

Об использовании методов решения проблем для реализации активных, волюционирующих информационных ресурсов

Active evolving information resources and their integration into an information system are considered. The possibility to support such resources is based on the approach using ontologies of information resources and library of domain-independent problem-solving methods of modification ontology's elements.

Рассматриваются информационные ресурсы и их интегрирование в информационную систему. Возможность поддержать такие ресурсы основывается на подходе, использующем онтологию информационных ресурсов и библиотеку методов решения проблем и модификацию онтологии элементов.

Введение

Разработка информационных систем (ИС) путем интеграции и адаптации уже существующих информационных ресурсов (ИР) составляет альтернативу разработке ИС «с нуля», с точки зрения экономии времени и усилий по созданию нужных ИР. На сегодняшний день существует ряд действующих ИС, в рамках которых обеспечивается семантическая интеграция неоднородных баз данных (TSIMMIS[1], SIMS[2], HERMES[3], InfoSleuth[4], OBSERVER[5], OntoSeek[6] и др.), компонентов знаний (IBROW3[7] и др.). Все эти системы используют идеологию медиатора [8], т.е. посредника между ИР и пользователями ИС. Некоторые системы реализованы в виде мультиагентных сред (SIMS, InfoSleuth, TSIMMIS).

Однако любой ИР является отражением некоторой части реального мира, называемой предметной областью, и должен быть адекватен ей и с ней согласован. Это значит, что ИР должен изменяться или эволюционировать при изменениях в структуре или в функциях предметной области. Поэтому помимо задач, связанных непосредственно с достижением семантической интероперабельности в рамках ИС, возникают задачи отслеживания эволюции ИР и поддержания работы ИС в условиях эволюции ИР.

Будем понимать под *активностью ИР* наличие компоненты реагирования этого ИР на изменения, происходящие в его предметной области. Активность ИР должна, таким образом, обеспечивать адекватность и согласованность ИР предметной области, т.е. обеспечивать *эволюцию ИР* в соответствии с эволюцией предметной области. Следует отметить, что под *эволюцией* здесь и далее в статье понимается не переход ИР из одного допустимого состояния в другое

при неизменной структуре ресурса («концептуальной схеме», согласно [9]), а изменение самой структуры ИС (появление / исчезновение типов сущностей, атрибутов, изменение условий связности сущностей и др.).

Проблемы поддержания эволюции ИР уже исследовались в рамках объектных баз данных. В [10] изучены пути возможных модификаций схем объектных баз данных, работы [11], [12] предлагают механизмы поддержания целостности схем объектных баз данных при модификации элементов схем. Некоторые из механизмов модификации нашли воплощение в таких объектно-ориентированных СУБД, как O₂, Gemstone, ORION. Более подробную информацию об этих системах и о подходах, применяемых к модификациям схем объектных баз данных, можно получить в [13]. Подобные механизмы можно найти и для других типов ИР. Например, для реляционных схем – поддержка системных таблиц в практически всех коммерческих реляционных СУБД и инструментарий для работы с этими таблицами (см. [14]). Таким образом, для отдельных ИР задача поддержания эволюции уже исследовалась.

Для ИС, интегрирующих неоднородные эволюционирующие ИР, задача поддержания эволюции как всей системы, так и отдельных ИР в рамках системы требует своего исследования. Предлагается подход к поддержанию эволюционирующих ИР в рамках системы семантической интеграции неоднородных эволюционирующих ресурсов. Подход базируется на использовании двух типов онтологий предметных областей: *общая онтология* всей ИС и онтологии отдельных ИР (*локальные онтологии*). Методы модификации представляются в библиотеке домен-независимых методов решения проблем. Библиотека методов решения проблем помимо самих методов модификации элементов онтологий должна иметь собственную *онтологию класса задач* (task ontology), т.е. описание терминов, в которых заданы сами методы модификации. Наличие такой библиотеки позволит использовать ее в произвольных системах семантической интеграции неоднородных эволюционирующих ИР, основанных на использовании онтологий.

Статья организована следующим образом: в разделе 1 – конкретизируется использование онтологий для решения поставленной задачи; в разделе 2 – дается анализ возможных модификаций элементов онтологий; в разделе 3 – рассматриваются методы решения проблем для модификации онтологий; раздел 4 – содержит примеры архитектур взаимодействия эволюционирующих ИР с ИС. В заключении содержатся выводы и открытые для исследования вопросы.

