



УДК 004.82

ОБЗОР И АНАЛИЗ МЕТОДОЛОГИЙ И МЕТОДОВ ПОСТРОЕНИЯ ОНТОЛОГИЙ

Хала Е.А.

** Международный научно-учебный центр информационных технологий и систем НАН и МОН
Украины, г. Киев, Украина
cecergreat@ukr.net*

В работе выполнен обзор и сравнение методик создания онтологических описаний, приведены результаты их анализа. Приводится общая характеристика и недостатки автоматических методов построения онтологий. Описаны основные языки представления онтологий и наиболее значимые существующие онтологические ресурсы.

Ключевые слова: онтология, онтологические модели, онтологические системы, языки представления онтологий.

Введение

Понятие онтологии, заимствованное из философии, в настоящее время активно применяется в искусственном интеллекте и информатике. Все больше интеллектуальных задач, решение которых связано с обработкой знаний, решаются с применением онтологий [Гаврилова и др., 2000], [Куршев, 2007], [Найханова, 2008], [Нариньяни и др., 2001], [Россеева и др., 2000] и др. По всей вероятности в недалекой перспективе онтологии будут использоваться при решении очень многих интеллектуальных задач.

Самым распространенным определением онтологии на данный период является определение T.R. Gruber [Gruber, 1993], [Gruber, 1995], согласно которому онтология является точной спецификацией концептуализации. С этой точки зрения для каждой из баз данных или баз знаний, или систем, основанных на знаниях, или агентов знаний должны быть построены спецификации, основанные на некоторой концептуализации. Множества объектов и отношений между ними должны быть описаны в некотором словаре, в котором система, основанная на знаниях, представляет свои знания. К средствам создания онтологий относятся системы Ontolingua [Farquhar, 1997], Protege [Musen, 1998], OntoEdit [Sure, 2002], OilEd [Bechhofer, 2001], Web-Deso [Naichanova, 2005], Ontogrid [Загоруйко и др., 2005] и другие. В работах [Нариньяни и др., 2002] и [Овдей и др., 2004] приведены сравнительные оценки перечисленных средств с точки зрения их внешней и внутренней организации.

1. Классификация онтологий и языки описания онтологий

В проектировании онтологий условно можно выделить два направления, до некоторого времени развивавшихся отдельно. Первое связано с представлением онтологии как формальной системы, основанной на математически точных аксиомах. Второе направление развивалось в рамках компьютерной лингвистики и когнитивной науки. Там онтология понималась, как система абстрактных понятий, существующих только в сознании человека, которая может быть выражена на естественном языке.

Таким образом, существует два подхода к созданию и исследованию онтологий. Первый (формальный) основан на логике. Второй (лингвистический) основан на изучении естественного языка (семантики) и построении онтологий на больших текстовых массивах, так называемых корпусах.

В настоящее время данные подходы тесно взаимодействуют, и идет активный поиск связей, позволяющих комбинировать соответствующие методы. Поэтому иногда бывает сложно отделить лексические онтологии с элементами формальных аксиоматик от логических систем с включениями лингвистических знаний. Независимо от различных подходов можно выделить три основных принципа классификации онтологий [Соловьев, 2006]:

- по степени формальности;
- по наполнению, содержанию;
- по цели создания.

Ключевым моментом в проектировании онтологий является выбор соответствующего языка спецификации онтологий (Ontology specification language). Цель таких языков - предоставить возможность указывать дополнительную машинно-интерпретируемую семантику ресурсов, сделать машинное представление данных более похожим на положение вещей в реальном мире.

На сегодняшний момент выделяют три основных класса языков описания онтологии [Соловьев, 2006]:

- традиционные языки спецификации онтологий: Ontolingua, CycL и языки, основанные на дескрипционной логике (такие как LOOM), также языки, основанные на фреймах (OKBC, OCML, Flogic);
- более поздние языки, основанные на Web - стандартах (XOL, SHOE, UPML);
- специальные языки для обмена онтологией через Web: RDF(S), DAML, OIL, OWL.

