

К ВОПРОСУ СИСТЕМНО-ОНТОЛОГИЧЕСКОЙ ИНТЕГРАЦИИ ЗНАНИЙ ПРЕДМЕТНОЙ ОБЛАСТИ

Abstract: Some aspects of system integration of knowledge of a subject domain and their orientation to are researched in this work knowledge-based of information technologies of natural language processing objects.

Key words: ontology, natural language processing, an ontology-language picture of the worlds.

Анотація: У роботі досліджено деякі аспекти системної інтеграції знань предметної області та їх орієнтацію на ефективну реалізацію знанняорієнтованих інформаційних технологій обробки природномовних об'єктів.

Ключові слова: онтологія, комп'ютерна обробка природної мови, мовно-онтологічна картина світу.

Аннотация: В работе исследованы некоторые аспекты системной интеграции знаний предметной области и их ориентация на эффективную реализацию знаниеориентированных информационных технологий обработки естественных языковых объектов.

Ключевые слова: онтология, компьютерная обработка естественного языка, языково-онтологическая картина мира.

1. Введение

Разработка и коммерческое использование баз знаний (БЗ) и соответствующего инструментария во многих прикладных областях (Knowledge based engineering systems) на основе онтолого-управляемых информационных систем во многих научно-исследовательских центрах и корпорациях привели к росту теоретических разработок формализованных методологий проектирования онтологических структур. При этом суть указанных методологий сводилась к формальному обоснованию структурирования иерархического дерева онтологии (формализованного построения наборов концептов и связывающих их концептуальных отношений или категоризации) и разработке формальных языков представления знаний, которые описывают аксиоматизацию концептов предметной области (ПрО). Следующим шагом развития теории баз знаний являлась необходимость теоретически обоснованного объединения (или *системной интеграции*) уже разработанных как общедоступных онтологий, так и коммерческих баз знаний для разнообразных прикладных задач, проблем, целых предметных областей и мультидисциплинарных знаний общего характера.

2. Постановка задачи

В общем виде процесс, обеспечивающий системную интеграцию множества онтологий, можно описать следующей формулой:

$$O = \prod_i^U O_i, \quad i = \overline{1, N}, \quad (1)$$

где \prod_i – знак концептуального объединения. Смысл этого знака состоит в системной интеграции исходных онтологических графов (ОГ) с учётом областей определений $O_i (i = \overline{1, N})$ и их взаимосвязи (взаимодействия).

Обобщенная архитектура знаниеориентированной системы, эффективно реализующей цепочку технологий *Natural Language Processing (NLP) → Knowledge Representation (KR) →*

Knowledge Processing (KP), в том числе процедуру системной интеграции знаний в предметных областях, представлена на рис. 1.

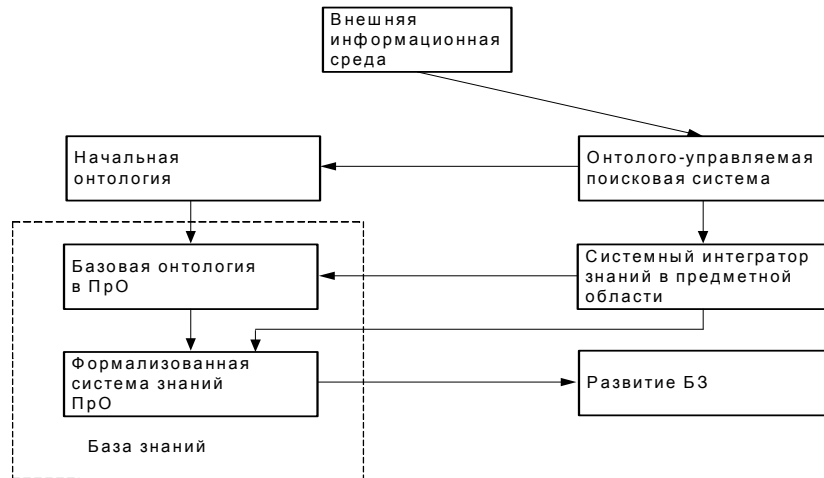


Рис. 1. Архитектура развивающихся знаниеориентированных систем

Реализация указанных технологий и процесса системной интеграции знаний предполагает учёт различных формально-методологических требований, критериев и оценок. Приведём основные из них.

1. Сущность методов онтологического исследования как раздела системного анализа предполагает:

- а) построение статических и динамических моделей;
- б) исследование структуры ПрО и взаимосвязи компонент знаний;
- в) исследование поведения системы, функционирующей в ПрО – разработка алгоритмов процедур композиций знаний и диаграмм состояний.

2. Влияние предметной области на вид онтологии.

3. Правила формирования достоверных утверждений и выводов, описывающих систему и её поведение.

4. Категориальный уровень онтологии ПрО и соответствующая ему система отношений.

5. Параметры онтологического дерева (средний коэффициент ветвления i -ой вершины, количество уровней ветвления, количество типов отношений, в том числе по подтипам).