1. Онтологии информационных ресурсов

Термин «онтология», употребляемый в области искусственного интеллекта, имеет следующий смысл [15]:

«Онтология – некоторая логическая теория, отвечающая за отражение подразумеваемого смысла терминов некоторого формального словаря».

На сегодня существует несколько типов онтологий. Различают [15] онтологии предметных областей (domain ontology), онтологии классов задач (task ontology), онтологии приложений (application ontology), онтологии верхнего уровня (top-level ontology). К отдельному типу также относят онтологии методов решения проблем [16].

Онтология зависит как от языка своего описания, так и от того, к какому типу она относится. Так, например, известны предложения по формализации онтологии произвольной предметной области на основании теории множеств NBG [9]; для онтологий верхнего уровня есть частичная формализация базовых концепций (таких, как субстанции и события) в рамках мереологии и топологии [17]. В качестве формальных языков описания онтологий используется алетическая модальная логика [18], [9], деонтическая логика [19], а также классическое исчисление предикатов 1-го порядка, многосортная логика.

Известные промышленные языки описания онтологий также основываются на многосортном исчислении предикатов 1-го порядка. К ним относятся: KIF [20], CycL[21], лексические концептуальные графы [22], СИНТЕЗ [23], LDL [24], ODL I3 [8], другие. Подробную информацию о промышленных языках можно найти в [25].

Рассмотрим теперь структуру произвольной онтологии ИР (т.е. онтологию предметной области этого ресурса). Онтология ИР является формальной теорией 1-го порядка. Унарные предикаты соответствуют терминам формального словаря ИР (концепциям), бинарные предикаты соответствуют отношениям между концепциями. В самом простом случае онтология описывает иерархию концепций, связанных отношением «подтип-супертип». В более сложных случаях, помимо определений самих концепций в онтологию добавляются аксиомы, выражающие другие семантические связи между концепциями (например, меронимные, отношения следования и др. [26]), а также аксиомы, ограничивающие интерпретацию концепций и связей.

Анализ систем семантической интеграции неоднородных ИР ([2],[4],[5]) показал, что наиболее привлекательным подходом является наличие *общей* онтологии ИС и *локальных* онтологий ИР. Такая идеология получает естественное выражение в мультиагентной архитектуре ИС. *Локальная онтология ИР* создается поставщиком ресурса либо генерируется агентом ресурса при подключении ресурса в ИС, на основании общей онтологии ИС. Эта онтология выражается в терминах ИР. Затем агент ресурса строит модель соответствия между локальной онтологией в терминах ИР и общей онтологией системы. Назовем такую модель *канонической онтологией ИР*, т.е. каноническая онтология ИР – локальная онтология ИР, выраженная в терминах общей онтологии ИС. Каноническая онтология может либо присутствовать явно (как в SIMS, OBSERVER – в виде модели ресурсов), либо быть доступной опосредовано, через агента ресурса (в InfoSleuth – информация о соответствии встроена в структуру каждого агента ресурса). Будем предполагать, что структура канонической онтологии такая же, как и у любой онтологии ИР – словарь концепций вместе с их формальными определениями, аксиомы, задающие семантические связи между концепциями и ограничения

на возможные интерпретации концепций, т.е. каноническая онтология каждого ИР выражена в явном виде.

2. Модификация онтологий

Проанализируем возможные пути модификации онтологий, не касаясь вопросов, связанных с языком описания онтологий. Рассмотрим онтологию некоторой предметной области, построенную на основании методики, предложенной в [9]. Пусть эта онтология является простой иерархией классов (*E*-классов), обладающих наборами атрибутов. Между классами могут существовать связи по атрибутам, когда множество значений атрибута принадлежит экстенсionалу некоторого другого *E*-класса. Модификации могут подвергаться атрибуты классов, структура класса, граф наследования классов, связи по атрибутам.

