понятие онтологии, описывающее смысл семантической разметки.

Под онтологией (в наиболее общих чертах) понимают систему понятий некоторой предметной области, которая представляется как набор сущностей, соединенных различными отношениями [1]. Онтологии используются для формальной спецификации понятий и отношений, которые характеризуют определенную область знаний. Основным преимуществом онтологий как способа представления знаний является их формальная структура, которая упрощает их компьютерную обработку.

Онтологии используются как источники данных для многих компьютерных приложений (для информационного поиска, анализа текстов, извлечения знаний и т. д.), позволяя более эффективно обрабатывать сложную и разнообразную информацию. Этот способ представления знаний позволяет приложениям распознавать те семантические отличия, которые являются само собой разумеющимися для людей, но не известны компьютеру [2].

Онтологии получили широкое распространение в решении проблем представления знаний и инженерии знаний, семантической интеграции информационных ресурсов, информационного поиска и т. д. Интеллекту-

альные системы на основе онтологий показали на практике свою эффективность, однако построение онтологий требует экспертных знаний в исследуемой предметной области и занимает существенный объем времени (возникает так называемая проблема «бутылочного горлышка»), поэтому актуальной задачей является автоматизация процесса построения онтологий.

Цель работы

Цель работы: рассмотреть существующие подходы к моделированию семантических отношений. Показать использование алгебры конечных предикатов для моделирования семантических отношений в онтологии, что позволит улучшить качество автоматического создания онтологий или автоматического пополнения онтологий и, соответственно, повысить эффективность процессов обработки данных и знаний в компьютерных системах и сетях.

1. Онтологии как основа создания семантической сети

Большинство электронных документов находится в сети Интернет. Современный взгляд на Интернет предполагает создание семантической паутины – это надстройка над существующей Всемирной паутиной, которая призвана сделать размещенную в ней информацию более понятной для компьютеров. Машинная обработка возможна в семантической паутине благодаря двум её важнейшим характеристикам:

1. Повсеместное использование унифицированных идентификаторов ресурсов (URI). Традиционно в Интернете эти идентификаторы используются для установки ссылок на адресуемый объект (например, веб-страницу, файл или ящик электронной почты). Поскольку URI глобально уникальны, они позволяют называть одни и те же предметы в разных местах в семантической паутине. При этом URI протокола HTTP можно одновременно использовать как адреса документов, содержащих машиночитаемые описания этих предметов [3].

2. Использование семантических сетей и онтологий. Современные методы автоматической обработки данных, доступных в Интернете, как правило, основаны на

частотном и лексическом анализе текстового содержания, которое, прежде всего, предназначено для восприятия человеком. В семантической паутине вместо этого используется стандарт RDF, описывающий семантические сети (графы), в которых узлы и дуги имеют URI. Утверждения, кодируемые с помощью RDF, в дальнейшем можно интерпретировать с помощью онтологий, созданных по стандартам RDF Schema и OWL, чтобы получать из них логические заключения. В основе онтологий лежат математические формализмы, называемые дескрипционными логиками, которые представляют собой семейство языков, позволяющих формально и однозначно описывать понятия. Структура семантической паутины представлена на рис. 1.