3. Компоненты (онтологии) интеграции знаний

В настоящее время в мире известны несколько широко используемых онтологий, на основе которых разрабатываются информационные системы управления базами знаний в различных предметных областях. Кратко рассмотрим их.

WordNet. Онтология разработана в Принстонском университете (штат Нью Джерси), свободно распространяется в сети Интернет и представляет собой лексикографическую базу данных, структурированную в соответствии с лингво-семантическими отношениями: синонимия, антонимия, гипонимия, меронимия и морфологические отношения [1, 2]. База данных разделена по частям речи на существительное, глагол, прилагательное, наречие и функциональные слова. Части речи организованы в иерархию узлов, где каждый узел представляет смысл слова, или синсет. У WordNet существует более 66 тыс. синсетов существительных, около 18 тыс. синсетов

прилагательных, более 3600 синсетов наречий и более 12 тыс. синсетов глаголов (на сегодняшний день она включает более 130 тыс. английских слов и фраз).

Технически WordNet есть электронный тезаурус, определяющий широкий класс значений слов, связанных между собой семантическими указателями. Логическая структура WordNet показана на рис. 2.

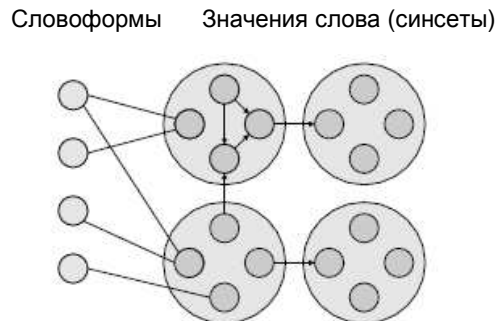


Рис. 2. Логическая структура WordNet

Рассмотрим пример записи синсета в базе данных.

```
00047131 04 n 02 accession 0 addition 0 001 @
09536731 n 0000 | something added to what you
have already; "the librarian shelved the new accessions";
"he was a new addition to the staff"
```

Первая часть записи утверждает, что число 00047131 является уникальным идентификатором синсета существительного {accession, addition}. Часть записи между символами "@" и "|" указывает, что этот синсет непосредственно подчинён синсету с идентификатором 09536731. Последний синсет соответствует значению "acquisition". И последняя часть записи (после символа "|") представляет толкование синсета и некоторые примеры употребления слов, входящих в синсет.

Онтология WordNet очень популярна в научных исследованиях по компьютерной обработке естественного языка и поиску информации в сети (Information Retrieval). Она постоянно обновляется, и уже доступна вторая редакция программного продукта. Большое количество примеров её использования приведено в [1].

Sensus. Онтология разрабатывалась группой исследователей по компьютерной обработке естественного языка (Natural Language Group) из Information Sciences Institute of University of Southern California (ISI USC). Основными направлениями исследований группы являются машинный перевод (Machine Translation) и реферирование текстов (Text Summarization). Онтология Sensus представляет собой расширение и реорганизацию основных ветвей дерева WordNet и содержит более 70 тыс. узлов таксономически упорядоченных концептов [3]. Кроме того, на верхнем уровне добавлены узлы из онтологии Penman Upper Model. Для Sensus разработаны эффективные алгоритмы реализации онтолого-связанных перекрестных ссылок из нескольких языков к концептам онтологии, существенно упрощающих соотнесение слов в системе машинного перевода. Для онтологии разработан Ontosaurus – браузер для просмотра.

Omega. Онтология разрабатывалась также в ISI USC как развитие Sensus. Она представляет собой объединение и реструктуризацию онтологий WordNet 2.0 и Mikrokosmos (ориентированную на испанский язык) и содержит около 120 тыс. концептных узлов, 156 тыс.

англоязычных наименований, 28 тыс. испаноязычных наименований и 270 тыс. смыслов [4]. Omega может быть охарактеризована как поверхностная, лексически ориентированная онтология, большинство таксономических терминов которой соответствуют её концептам, выраженным одним словом в английском языке. Онтология не содержит формальных концептных определений, а только связывает концепты несколькими внутренними взаимными связями (семантическими отношениями). Основными прикладными областями применения для неё являются вопросно-ответные системы и информационная интеграция.

Главными компонентами Omega, как указывалось, являются онтологии WordNet 2.0 и Mikrokosmos, имеющие взаимно противоположные преимущества и недостатки. Первая имеет широко развитую структуру среднего и нижнего уровней, хорошо обеспеченные лексикографическими данными, в то время как вторая имеет намного меньшую сетевую структуру, но богато развитые концептуальные связи категориального уровня. Результатом декомпозиции указанных онтологий (с учётом преимуществ и недостатков) является структура онтологии Omega. При этом верхний уровень иерархии получил название New Upper Model и представлен на рис. 3.