Классификация модификаций такой онтологии может быть адаптирована из [13]:

1. Модификация атрибутов класса
 - а) создание атрибута;
 - б) удаление атрибута;
 - в) переименование атрибута;
 - г) изменение домена атрибута;
 - д) изменение категории атрибута (обязательность, множественность и т.д.)
2. Модификация класса
 - а) создание класса;
 - б) удаление класса;
 - в) переименование класса;
3. Модификация графа наследования класса
 - а) создание нового отношения «подкласс-суперкласс» между классами;
 - б) удаление отношения «подкласс-суперкласс» между соседними классами;

Однако поскольку онтология является логической теорией, то наряду с описанием структуры и семантики понятий, выражаемых концепциями и семантическими связями, она может содержать и аксиомы, накладывающие некоторые ограничения. Таким образом, к указанной выше классификации следует добавить такой раздел:

4. Модификация прикладных аксиом онтологии
 - а) добавление прикладной аксиомы;
 - б) удаление прикладной аксиомы.

Любые модификации онтологии должны оставлять онтологию согласованной, т.е. должна выполняться непротиворечивость онтологии и в исходном состоянии, и в модифицированном. Очевидно также, что модифицированная каноническая онтология должна быть согласована и с самим ИР. Следовательно, для поддержания согласованности онтологии ИР с самой предметной областью этого ИР и непротиворечивости самой онтологии, необходимо ввести некоторый набор метаправил – инвариантов онтологий. Эти пра-

вила, сформулированные на метауровне по отношению к языку описания онтологий, позволят поддерживать непротиворечивую модификацию онтологий.

Пример набора инвариантов, касающихся изменения схемы объектной базы данных, дан в [13]:

- i1: инвариант иерархии классов – иерархия классов есть связный ориентированный ациклический граф с одним корнем. Каждый класс в иерархии имеет уникальное имя.
- i2: инвариант уникальности имен – все атрибуты класса, наследованные или собственные, имеют разные имена.
- i3: инвариант единственности происхождения – все атрибуты класса имеют единственный (различный для каждого) прототип.
- i4: инвариант полного наследования – класс наследует все атрибуты из всех своих суперклассов, если только это не нарушает требований i2 и i3.
- i5: инвариант совместимости доменов – если атрибут A появляется и в классе X, и в классе Y, причем X – подкласс Y, тогда домен A в X должен быть либо равен домену A в Y, либо быть его поддоменом.

Подобный набор инвариантов можно разработать и для модификации онтологий. Особый интерес вызывает вопрос наличия инвариантов модификации прикладных аксиом онтологии ИР. В качестве предварительного рассуждения отметим, что в рамках прикладных формальных теорий 1-го порядка можно потребовать непротиворечивости системы прикладных аксиом в любом состоянии онтологии ИР.

3. Методы решения проблем в задаче модификации онтологий информационных ресурсов

В разделе 2 было отмечено, что любая онтология, независимо от языка реализации и предметной области ИР, может быть подвержена ограниченному числу видов модификации. Это позволяет применить подход к описанию алгоритмов модификации онтологий в виде библиотеки методов решения проблем. Методы решения проблем (МРП)[16] – домен-независимые компоненты рассуждения, которые описывают образцы поведения и могут быть использованы для разных областей. МРП предоставляют уже систематизированные знания по решению целых классов задач. Стандарты архитектуры и языка описания МРП сегодня разрабатываются в рамках проектов IBROW3[7], НРКВ [27]. Так, в рамках класса задач предлагалась архитектура Task-Method-Domain-Application (TMDA[28]). Усовершенствованием TMDA является архитектура PPA (Problems, Paradigms, Assumptions)[16], которая позволяет уточнять МРП под конкретный класс задач, создавая для него адаптер.

Используем подход МРП к поставленной задаче.

Для выполнения инвариантов онтологии необходимо сформулировать правила разрешения конфликтов, связанные с модификациями элементов онтологий. Эти правила могут быть реализованы внутри агента ИР на про-

граммном уровне. Тогда, хотя канонические онтологии ИР при этом будут выражены унифицировано, необходимо будет разрабатывать различных агентов для разных типов ИР. Если же реализовать разработанные правила в виде домен-независимых МРП модификации элементов онтологии ИР, то набор таких методов сформирует библиотеку, которую можно будет использовать для разных приложений и предметных областей. Библиотека таких МРП, помимо самих методов модификации онтологий, должна иметь собственную онтологию класса задач, т.е. описание терминов, в которых заданы сами методы. Тогда такая библиотека будет независимым компонентом ИС интеграции неоднородных эволюционирующих ресурсов и позволит реализовывать активность произвольного ИР в рамках ИС.