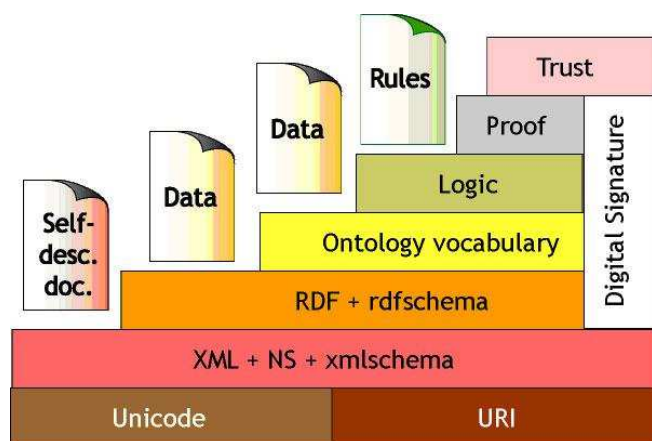


Рис. 1. Составляющие семантической паутины

Онтологии широко используются во всех областях, занимающихся обработкой данных на естественном языке. В связи с использованием онтологий в различных приложениях возникла необходимость создания стандартизированных способов их представления. Началось развитие разнообразных языков, которые могли бы применяться повсеместно во всех системах (самыми известными являются RDF и OWL). Возникло также большое количество редакторов для создания, пополнения и изменения онтологий. Каждое из этих средств, обычно, направлено на работу с определенным форматом данных и обладает своими особенностями.

Однако, несмотря на это, на сегодняшний день для решения практически любой задачи автоматической обработки естественного языка требуется онтология.

2. Семантическая сочетаемость слов

Основной лингвистической единицей, вокруг которой разрабатывается тот или иной прием определения смысла или содержания, является слово. Слова в предложении связаны друг с другом грамматически и по смыслу. Грамматические связи обеспечивают грамматическую правильность речи. Примеры грамматических связей: согласование прилагательного-определения с определяемым существительным в роде, числе и падеже, глагольное управление существительными. Задача грамматической сочетаемости, благодаря морфологическому анализу и специфическим маркерам (например, изменение окончания), является практически иденти-

фицируемой, особенно для часто используемых конструкций словосочетаний.

Смысловые связи обеспечивают правильность высказывания по смыслу и смысловые отношения слов редко выражаются при их написании. Тем не менее, употребляя слово, мы должны согласовывать его по смыслу с другими словами. Это смысловое согласование выражается в двух типах словесной сочетаемости – семантической и лексической.

Семантическая сочетаемость слова – это его способность вступать в сочетания с целыми классами слов, объединяемых общностью смысла. Например, глаголы думать, полагать, радоваться, смеяться, грустить и другие описывают различные состояния человека; значит, и сочетаться они могут лишь с такими словами, которые обозначают человека (это и есть один из семантических классов): мальчик, старик, прохожий, врач, учительница и т.п. Только в сказке или в фантастической повести возможны выражения вроде коза задумалась или шкаф засмеялся.

Лексическая сочетаемость слова – это его способность вступать в сочетания не с любым словом из какого-либо семантического класса, а только с некоторыми. Например, существует класс слов, объединяемых общим смыслом «множество, совокупность»: стадо, табун, стая, рой, косяк и т.п. При необходимости обозначить множество каких-нибудь животных мы не можем сочетать название любого животного с любым из этих слов. Говорят: стадо коров, табун лошадей, стая птиц, рой пчел, косяк рыбы (но не «стая рыбы или рыб», «стадо пчел», «рой лошадей» и т.п.).

3. Типы связей в онтологии

Считается, что связи между понятиями в реальном мире и его моделях бывают различного типа, например, таксономические связи, выражающие отношения «является видом» или отношение «общее/специфичное», композиционные связи, выражающие отношения «является частью», и «топологические» связи (Topological relationships) или проблемно-специфичные (Domain specific relations), показывающие «пути» физических взаимодействий (interactions) между понятиями или обеспечивающие информацией о пространственном их расположении [4-6]. Каждая из таких типов связей представлена в конкретных онтологиях множеством вариантов.

Таксономические связи. К наиболее типичным таксономическим отношениям относятся отношение, ставящее две сущности в зависимость «частное – общее» (является потомком) [3-5], исчерпывающее разбиение на подклассы (Exhaustive-Subclass-Partition). К таксономическим может быть отнесено и отношение «является представителем» (Instance-of) [5, 6].

Композиционные связи. Согласно [4, 6] в «искусственном интеллекте» и концептуальном моделировании различают несколько типов отношений часть-целое (part-whole), среди которых наиболее популярны:

- *компонент-объект* (component-object) – физическая или концептуальная часть [3, 5];
- *ингредиент, вещество* – для объекта (stuff-object) (например, мука в хлебе, пластик – материал для

монитора) [4, 5];

- *член из набора* (member-collection) – сущность является членом множества (набора, коллекции) однородных сущностей [3, 4]).