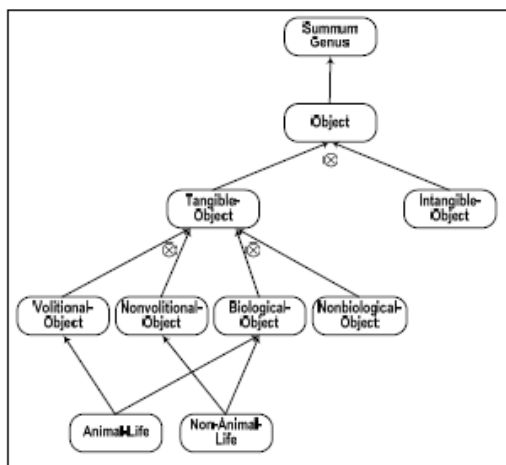


Рис. 3. Верхний уровень онтологии Omega

Подобно многим онтологиям, ядром Omega является сеть концептов, связанных множеством приписанных взаимоотношений. Пространство концептов объединено основными отношениями, такими как IS-A, PART-OF, SUBSTANCE-OF и ELEMENT-OF, и вспомогательными THEME, INSTRUMENT PERTAINS-TO. Концепты также обладают несемантическими атрибутами, например, GLOSS [4].

Mikrokosmos. Онтология разработана в Computing Research Laboratory (CRL) University of New Mexico State, ориентирована (в первую очередь) на лексику испанского языка и содержит лексикон из 7 тыс. слов, поддержанных в онтологии более чем 5 тыс. концептов [5, 6]. Она определяется как модель мира и основные знания о мире, иерархически структурированные как направленный онтограф или дерево со многими связями. Все знания разделены на две (внутри связанные) базы знаний. Первая БЗ относится к онтологии и содержит знания о концептах. Вторая база знаний, названная ономастиком, идентифицирует экземпляры тех концептов, которые включены в тело общих, но доменно-зависимых знаний.

На самом вершине онтологии концепты разделены на Object, Event и Rproperty и организованы в простую таксономию (IS-A иерархия). В реальной онтологии каждый узел может иметь некоторые другие связи, структурирующие его с другими узлами в более сложную онтологию, например, IS-PART-OF, IS-AN-OCCUPANT-OF, MANUFACTURED-BY и др. Эти связи (отношения) представлены как слоты на концептах, или графически, как помеченные связи между концептами. Например, концепт EAT может иметь слоты, такие как AGENT и THEME (описывающие того, кто ест и съеденные продукты).

В определении отношения существуют ограничения, не допустимые размещения концептов в слот для конкретных прикладных областей; эти ограничения также являются концептами из онтологии. Ограничения на разрешенные заполнители различных слотов называются семантическими фасетами слота, в то время как заполнители сами по себе являются фасетами значения. Аксиоматика онтологии состоит из 36 аксиом, записанных формулами исчисления предикатов первого порядка. Аксиомы определяют, что представляет правильное и непротиворечивое представление в онтологии, а что нет. На рис. 4 приведены первые два уровня онтологии Mikrokosmos.

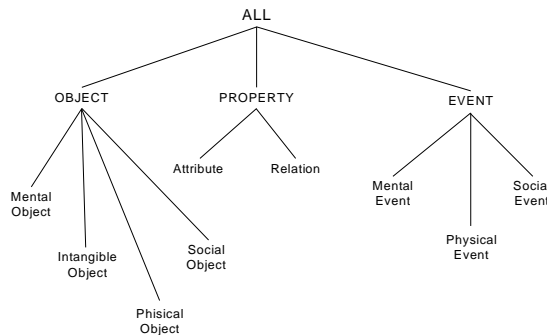


Рис.4. Верхний уровень онтологии Mikrokosmos

с концептуальными отношениями и описана на формальном языке Сус. База знаний Сус, разбитая на уровни по степени общности, изображена на рис. 5 и содержит от небольшого (по количеству категорий) уровня абстрактных концептов наверху и широкого уровня реальных фактов внизу. Посредине находятся общие теории для группы прикладных областей и конкретные (доменно-зависимые) теории.

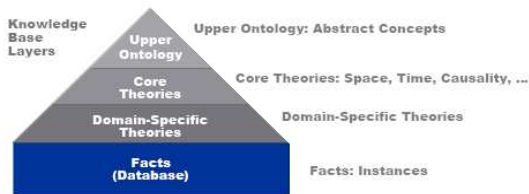


Рис. 5. Структура базы знаний Сус

Сус включает около 87 тыс. разных событийных типов, которые описывают большинство событий, происходящих в мире. Их ранг колеблется от наиболее общих до специфических, конкретных. Информация в Сус представлена тремя общими типами: абстрактные последовательности и знаки, пропозициональное содержание и концептуальные механизмы-согласователи. На рис. 6 представлена организация тем верхнего уровня в онтологии Сус.