2. Подходы и методологии построения онтологий

Подходы к построению онтологий можно условно разделить на классический и современный.

Под классическим подходом понимается построение онтологий в соответствии со

стандартом онтологического анализа IDEF5 [ИСЕ, 1994]. Современный подход к онтологическому анализу подразумевает, в основном, построение web-онтологий – онтологий в контексте семантической паутины и сводится к разработкам консорциума The World Wide Web Consortium [Андреева и др., 2009]. Среди методологий и методов построения онтологий можно выделить следующие:

- метод Усколда и Кинга [Uschold et al, 1998];
- методология Грюнингера и Фокса [Gruninger et al, 1995];
- подход Cyc [Lenat et al, 1989];
- метод KACTUS [Schreibe et al, 1995];
- метод SENSUS [Swartout et al, 1997];
- методология On-To-Knowledge [Staab et al, 2001];
- методология METHONTOLOGY [Fernandez et al, 2006].
- и др. методологии приведенные в таблице 1.

Рассмотренные методики имеют ряд общих характеристик – выбор основных понятий, сущностей и связей между ними, а затем их кодирование посредством различных инструментов. Некоторые методологии содержат этап, предшествующий выборке понятий – определение специфики предметной области, целей и сфер применения построенной онтологии.

Methodologies	Type of development	Collaborative construction	Reusability support	Degree of application dependency	Life cycle recommendation	Strategies for identifying concepts	Methodology details	Interoperability support
TOVE	Stage based	No	Yes	Application semi independent	No	Middle out strategy	Some details	No
Enterprise model approach	Stage based	No	Yes	Application independent	No	Middle out strategy	Some details	No
METHONTOLOGY	Evolving prototype	No	Yes	Application independent	Yes	Middle out strategy	Sufficient details	No
KBSI IDEF5	Evolving prototype	No	Yes	Application independent	No	Not clear	Some details	No
Ontolingua	Modular development	Yes	Yes	Application independent	No	Not clear	Some details	Yes
Common KADS and KACTUS	Modular development	No	Yes	Application dependent	No	Top down strategy	Insufficient details	No
PLINIUS	Guidelines	No	No	Application independent	No	Bottom up strategy	Some details	No
ONIONS	Modular development / Guidelines	No	No	Application dependent	No	Not clear	Insufficient details	Yes
Mikrokosmos	Guidelines	No	No	Application dependent	No	Rule based strategy	Some details	No
MENELAS	Guidelines	No	No	Application dependent	No	Concepts Graphs (CG)	Insufficient details	No
SENSUS	does not mention any preference	Yes	Yes	Application semi independent	No	Bottom up	Some details	Yes
Cyc methodology	Evolving prototype	No	Yes	Application independent	No	Not clear	Some details	No
UPON	Evolving prototype	No	Yes	Application independent	Yes	Middle out strategy	Some details	No
101 method	Evolving prototype	No	Yes	Application independent	No	Developer's consent	Some details	No
On-To-Knowledge	Evolving prototype	No	No	Application dependent	Yes	Middle out strategy	Some details	No

Немаловажной частью методик являются алгоритмы и инструкции, по которым происходит выделение второстепенных сущностей и интеграции их в уже имеющуюся онтологию, а также метод построения основания будущей онтологии, используемой для последующего наращивания. Заключительным этапом является определение компетентности онтологии, проверка соответствия поставленным требованиям и решаемым задачам, а также сфер применения и использование построенной онтологии.

К сожалению ни одна из методологий не является полной, если сравнивать их со стандартом IEEE, а все предлагаемые решения не унифицированы. В настоящее время каждая группа применяет свою собственную методологию, что усугубляется недостаточной полнотой описания методологий. Поэтому, необходимы усилия по унификации методологий.