4. Механизмы поддержания эволюции информационных ресурсов в рамках информационной системы

Рассмотрим теперь возможные архитектуры ИС, интегрирующей неоднородные эволюционирующие ИР с точки зрения взаимодействия агента ресурса с общей и локальными онтологиями ИР.

Самой простой архитектурой является такая, в которой отсутствуют в явном виде канонические онтологии ИР и нет механизмов поддержания эволюции ИР. Тогда при любой модификации локальной онтологии любого ИР происходит перерегистрация ИР в ИС. Достоинством такой архитектуры является простота реализации. Недостаток: при большом количестве ИР с объемными локальными онтологиями любая модификация элемента онтологии влечет за собой перестройку общей онтологии ИС. Пусть теперь для любого ИР есть явная каноническая онтология, но механизмы ее модификации отсутствуют. Тогда при модификации элемента локальной онтологии агент ИР перерегистрирует только каноническую онтологию, перестройка общей онтологии может потребоваться только в случаях глобального изменения онтологии ИР. Наконец, пусть для любого ИР есть явная каноническая онтология и есть явно сформулированные механизмы поддержания эволюции ИР в виде библиотеки МРП модификации элементов онтологий. Тогда, при модификации элемента локальной онтологии агент ИР модифицирует соответствующие элементы канонической онтологии и, возможно, генерирует сообщение об изменениях агенту общей онтологии. Перерегистрация канонической онтологии требуется только в случаях серьезных изменений в локальной онтологии ИР.

Таким образом, последний подход с использованием явных канонических онтологий и явной библиотеки МРП модификации онтологий является альтернативой двум первым, поскольку предлагает унифицированный подход к модификациям как канонических онтологий ИР, так и общей онтологии.

Заключение

В статье поднимается вопрос реализации информационных систем, интегрирующих на семантическом уровне неоднородные эволюционирующие информационные ресурсы. При этом под эволюцией ресурса понимается изменение структуры (онтологии) этого ресурса с течением времени в связи с изменениями в структуре его предметной области и классов задач, для которых предназначался этот ресурс. Для решения этой задачи предлагается подход явного использования двух типов онтологий предметных областей – общей онтологии информационной системы и канонических онтологий отдельных информационных ресурсов. Для поддержки эволюции информационных ресурсов предлагается разработать библиотеку домен-независимых методов модификации онтологий в виде библиотеки методов решения проблем. Помимо описания самих методов, предлагается разработать онтологию класса задач модификации онтологий, содержащую всю терминологию, в которой описаны методы модификации.

Остаются открытыми такие вопросы:

1. Разработка и формализация правил модификации онтологий в виде библиотеки методов решения проблем.
2. Структура агента информационного ресурса, поддерживающего эволюцию ресурса.
3. Выяснения критериев «терпимости» агента информационного ресурса к эволюции локальной онтологии ресурса.
4. Архитектура информационной системы, интегрирующей на семантическом уровне неоднородные эволюционирующие ресурсы.

Литература

1. H. Garcia-Molino et. al. The TSIMMIS Approach to Mediation: Data Models and Languages. In Proceedings of the NGITS (Next Generation Information Technologies and Systems), June 1995.
2. Y. Arens, C.A. Knoblock, W. Shen. Query Reformulation for Dynamic Information Integration. Journal of Intelligent Information Systems, 1996.
3. S. Adali, V.S. Subrahmanian. Amalgamating knowledge bases, II. Distributed mediators. International Journal of Intelligent and Cooperative Information Systems 3(4): 349-383, 1994.
4. R.J. Bayardo et al. InfoSleuth: Semantic Integration of Information in Open and Dynamic Environment. In Proceedings of the 1997 ACM International Conference on the Management of Data (SIGMOD), Tucson, Arisona, May 1997.
5. E. Mena, V. Kashyap, A. Sheth, A. Illaramendi. OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-Existing Ontologies. In Proceedings of the First IFCIS International Conference on Cooperative Information Systems (CoopIS'96), Brussels (Belgium), June. IEEE Computer Society Press, 1996.
6. N. Guarino, C. Masolo, G. Vetere. Content-Based Access to the Web. IEEE Intelligent Systems, May/June 1999, p.70-80.
7. V.R. Benjamins et al. IBROW3: An Intelligent Brokering Service for Knowledge-Component Reuse on the World – Wide Web. Proceedings of the 11th Workshop on Knowledge Acquisition, Modeling and Management, KAW'98.