DOLCE. Представляет собой дескриптивную онтологию для лингвистики и когнитивного инжиниринга. Она разработана в Laboratory for Applied Ontology of Institute of Cognitive Science and Technology, Italy и представляет первый модуль в библиотеке базовых онтологий (WonderWeb Foundational Ontologies Library) [8]. Указанная библиотека должна служить как:

- отправная точка для построения новых онтологий. Одним из наиболее важных и критических вопросов при проектировании новой онтологии есть определение того, какие сущности входят в прикладную область и должны быть смоделированы. Адаптация верхнего уровня онтологии существенно упрощает нахождение ответа на этот вопрос;

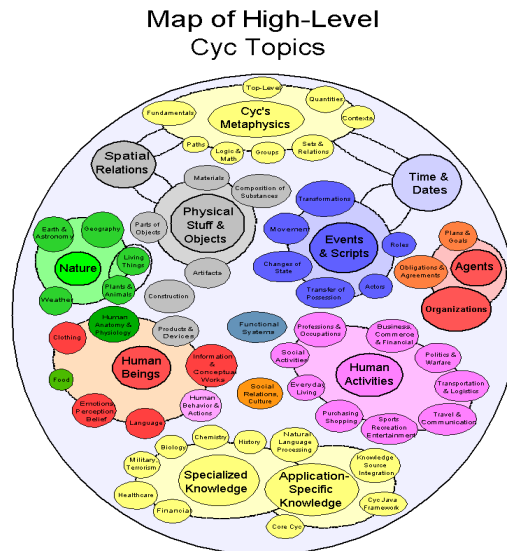


Рис. 6. Карта верхнего уровня Cyc

- базовая структура для анализа и интеграции существующих онтологий и стандартных метаданных (отображающих существующие концепты в категории, принятые в некотором модуле (-ях) библиотеки).

Кроме того, библиотека должна быть:

- минимальной – как противоположность другим всеобъемлющим онтологиям верхнего уровня. В библиотеку включены только многократно используемые и широко применяемые категории верхнего уровня;
- строгой – широкое использование формальных аксиоматизаций;
- экстенсивно исследованной – каждый модуль при добавлении в библиотеку проходит всестороннюю экспертную оценку; затем он адаптируется и снабжается соответствующими ссылками.

Онтология DOLCE основывается на функциональном различии между “продолжительностями” и “событийностями”. Главное отличие между сущностями указанных категорий состоит в том, что первые существуют во времени, а вторые случаются во времени. Главным отличием между сущностями первого и второго вида является отношение участия. Первые два уровня онтологии DOLCE представлены на рис. 7.

PropBank. Является онтологией глаголов английского языка и представляет собой развитие TreeBank II, разработанной в университете штата Пенсильвания. Она содержит около 3200 глагольных лемм, описывающих более 4300 главных смыслов [9, 10]. Система ориентирована на синтаксический и семантический анализ входного текста и построение простых предикатно-аргументных структур.

Конечной целью разработчиков системы является онтолого-информационная интеграция PropBank с FrameNet и VerbNet (основанная на классах глаголов Левина).

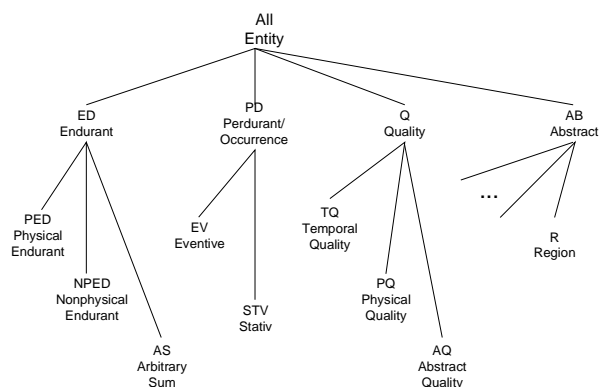


Рис. 7. Таксономия базовых категорий DOLCE

FrameNet. Онтология разработана в рамках проекта “Tools for Lexicon Building” в международном институте информатики в Беркли, Калифорния. Основным назначением проекта является связывание лингвистического корпуса (100 млн. British National Corpus) в семантические и синтаксические обобщения, представление валентностей целевых слов (в основном существительных, прилагательных и глаголов), в которых смысловая часть предполагает использование фреймовой

семантики. Результирующая база данных содержит описания семантических фреймов, лежащих в основе значений описанных слов, представление валентности (семантической и синтаксической) нескольких тысяч слов и фраз, при этом каждая сопровождается репрезентативным набором аннотированного корпуса. База данных содержит более 5 тыс. лексических входов совместно с аннотированным лингвистическим корпусом и представлена в формате, совместимом для интеграции с другими приложениями [11, 12].