3. Методы автоматического построения онтологии

3.1. Представление онтологии в виде конечного автомата

Если представить онтологию в виде орграфа, где вершины – это множество предметных областей, а ребра – бинарное отношение между этими предметными областями, то с каждым таким орграфом будем ассоциировать конечный частичный детерминированный автомат без выходов [Крывый, 2008]. Представление онтологий в виде конечного автомата без выходов позволяет ввести операции на онтологиях. Операции на автоматах означают операции на регулярных языках, которые акцептируются этими автоматами [Крывый, 2008], [Бениаминов, 2003]. Причем множество операций (в случае надобности) можно расширять операциями на графах и алгеброй отношений. Поскольку каждая онтология является представлением некоторой совокупности отношений, то можно вводить операции реляционной алгебры.

3.2. Подход на основе лексико-синтаксических шаблонов

Данный подход был предложен в [Рабчевский, 2009] и относится к группе методов автоматического построения онтологий, использующих лингвистические средства. Сторонники подхода утверждают, что для построения онтологий следует активно использовать все уровни анализа естественного языка: морфологию, синтаксис и семантику. На основе лексико-синтаксических шаблонов выделяются онтологические конструкции [Анисимов, 2002], [Рабчевский, 2009].

Подход на основе лексико-синтаксических шаблонов не является специализированной на определенную предметную область и это его достоинством, однако лексико-синтаксические шаблоны как метод семантического анализа текстов

на естественном языке является не очень эффективным средством для автоматического построения онтологий.

3.3. Подход на основе системы продукций

Данный подход относится к группе методов автоматического построения онтологий, в основе которых лежат подходы из области искусственного интеллекта. В [Найханова, 2008] предлагается модель автоматического построения онтологий в виде системы продукций и применении генетического и автоматного программирования для создания необходимых моделей. Для создания методов автоматического построения онтологий автор разрабатывает модель генерации системы продукций (генетическое программирование), модель генерации преобразователей и модель генерации систем логического вывода (генетическое и автоматное программирование), модель аппарата активации продукций (автоматное программирование).

Применение продукционных правил обеспечивает следующие преимущества: простота и высокое быстродействие, модульность, удобство модификации, ясность, прозрачность, возможность постепенного наращивания. Среди недостатков можно выделить недостаточную семантическую связность между правилами.

3.4. Автоматическое построение онтологий по коллекции текстовых документов

В работе [Мозжерина, 2011] предлагается подход к решению проблемы автоматического построения онтологий, преимущественно основанный на статистических методах анализа текстов на естественном языке.

Для улучшения получаемой в результате работы системы онтологии, автор предлагает провести предварительную кластеризацию документов коллекции таким образом, чтобы в один кластер попадали тематически близкие документы, а дальнейшую работу проводить отдельно с каждым полученным кластером.

Также [Мозжерина, 2011] предполагает, что существующие статистические методы извлечения терминов из текстов на естественном языке могут показать лучшие результаты, если дополнить их определенными эвристиками, которые детальней описываются в работе автора.

Статистический подход является достаточно универсальными, но позволяет выделить только базовые отношения, необходимые для построения онтологии, что является его недостатком.

Заключение

Создание онтологий является перспективным направлением современных исследований по обработке информации, представляемой на естественном языке. В рамках работы выполнен

обзор методов онтологического инжиниринга и сформулирован обобщенный подход для построения онтологии, рассмотрены различные классификации онтологий. Уже сейчас существует ряд обширных онтологий, построенных как в рамках отдельных предметных областей, так и для незамкнутых областей знания. Наиболее перспективной является автоматизация создания онтологий, однако на данном этапе еще не разработаны эффективные процедуры, применение которых позволит сократить долю ошибок. Однако уже сейчас существует ряд приложений, успешно использующих онтологии в своей работе.