«Топологические» связи. Согласно анализу реальных онтологий [4, 5], «топологическими» (или проблемно-специфичными) являются, например, такие связи: сущность имеет атрибутом что-то (отношение объект – его свойство);

Следующая «группа» связей – **связи сущностей с процессами и связи сущностей с сущностями посредством процессов**: сущность поддерживает процесс, сущность является инструментом процесса, актор/сущность участвует в процессе, актор/сущность выполняет процесс, актор руководит/управляет процессом, и т.п.

Другая группа связей – **причинно-следственные связи**: процесс влияет на прохождение другого процесса (например, препятствует, оберегает от), процесс влияет на сущность, процесс вызывает другой процесс, абстракция является предпосылкой для абстракции и др.

Временные связи (или связи сущностей, протекающих во времени): процесс связан по времени с другим процессом, временной интервал является частью другого, временной интервал связан с другим интервалом, процесс является частью другого процесса.

Пространственные связи: сущность частично располагается в другой сущности, Объект пространственно связан с чем-то (находится рядом, пересекается...), организм обитает в сущности, сущность наполняет другую сущность.

Все рассматриваемые отношения являются бинарными отношениями. Для анализа структуры этого достаточно, поскольку к таковым можно свести любые парные отношения.

4. Описание онтологических отношений средствами алгебры конечных предикатов

Между словами в словосочетаниях существуют различные типы отношений (рис. 2). Отношения иерархии как разновидность сильных парадигматических отношений соответствуют отношениям подчинения или отношениям типа «выше-ниже». В пределах иерархических отношений различают отношения типа «род-вид» и «целое-часть» [7].

Отношение «род-вид» является одним из важнейших видов связей между понятиями. При этом родовым (подчиняющим) называется понятие, выражающее существенные признаки класса предметов, являющихся видами этого рода.

Соответственно видовым (подчиненным) называется понятие, которое отображает существенные признаки класса предметов, являющегося видом какого-либо рода.

В качестве математического аппарата для моделирования семантики свободных словосочетаний с родовидовыми отношениями предлагается использовать алгебру конечных предикатов (АКП). Это универсальный математический аппарат для описания детерминированных, дискретных и конечных информационных

объектов. Этот аппарат позволяет описывать различные языковые закономерности в рамках единой математической модели.

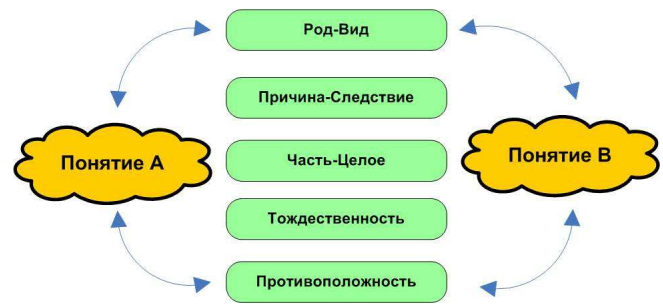


Рис. 2. Виды логических связей между понятиями

Аппарат алгебры конечных предикатов (АКП) является обобщением аппарата булевых функций и аппарата многозначной логики. АКП дает возможность перехода от алгоритмического описания лингвистических закономерностей к описанию их в виде уравнений.

Словоформы принимают свои сочетания в контексте. Контекстом в нашем случае служит словосочетание в определенной предметной области. Поэтому значение словоформы определяется лишь теми признаками, которые характеризуют ее связь со значениями слов в контексте, а также с объектами предметной области.

Семантическая сочетаемость словоформ зависит от знаний предметной области и от родо-видовой структуры [8].

Родо-видовое отношение в русском языке – это бинарное отношение R , заданное на множестве словоформ D такое, что для любого $d \in D$, всегда существует $d^1 R d$ и, кроме того, R не рефлексивно ни для одного $d \in D$, антисимметрично и транзитивно. Отношение R называют еще отношением древовидного порядка. Для иерархического порядка, если $d R d^1$, то говорят, что d^1 следует за d . Если, кроме того, нет такого d^{11} , что $d R d^{11}$ и $d^{11} R d^1$ то говорят, что d^1 непосредственно следует за d .