В настоящее время разрабатывается вторая очередь *FrameNet II*, включающая около 900 фреймов [13]. Фрейм представляет собой множество концептов, ассоциированных с событием или состоянием и упорядоченных от простого к сложному. Для каждого фрейма определено около 10 множеств ролей или аргументов, названных фреймовыми элементами (FE). Слово может быть соотнесено к фрейму, а его синтаксические зависимости могут заполнять FE слоты. Семантические отношения между фреймами зафиксированы у фреймовых отношениях, каждый с FE ↔ FE отображением. Все фреймы покрыты более 10 тыс. лексических единиц или смыслов слов. Для них включено более 135 тыс. аннотированных примеров предложений, используемых для построения обучающих выборок у *FrameNet*.

SUMO. Начиная с 2000 года, исследователи из различных областей знаний, таких как искусственный интеллект, информатика, лингвистика, библиотечное дело и др., объединились под эгидой IEEE комитета в рабочую группу по созданию стандарта онтологии верхнего уровня (SUO WG). Под онтологией они понимают формальную, эксплицитно заданную спецификацию совместно используемой концептуализации. Она является абстрактной моделью некоторой совокупности явлений в мире, эксплицитно представленных как концепты, отношения и ограничения, которые являются машинно-читаемыми и встроенными в концептуальные знания некоторой области знаний. Позже проект был одобрен и рабочая группа получила шифр P1600.1 [14, 15]. Объем онтологии должен составлять от 1000 до 2500 терминов и около 10 аксиоматизированных утверждений для каждого термина.

Основными целями SUO являются:

- проектирование новых баз знаний и баз данных. Разработчики могут определять новые знания и новые элементы данных в терминах общей онтологии и таким образом получить функциональную совместимость с другими аналогичными системами;

- многократное использование и интегрирование совместимых баз данных. Элементы данных из существующих систем могут только однократно быть отображены в общей онтологии;
- интегрирование проблемно-ориентированных онтологий. Такие онтологии (совместимые с SUO) смогут взаимодействовать между собой на основе совместно используемых терминов и определений (в таком смысле некоторой аналогией является итальянский проект по созданию библиотеки онтологий WonderWeb Foundational Ontologies Library и её первого модуля DOLCE).

Рабочий проект SUO был создан на базе корпорации Teknowledge Corporation и получил название Suggested Upper Merged Ontology (SUMO). SUMO объединяет общедоступные, широкоизвестные онтологические контенты (онтологии на серверах Ontolingua и Institute of Cognitive Sciences and Technology (ICST, Italy), онтологию верхнего уровня Дж. Совы [16] и некоторые другие). Языком представления знаний в SUMO был выбран KIF-язык, который после доработок получил название SUO-KIF. Он предложен как самостоятельная единица стандарта [17]. Рабочая группа с начала разработки проекта уже выпустила многочисленные версии SUMO (в настоящее время доступна версия 1.75). SUMO включает в себя две части: онтологию верхнего уровня (переработанные онтологии Дж. Совы и Рассела-Норвига) и онтологию среднего уровня (реструктурированные остальные онтологии). После завершения создания каждой части в отдельности они были объединены в единую концептуальную структуру. На рис. 8 приведен верхний уровень онтологии SUMO.

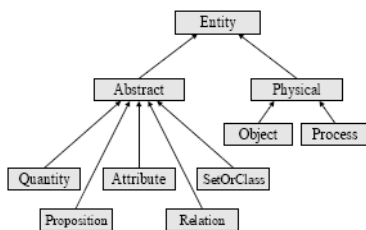


Рис. 8. Категории верхнего уровня SUMO

Помимо онтологии верхнего и среднего уровней, разработчики SUMO предоставляют пользователям доменно-зависимые онтологии, такие как Communications, Countries and Regions, distributed computing, Economy, Finance, engineering component, Geography, Governement, People, Transportations, World Airports и другие.

Языково-онтологическая картина мира (ЯОКМ). Представляет собой лингвистическую онтологию, включающую все полнозначные лексемы украинского языка, а верхний уровень которой адаптирован из онтологии Дж. Совы. Подробно ЯОКМ описана в ряде работ [18 – 22].

Некоторые данные, характеризующие описанные онтологии, приведены в таблице.

Таблица. Основные характеристики онтологий

№ п/п	Наименование онтологии	Тип онтологии	Разработчик	Браузер	Количество концептов (лексем)	Формальная аксиоматизация	Интеграция с другими онтологиями
1	WordNet	Лингвистическая	Princeton University	WordNet	Около 100 тыс.	-	SUMO, FrameNet
2	Sensus	Смешанная	ISI USC	Ontosaurus	Более 70 тыс.	-	-
3	Omega	Лингвистическая	ISI USC	Omega	Около 120 тыс.	-	WordNet, Mikrokosmos
4	Mikrokosmos	Лингвистическая	CLR UNMS	Mikrokosmos	Более 7 тыс.	+	WordNet, Omega