Библиографический список

- [Андреева, 2009] Андреева, Н.В. Выбор методов и средств онтологического анализа стандартов информатической безопасности / Н.В.Андреева, А.В.Любимов // VI Всероссийская межвузовская конференция молодых ученых. — СПб: СПбГУ ИТМО, 2009. Вып. 6. - С. 29-33.
- [Анисимов, 2002] Анисимов, А.В. Система обработки текстов на естественном языке / А.В.Анисимов, А.А.Марченко // Искусственный интеллект. — 2002. — № 4. — С. 157 — 163.
- [Бениаминов, 2003] Бениаминов, Е.М. Алгебраические методы в теории баз данных и представлении знаний / Е.М.Бениаминов. — М.: Научный мир, 2003 — 184 с.
- [Гаврилова, 2000] Гаврилова, Т.А. Базы знаний интеллектуальных систем / Т.А.Гаврилова, В.Ф.Хорошевский. — СПб: Питер, 2000. — 384 с.
- [Загоруйко, 2005] Загоруйко, Н.Г. и др. Система "Ontogrid" для построения онтологий / Н.Г.Загоруйко // Компьютерная лингвистика и интеллектуальные технологии. Тр. междунар. Конференции Диалог'2005. М., 2005. С. 146-152.
- [Крывый, 2008] Крывый, С.Л. Автоматное представление онтологий и операции на онтологиях / С.Л.Крывый, А.Н.Ходзинский — International Book Series. — N 1. — Algorithmic and Mathematical Foundations of the Artificial Intelligence. — ITHEA: Sofia. -2008. — PP. 173-179.
- [Куршев, 2007] Куршев, Е.П. Исследование методов извлечения информации из текстов с использованием автоматического обучения и реализация исследовательского прототипа системы извлечения информации / Е.П.Куршев, Д.А.Кормалев, Е.А.Сулейманова, И.В.Трофимов // Математические методы распознавания образов: 13-я Всерос. конф.: сборник докладов. — М.: МАКС Пресс, 2007. — С.602-605.
- [Мозжерина, 2011] Мозжерина, Е.С. Автоматическое построение онтологий по коллекции текстовых документов / Е.С.Мозжерина // Электронные библиотеки: Перспективные Методы и Технологии, Электронные коллекции — RCDL 2011 — Воронеж, 2011 — С. 293 — 298.
- [Найханова, 2008] Найханова, Л.В. Методы и модели автоматического построения онтологий на основе генетического и автоматного программирования / Л.В.Найханова. - Красноярск, 2008. - 36 с.
- [Нариньяни, 2001] Нариньяни, А.С. Кентавр по имени ТЕОН: тезаурус+онтология / А.С.Нариньяни // Междунар. семинар Диалог'2001 по компьютерной лингвистике и ее приложениям. — Аксаково, 2001. — Т. 1. — С. 184—188.
- [Нариньяни, 2002] Нариньяни, А.С. ТЕОН-2: от тезауруса к онтологии и обратно / А.С.Нариньяни // Компьютерная лингвистика и интеллектуальные технологии: междунар. семинар Диалог'2002 — М.: Наука, 2002. — Т. 1. — С. 307-313.
- [Овдей, 2004] Овдей, О.М. Обзор инструментов инженерии онтологий / О.М.Овдей, Г.Ю.Проскудина // Журнал ЭБ. 2004 — №4
- [Рабчевский, 2009] Рабчевский, Е.А. Автоматическое построение онтологий на основе лексико-синтаксических шаблонов для информационного поиска / Е.А.Рабчевский // Труды XI Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». — Петрозаводск, 2009.
- [Соловьев, 2006] Соловьев, В.Д. Онтологии и тезаурусы:

учеб. пособие / В.Д.Соловьев, Б.В.Добров, В.В.Иванов, Н.В.Лукашевич; Казанский гос. ун-т, МГУ им. М.В. Ломоносова Казань. — М.: Казань, 2006. — 157 с.

[Bechhofer, 2001] Bechhofer, S. OilEd: A Reason-able Ontology Editor for the Semantic Web / S.Bechhofer, I.Horrocks, C.Goble, R.Stevens // Joint German/Austrian conf. on Artificial Intelligence (KI'01). Lecture Notes in Artificial Intelligence LNAI 2174, Springer-Verlag, Berlin, pages.396-408, 2001.