Использование аппарата АКП требует перехода от двухместного родо-видового отношения к n -местному. Каждому n -местному отношению R , заданному на n -той декартовой степени D^n множества словоформ D , можно поставить в соответствие каноническое уравнение, т.е. уравнение вида:

$$f_R(x_1, x_2, \dots, x_n) = 1, \quad (1)$$

которое связывает переменные x_1, x_2, \dots, x_n точно так же, как и отношение R . Предикат f_R выбирается следующим образом:

$$f_R(x_1, x_2, \dots, x_n) = \begin{cases} 1, & \text{если } (x_1, x_2, \dots, x_n) \in R \\ 0, & \text{если } (x_1, x_2, \dots, x_n) \notin R, \end{cases} \quad (2)$$

где $(1, 2, \dots, n)$ – количество уровней иерархии. Записав предикат f_R в виде формулы АКП, получаем аналитическую запись отношения R в форме уравнения (1).

Каждый уровень иерархии можно представить следующим образом:

$$x_1^1 \sim (x_2^{a_1} \vee x_2^{a_2} \vee \dots \vee x_2^{a_i}) = 1, \text{ где } (a_1, a_2, \dots, a_i) \in A,$$

$$x_2^{a_1} \sim (x_3^{b_1} \vee x_3^{b_2} \vee \dots \vee x_3^{b_k}) = 1, (b_1, b_2, \dots, b_k) \in B, m \geq k,$$

$$x_n^{c_1} \sim (x_n^{z_1} \vee x_n^{z_2} \vee \dots \vee x_n^{z_l}) = 1, (z_1, z_2, \dots, z_l) \in Z, l \geq j,$$

где A, B, C, T – соответственно подмножество словоформ, словаря D , представленных на каждом уровне иерархии. Согласно теореме о разложении [8] уравнение (1) можно записать в следующем виде:

$$(x_1^a \sim \vee x_2^a)(x_2^a \sim \vee x_3^b) \dots (x_{n-1}^c \sim \vee x_n^l) = 1.$$

$$a = 1 \quad b = 1 \quad l = 1$$

Слова, связанные родо-видовыми отношениями, представляют собой лексико-семантические классы с иерархической структурой, описываемой в виде родо-видовых деревьев [7]. Родо-видовое дерево представляет собой иерархию элементов, называемых узлами. В роли узлов выступают словоформы. Пример родо-видового дерева представлен на рис. 3.

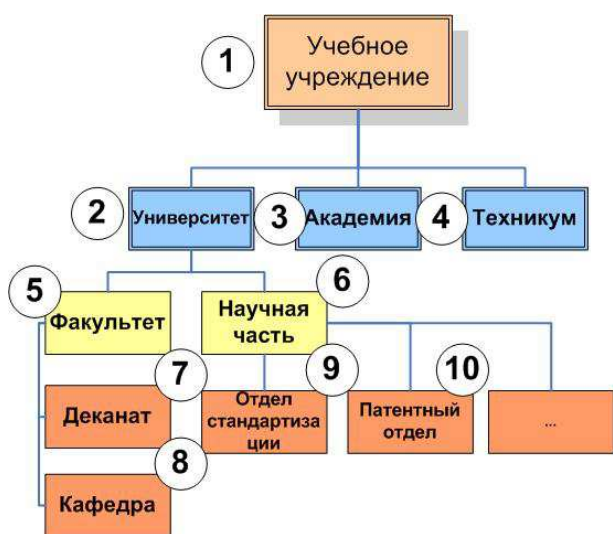


Рис. 3. Родо-видовое дерево лексико-семантического класса существительных

В зависимости от контекста родо-видовое дерево с этими же словоформами будет иметь другой вид. Для приведенного дерева (рис. 3) введем переменные и области их допустимых значений. Пусть $x_1 \in \{1\}$, $x_2 \in \{2,3,4\}$, $x_3 \in \{5,6\}$, $x_4 \in \{7,8,9,10\}$.