5	OpenCyc	Концептуальная	Cycorp	OpenCyc	Более 100 тыс.	+	WordNet
6	DOLCE	Концептуальная	LAO ICST	DOLCE	Около 4 тыс.	+	-
7	PropBank	Смешанная	University PennState	-	Более 4300	-	FrameNet VerbNet
8	FrameNet	Смешанная	ISI, Berkeley, CA	Frame-Grapher	Около 900 фреймов	+	WordNet PropBank SUMO
9	SUMO	Концептуальная	Teknowledge Corporation, SUO WG	SUMO	Более 1000	+	FrameNet, WordNet, EMELD
10	ЯОКМ	Лингвистическая	ИК НАНУ	-	Более 1000	-	-

4. Интеграция онтологий

4.1. Формализованные аспекты интеграции и её оценка

Объём знаний W в предметных областях можно оценить через характеристики (параметры) их формально-онтологических представлений. В частности, при представлении онтологическим графом (без учёта типов отношений и сложности функций интерпретации) величина W может характеризоваться числом вершин ОГ. В случае простой древовидной структуры это число может быть выражено формулой

$$W = \sum_i \sum_h \sum_l O_i \cdot S_{h,l}, \quad (2)$$

где O_i – онтограф i -ой предметной области, $i = \overline{1, N}$; $S_{h,l}$ – степень вершины, равная числу исходящих из неё рёбер; $h = \overline{1, H}$ – количество уровней ОГ; $l = \overline{1, L_h}$ – номер вершины на соответствующем (h -ом) уровне ОГ.

При равномерной плотности распределения ОГ, т.е. при $S_{h,l} = S$ ($h, l = 1, 2, \dots$), (2) сводится к известной формуле суммы геометрической прогрессии

$$W = \sum_i O_i \frac{1-S^h}{1-S}. \quad (3)$$

Учёт типов отношений и сложность функций интерпретации приводит к ОГ со взвешенными вершинами и ребрами. Выражение (2) при этом преобразуется в вид

$$W = \sum_i \sum_h O_i \cdot \left(\alpha_l + \sum_j \beta_{l,j} \right), \quad (4)$$

где α_l и $\beta_{l,j}$ – значения весовых функций соответствующих отношений и функций интерпретации, приписанные вершинам (α_l) и ребрам ($\beta_{l,j}$) ОГ. Выражение (4) даёт полную

оценку сложности ОГ, а отношение $\omega = W^O / W$ характеризует среднюю плотность взвешенного ОГ.

Рассмотренные оценки позволяют сравнивать различные варианты представления знаний о предметных областях, а также отслеживать процесс эволюции научных теорий.

Процесс развития знаний в любой ПрО связан с её анализом, концептуализацией и построением формальной теории. При этом формализация в общем случае относится к четырём основным видам представления информации:

$$I = I(V, A, T, G), \quad (5)$$

т.е. к вербальному (V), аналитическому (A), табличному (T) и графическому (G). Между ними существует взаимно-однозначное соответствие, не всегда на практике реализуемое строго и полностью. Поэтому все они находят своё, вполне определённое место при описании научной теории. Во многих случаях корректным оказывается ограничение этих видов до двух: вербальный и аналитический.

Как правило, процесс развития теорий сопровождается перераспределением объёма информации о предметной области между вербальной и формальной компонентами, т.е. между естественно-языковым описанием предмета исследования и формально-аналитическим A (формульным Φ , табличным T , графическим G представлением существа теории). Очевидно, что формализованное представление является более компактным, а главное, более строгим и пригодным для компьютерной обработки.

4.2. Практическая реализация интеграции онтологий

Накопившийся опыт практического использования SUMO и других онтологий (в первую очередь, WordNet и FrameNet) показал необходимость в объединении двух и более онтологий (построении взаимных отображений концептов) в единую концептуальную структуру. Два вида онтологий (лингвистическая и концептуальная) отражают две стороны концептуализации реального мира. Если у WordNet концептуализация отображается в терминах естественного языка, то у SUMO концепты организованы в логическую структуру. При этом объединение предполагает построение отображения между соответствующими узлами различных онтологий.

Отображение SUMO \leftrightarrow WordNet. Рабочая группа SUO WG разработала такое отображение. Оно сопоставляет каждому синсету в базе данных WordNet путём теггирования соответствующий концепт в SUMO. При этом вид отношения между синсетами WordNet и концептами SUMO может быть: синсет эквивалентен концепту; синсет является более высшим классом, чем концепт, и синсет является элементом концепта. Разработанные файлы отображения позволяют соотнести слова естественного языка в термины SUMO, используя синсеты WordNet как промежуточный уровень [23].

Приведём пример для первого вида отношения между синсетами WordNet и концептами SUMO:

```
00008864 03 n 03 plant 0 flora 0 plant_life 0 027
@ . . . | a living organism lacking the power of
locomotion &%Plant=
```

Префикс '&%' указывает на то, что термин взят из онтологии SUMO, а суффикс '= ' указывает, что отношение отображения - синонимия.

Мотивацией для построения таких отображений может быть следующее.