[IICE, 1994] Information Integration for Concurrent Engineering (IICE). IDEF5 Method Report. — Knowledge Based Systems, Inc., 1408 University Drive East College Station, Texas, USA. — September 21. — 1994.

[Farquhar, 1997] Farquhar, A. The Ontolingua server: A tool for collaborative ontology construction / A.Farquhar, R.Fikes, J.Rice // International Journal of Human-Computer Studies, 46(6), pages 707—728, 1997.

[Fernandez, 2006] Fernandez, M. METHONTOLOGY: From Ontological Art Towards Ontological Engineering / M. Fernandez, A.Gomez-Perez, N.Juristo // AAAI Technical Report SS-97-06, 2006.

[Gruber, 1993] Gruber, T.R. A translation approach to portable ontologies / T.R.Gruber // Knowledge Acquisition. 1993. —No 5(2). — С. 199-220.

[Gruber, 1995] Gruber, T. Towards principles for the design of Ontologies used for knowledge sharing / T.Gruber // International Journal of Human-Computer Studies. —1995. —No 43(5/6). —С. 907-928

[Gruninger, 1995] Gruninger, M. Methodology for the design and evaluation of ontologies / M.Gruninger, M.S.Fox // Proceeding of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI. 1995.

[Lenat, 1989] Lenat D. Building Large Knowledge Based Systems: Representation and Inference in the Cyc Project / D.Lenat, R.V.Guha // Addison-Wesley, 1989.

[Musen, 1998] Musen, M. Domain Ontologies in Software Engineering: Use of Protege with the EON Architecture / M.Musen, // Methods of Inform. in Medicine, pages 540-550,1998.

[Naichanova, 2005] Naichanova, L.V. Knowledge presentation of inference system based on knowledge/ L.V.Naichanova // (IT@IES'2005): Materials of the international scientific conference / edited by A.N. Tikhonov (chair.) and others; SIIT@T Informika. — Moscow: VIZCOM, 2005. — P. 63—66.

[Uschold, 1998] Uschold, M. The Enterprise Ontology The Knowledge Engineering Review / M.Uschold, M.King, S.Moralee, Y.Zorgios // Vol. 13, Special Issue on Putting Ontologies to Use,1998.

[Schreibe, 1995] Schreibe, G. The KACTUS View on the 'O' World / G.Schreibe, B.Wielinga, W.Jansweijer // In Proceedings of the National Dutch AI Conference. NAIC'95. 1995.

[Staab, 2001] Staab, S. Knowledge processes and ontologies / S.Staab, H.P.Schnurr, R.Studer, Y.Sure // IEEE Intelligent Systems - 2001. - V. 16. - № 1. - P. 26-34.

[Sure, 2002] Sure, Y. OntoEdit: Collaborative ontology development for the Semantic Web / Y.Sure, M.Erdmann, J.Angelle, S.Staab, R.Studer, D.Wenke // In Proc. of the Inter. Semantic Web Conference (ISWC 2002), Sardinia, Italia, June 2002.

[Swartout, 1997] Swartout, B. Toward Distributed Use of Large-Scale Ontologies / B.Swartout, P.Ramesh, K.Knight, T.Russ // Symposium on Ontological Engineering of AAAI. Mars, 1997.

REVIEW AND ANALYSIS METHODOLOGIES AND TECHNIQUES OF ONTOLOGY BUILDING

Khala C.A.

International Research and Training Center of Information Technologies and Systems NAS and MES of Ukraine, Kiev, Ukraine

cecerongreat@ukr.net

The paper gives an overview and comparison of techniques for building ontological descriptions, the results of their analysis. Provides a general description and disadvantages of automatic methods for constructing ontologies. The basic ontology representation languages and the most significant existing ontological resources.