8. P. Buneman, L. Raschid, J.Ullman. Mediator Languages – a Proposal for a Standard. Report of an F/POB working group held at the University of Maryland, April 1996. <ftp://ftp.umiacs.umd.edu/pub/ONRrept/medmodel96.ps>
9. Андон Ф.И., Яшунин А.Е., Резниченко В.А.. Логические модели интеллектуальных информационных систем. –Киев: Наукова думка.–1999. – С. 396.
10. J. Banerjee, W. Kim, H.-J. Kim, H.F. Korth. Semantics and Implementation of Scheme Evolution in Object-Oriented Databases. In: U. Dayal, I. Traiger (eds.), Proceedings of ACM SIGMOD International Conference on Management of Data, pp.311-322, San Francisco, May 1987.
11. R. Zicari. A Framework for Schema Updates in an Object Oriented Database System. In: Proceedings of the International Conference on Data Engineering, pp.2-13, April 1991.
12. C. Delcourt, R. Zicari. The Design of an Integrity Consistency Checker (ICC) for an Object Oriented Database System. In: P. America (ed.), Proceedings European Conference on Object-Oriented Programming., pp.97-117, Geneva, Switzerland, July 1991. Springer Verlag, New York, LNCS 512.
13. C. Zaniolo, S. Ceri, C. Faloutsos, R.T. Snodgrass, V.S. Subrahmanian, R. Zicari. Advanced Database Systems. Morgan Kaufmann Publishers, San Francisco, 1997.
14. V.A. Ermolayev. Object Oriented Dynamic Data Modeling and Active Data Dictionaries: Some Crosspoints. //Вісник Запорізького державного університету. Запоріжжя, 1998, №2. – С.53-63
15. N. Guarino. Formal Ontologies and Information Systems. In: N. Guarino (eds.), proceedings of the First International Conference on Formal Ontologies in Information Systems (FOIS'98), pp.3-15, Trento, Italy 1998.
16. D. Fensel, E. Motta. Structured Development of Problem Solving Methods. In: Proceedings of the 11th Banff Knowledge Acquisition for Knowledge Based System Workshop (KAW'98), Banff, Canada, April 1998.
17. B. Smith. Basic Concepts of Formal Ontology. In: N. Guarino (eds.), Proceedings of the First International Conference on Formal Ontologies in Information Systems (FOIS'98), pp.19-28, Trento, Italy 1998.
18. N. Guarino, M. Carrara, P. Giaretta. An Ontology of Meta-Level Categories. In: Proceedings of the 4th International Conference of Knowledge Representation and Reasoning (KR'94), Morgan Kaufmann, San Mateo, CA, 1994.
19. P. Johanneson, P. Wohed. Deontic Specification Patterns. In: N. Guarino (eds.), Proceedings of the First International Conference on Formal Ontologies in Information Systems (FOIS'98), pp.95-107, Trento, Italy 1998.
20. M.R. Genesereth, R.E. Fikes et.al. Knowledge Interchange Format Version 3.0 Reference Manual. Logic-92-1, Stanford University Logic Group, 1992.
21. D. Lenat et al. CYC: Toward programs with Common Sense, Communications of the ACM, Vol.33, No.8, August 1990, p. 30-49.
22. J. Sowa. Conceptual Structures: Information processing in Mind and Machine. Addison-Wesley, Reading, Mass.,1984.
23. Калиниченко Л.А. СИНТЕЗ: язык определения, проектирования и программирования интероперабельных сред интеграции неоднородных информационных ресурсов (вторая редакция). – Москва: ИПИ РАН, 1993.
24. C. Zaniolo. The Logical Data Language (LDL): An Integrated Approach to Logic and Databases. MCC Tech.Rep. STP-LD-328-91, 1991.
25. M. Uschold, M. Gruninger. Ontologies: Principles, Methods and Applications. Knowledge Engineering Review, 11(2), 1996.
26. V.C. Storey, «Understanding Semantic Relationships», Very Large Data Bases (VLDB) Journal, Vol.2, No.4, October 1993. – P. 455-488.
27. High Performance Knowledge Bases. <http://www.tekknowledge.com:80/HPKB>
28. E. Motta, Z. Zdrahal. An approach to the organization of a library of problem solving methods which integrates the search paradigm with task and method ontologies. International Journal of Human-Computer Studies, 1997.