Отображения могут функционировать как индексы естественного языка к концептам в онтологии, как мост между этими структурированными концептами и нетекстовой лингвистической структурой. Разработчики создали специальный инструмент и встроили его в SUMO-браузер, который позволяет пользователю задавать термины естественного языка и видеть на экране компьютера соответствующие концепты онтологии SUMO с соответствующей формальной аксиоматизацией. Это позволит значительно упростить процесс инжиниринга знаний и моделирования данных. Кроме того, отображения могут служить важным источником знаний для применений NLP, в частности, при реферировании текстов и семантического поиска документов в сети. С помощью соответствующего инструментария можно проверить на полноту степень завершенности проектируемой онтологии.

Отображение SUMO ↔ FrameNet. Онтологии SUMO и FrameNet являются относительно сформировавшимися средствами, но представляется целесообразным объединить их сильные стороны для существенного упрощения NLP. В частности, NLP-применения, использующие FrameNet, требуют знаний о возможных заполнителях элементов фрейма (FEs). Например, семантическому фреймовому анализатору необходимо знать, является ли определённый фрагмент текста (или именованная сущность) строгим заполнителем для FE, т.е. ему необходимо проверить, является ли тип заполнителя FE совместимым с типом именованной сущности. Поэтому у FrameNet представлено около 40 семантических типов (FT) как ограничения на заполнители FEs и имеющие соответствия с SUMO классами как FTs [24]. Фрагмент связывания SUMO и FrameNet представлен на рис. 9.

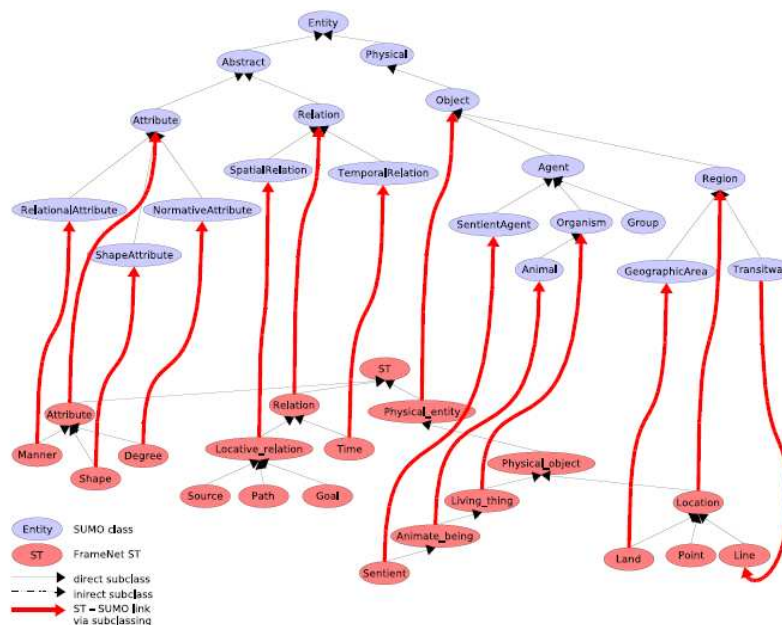


Рис. 9. Фрагмент связывания SUMO и FrameNet

Следует отметить, что существуют и другие проекты связывания лексикон-онтологии. Но рассмотренное связывание SUMO и FrameNet имеет ряд преимуществ. Дело в том, что FrameNet, в

отличие от WordNet, моделирует семантические и синтаксические валентности, снабжая их высококачественными аннотациями. Фреймовые семантики, естественно, представляют кросс-лингвистические абстракции и нормализацию парафраз. Что касается SUMO, то она значительно шире DOLCE, а в отличие от Сус является открытой для доступа.

При интеграции знаний следует учитывать прикладные аспекты. Принципиально различают области:

а) взаимодействия с пользователем – когнитивизация представления, поиск информации, создание начальной онтологии и начальной системы знаний;

б) развития системы знаний на основе имеющихся, генерация новых знаний.

5. Выводы

В работе исследованы вопросы системной интеграции онтологических знаний прикладных областей, в том числе предложена архитектура развивающихся знаниеориентированных систем, эффективно реализующих технологии компьютерной обработки знаний. Выполнен анализ широко используемых концептуальных и лингвистических онтологий, обобщены их системные характеристики, непосредственно влияющие на качественные показатели системной интеграции знаний. Анализ данных по известным проектам, ориентированным на реализацию системной интеграции онтологических знаний, показал, что наиболее востребованными являются онтологии: из концептуальных – SUMO, лингвистических – WordNet и смешанных – FrameNet. Очевидно существенное преобладание англоязычных проектов по сравнению с другими языками. Выполнен обзор известных технологий и их объединений, ориентированных на поддержку процедур манипуляции знаниями.

Основные результаты проведенного исследования предполагают выполнение дальнейших работ по разработке формальной методологии проектирования онтологии предметной области, алгоритмов и процедур системной интеграции знаний и соответствующих инструментальных средств, ориентированных на флективные языки.