Уравнение, описывающее дерево, имеет следующий вид:

$$(x_1^1 \sim x_2^2 \vee x_2^3 \vee x_2^4)(x_2^2 \sim x_3^5 \vee x_3^6)$$

$$(x_3^5 \sim x_4^7 \vee x_4^8)(x_3^6 \sim x_4^9 \vee x_4^{10}) = 1.$$

Решение данного уравнения соответствует узлам родо-видового дерева.

Выводы

В работе показано использование алгебры конечных предикатов для моделирования семантических отношений в онтологии, что позволит улучшить качество автоматического создания онтологий или автоматического пополнения онтологий и, соответственно, повысить эффективность процессов обработки данных и знаний в компьютерных системах и сетях. В работе представлено только один вид связи "род-вид". В даль-

нейшем планируется описать и другие типы связей. Моделирование семантических отношений может использоваться в задаче автоматического создания онтологии и других задачах автоматической обработки языка.

Список литературы: 1. *Simperl E., Mochol M.* Achieving Maturity: the State of Practice in Ontology Engineering in 2009. / In International Journal of Computer Science and Applications, Technomathematics Research Foundation Vol. 7 No. 1, pp. 45-65, 2010. 2. *Faatz A. and Steinmetz R.* Ontology enrichment with texts from the WWW. Semantic Web Mining 2nd Workshop at ECML/PKDD-2002, 20th August 2002, Helsinki, Finland 3. *Towards the Semantic Web: Ontology-driven Knowledge Management.* – John Wiley & Sons, 2003. – ISBN 0470848677. 4. *Gruber T.* What is an Ontology? <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>. 5. *Rogers J.E.* GALEN's Model of Parts and Wholes: Experience and Comparisons / Rogers J.E., Rector A.L. // Journal of the American Medical Informatics Association, Fall Symposium Special Issue, pp. 819-823, November, 2000. 6. *Нариньяни А. С.* ТЕОН-2: От тезауруса к онтологии и обратно // http://www.dialog-21.ru/archive_article.asp?param=7360&y=2002&vol=6077. 7. *Замаруева И. В.* Математические модели семантики свободных словосочетаний с родо-видовыми компонентами и их применение в АИС. – Дис... кан. техн. наук. – Харьков : ХТУРЭ, 1990. – 170 с. 8. *Шаронова Н.В.* Математические модели суффиксального словообразования и их использование в системах автоматической обработки текстов на русском языке. – Дис... кан. техн. наук. – Харьков : ХТУРЭ, 1984. – 217 с.

Поступила до редколлегии 04.03.2013

УДК 004.934

Моделивання семантичних відношень в онтології / Н. В. Бабкова, О.В. Каніщева, З.А. Кочуєва // Біоніка інтелекту: наук.-техн. журнал. 2013. №1 (80). С. 00-00.

У статті розглянутий підхід до моделювання семантичних відносин в онтології, проаналізовані теоретичні аспекти використання онтологій у семантичній павутині. У роботі показані типи зв'язків, які властиві онтологіям, дано опис онтологічних відносин засобами алгебри кінцевих предикатів, що дозволить поліпшити якість автоматичного створення онтологій або автоматичного поповнення онтологій. Наведено приклад родо-видового дерева для лексико-семантичного класу іменників.

Л. 3. Бібліогр.: 8 найм.

UDK 004.934

Modeling semantic ontology relationships / N. V. Babkova, O.V. Kanishcheva, Z.A. Kochueva // Bionica Intellecta: Sci. Mag. 2013. № 1(80). P. 00-00.

The paper discussed an approach to modeling semantic ontology relationships, analyzed the theoretical aspects of the use of ontologies in the semantic web. The paper shows the types of relationships that are inherent in the ontology, a description of the means of ontological predicates of finite algebra, which will allow to improve the quality of automatic creation of ontologies and automatic replenishment of ontologies. The example of genus-species tree for the lexical-semantic class of nouns.

Fig. 3. Ref.: 8 items.