СПИСОК ЛИТЕРАТУРЫ

1. Fellbaum, Christiane. WordNet: An Electronic Lexical Database // MIT Press. – 1998. – 445 p.
2. Miller, George A. (1995) WordNet: a lexical database for English // Communications of the ACM 38: 11, 39-41. – available at <http://www.cogsci.princeton.edu/-wn/>.
3. <http://www.isi.edu/natural-language/projects/ontologies.html>.
4. Philpot A., Hovy E. and Pantel P. The Omega Ontology // Information Sciences Institute of University of Southern California. – 2005. – 8 p. – or available at <http://www.isi.edu/div3/div3/pubs/papers/philpot/2005omega.pdf>.
5. <http://crl.nmsu.edu/Research/Projects/mikro/htmls/asis.paper-htmls/node1.html>.
6. <http://crl.nmsu.edu/Research/Projects/mikro/htmls/asis.paper-htmls/node4.html>.
7. <http://www.opencyc.org/doc>.
8. Masolo C., Borgo S., Gangemi A., Guarino N., Oltramari A. WonderWeb Deliverable D18: Ontology Library (final) // Laboratory For Applied Ontology – ISTC-CNR. – 2003. – 349 p. – available at <http://www.loa-cnr.it/Papers/D18.pdf>.
9. Giuglea A., Moschitti A. Knowledge Discovering using FrameNet, VerbNet and PropBank. – 2004. – 6 p. – available at <http://olp.dfki.de/pkdd04/giuglea-final.pdf>.
10. Kingsbury P., Palmer M. PropBank: the Next Level of the TreeBank // University of Pennsylvania, Department of Computer and Information Science. – 2003. – 12 p. – available at http://w3.msi.vxu.se/~rics/TLT2003/doc/kingsbury_palmer.pdf.
11. Fillmore C.J. Frame semantics and the nature of language // Annals of the New York Academy of Sciences. – 1976. – Vol. 280. – P. 20–32.
12. Baker C.F., Fillmore C.J., Lowe J.B. The Berkeley FrameNet Project // In proceeding of the COLING-ACL. – Montreal, Canada, 1998. – 6 p. – available at <http://framenet.icsi.berkeley.edu/~framenet/papers/acl98.pdf>.
13. <http://framenet.icsi.berkeley.edu/FrameGrapher>.
14. SUO, (2001). The IEEE Standard Upper Ontology web site. <http://suo.ieee.org>.

15. Niles I., Pease A. Towards a Standard Upper Ontology // In proceeding of the 2nd International Conference on Formal Ontology and Information Systems (FOIS-2001), Welty C. and Smith B., eds. – Ogunquit, Maine. – 2001. – 17–19 October. – 8 p. – available at <http://home.earthlink.net/~adampease/professional/FOIS.pdf>.
16. Sowa, John F. Knowledge Representation: Logical, Philosophical and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA, 2000. – 594 p.
17. <http://suo.ieee.org/suo-kif.html>.
18. Палагин А.В. Организация и функции «языковой» картины мира в смысловой интерпретации ЕЯ – сообщений // Information Theories and Application. – 2000. – Vol. 7, № 4. – С.155–163.
19. Палагин А.В., Яковлев Ю.С. Системная интеграция средств компьютерной техники. – Винница: «УНІВЕРСУМ-Вінниця», 2005. – 680 с.
20. Палагін О.В., Петренко М.Г. Модель категоріального рівня мовно-онтологічної картини світу // Математичні машини і системи. – 2006. – № 3. – С. 91–104.
21. Палагін О.В., Петренко М.Г. Архітектурно-онтологічні принципи розбудови інтелектуальних інформаційних систем // Математичні машини і системи. – 2006. – № 4. – С.15–20.
22. Палагін О.В., Петренко М.Г. Розбудова абстрактної моделі мовно-онтологічної інформаційної системи // Математичні машини і системи. – 2007. – № 1. – С. 42–50.
23. Niles I., Pease A. Linking Lexicons and Ontologies: Mapping WordNet to the suggested Upper Merged Ontology // Proc. of the 2003 International Conference on Information and Knowledge Engineering (IKE2003). – Las-Vegas, Nevada. – 2003. – June 23–26. – 6 p. – available at <http://home.earthlink.net/~adampease/professional/Niles-IKE.pdf>.
24. Scheffczyk I., Pease A., Ellsworth M. Linking FrameNet to the Suggested Upper Merged Ontology. – 2006. – 9 p. – available at <http://adampease.org/Articulate/publications/FOIS2006.pdf>.
25. Chow I.C., Webster J.J. Integration of Linguas Resources for Verb Classification: FrameNet, WordNet and SUMO // Fifth Mexican International Conference on Artificial Intelligence (MICAI'06). – 2006. – P. 262–268. – available at <http://csdl2.computer.org/persagen/DLAbstoc.jsp?resourcePth=/dl/proceedings/micai/&toc=comp/proceedings/micai/2006/2722/00/272>.

Стаття надійшла до редакції 05.06.2007