Организация информационного взаимодействия разнородных автоматизированных систем: онтологический подход

Organizing the Information Interaction between Heterogeneous Automated Systems: an Ontological Approach

Саенко / Saenko I.
Игорь Борисович
(ibsaen@mail.ru)
доктор технических наук, профессор.
ФГКВОУ ВО «Военная академия связи имени Маршала Советского Союза С. М. Буденного» МО РФ (ВАС им. С. М. Буденного),
профессор.
г. Санкт-Петербург

Брунилин / Brunilin A.
Алексей Александрович
(ale483@mail.ru)
ВАС им. С. М. Буденного,
адъюнкт.
г. Санкт-Петербург

Ефимов / Efimov V.
Вячеслав Викторович
(vve@loniis.ru)
кандидат технических наук, доцент.
Филиал ФГУП «Ленинградское отделение центрального научно-исследовательского института связи» (ЛО ЦНИИС),
и. о. директора.
г. Санкт-Петербург

Ясинский / Yasinsky S.
Сергей Александрович
(yasinsky777@mail.ru)
доктор технических наук, доцент.
Филиал ФГУП ЛО ЦНИИС,
научный консультант.
г. Санкт-Петербург

Ключевые слова: информационное взаимодействие – information interaction; автоматизированная система – automated system; онтология – ontology; логический вывод – logical conclusion; семантическая модель – semantic model.

Рассматривается новый подход к организации информационного взаимодействия разнородных автоматизированных систем, основанный на использовании онтологий. Приводится формальное описание онтологий. Обсуждаются результаты применения онтологического подхода к решению задачи интеграции разнородных баз данных. Предлагается гибридный способ взаимодействия онтологий.

A new approach to organizing information interaction of heterogeneous automated systems based on using ontologies is considered. A formal description of ontologies is given. Results of applying the ontological approach to solve the task of integrating heterogeneous databases are discussed. A hybrid technique for interaction of ontologies is suggested.

Введение

На современном уровне развития информационных технологий и широкого внедрения в различные отрасли экономики (включая критически важные инфраструктуры [1]) автоматизированных систем (АС) различного назначения остро встает вопрос обеспечения эффективного информационного взаимодействия последних.

Стремительно возрастающий объем информации, требуемой для выработки управленческих решений, ее разнородность, повышение наукоемкости методов аналитической обработки исходных данных и другие объективные факторы неизбежно ведут к разделению и специализации управления. Вместе с тем решение общих управленческих задач, стоящих перед предприятиями и организациями, остается результатом комплексного взаимодействия разнородных АС, основным видом которого является информационное взаимодействие.

Целью настоящей статьи является рассмотрение одного из наиболее перспективных направлений решения проблемы информационного взаимодействия разнородных АС, связанного с созданием и использований онтологий.

Проблема информационного взаимодействия разнородных автоматизированных систем

Внедрение в корпоративные системы управления множества разнородных узкопрофильных АС с последующей их интеграцией является более дешевой и простой альтернативой созданию крупномасштабных комплексных автоматизированных моносистем. Тем не менее менеджменту организации при выработке управляющих воздействий необходимо опираться на результаты комплексного анализа разнородной информации. Таким образом, при несомненных преимуществах идеи дробления АС по функциональному предназначению проблема организации их информационного взаимодействия становится все более очевидной.

Данный вопрос имеет большое практическое значение еще и в силу того, что с развитием информационных технологий в АС внедряются все новые и новые методы обработки данных. Реляционные базы данных, которые до недавнего времени считались наиболее прагматичными средствами хранения информации, уже нельзя более считать таковыми. Так, все большее распространение приобретают XML- и RDF-базы данных, а также комплексные хранилища, сочетающие все перечисленные типы баз данных [2].

Проблема информационного взаимодействия разнородных АС обоснована также тем, что зачастую АС дублируют функции друг друга, используя в ряде случаев одни и те же исходные данные и вырабатывая идентичные результаты. При этом применяются различные механизмы обработки информации, ее классификации и кодирования, вырабатываются результирующие данные различных, зачастую уникальных и несовместимых форматов. Обработка информации в некоторых АС ведется с использованием неоптимальных, устаревших алгоритмов. Регламент ввода исходных данных не поддерживается, что приводит к потере адекватности информации в разнородных АС.

Решение указанной проблемы видится в создании универсальной информационно-аналитической структуры, программная платформа которой предоставляла бы пользователю возможность доступа к разнородной первичной информации, а также к результатам ее предварительного компьютерного анализа, безотносительно к ее природе и способам обработки. Это позволяет предложить новый подход к организации информационного взаимодействия разнородных АС, основанный на использовании особой разновидности баз знаний, называемой онтологией.

Сущность онтологического подхода

Под онтологией понимается структура данных, содержащая классы объектов предметной области, их связи и имеющиеся между ними правила, которая обеспечивает формирование и обработку запросов на основе некоторого механизма логического вывода (как правило, основанного на дескрипционных логиках) [3].

Известны различные формальные описания для онтологий. Наиболее полное описание дано в работе [4], в которой онтология представляется в виде кортежа $O = <C, P, A, H^C, prop, att>$, где C и P – непересекающиеся наборы концептов и идентификаторов связей, соответственно; H^C – иерархия концептов, называемая также таксономией концептов; $prop$ – функция вида $prop: P \rightarrow C \times C$, которая отражает неиерархические связи между концептами; A – множество атрибутов; att – функция вида $att: A \rightarrow C$, связывающая атрибуты с концептами.

Более простой формализм предлагается в работе [5], в которой под онтологией понимается тройка вида $O = <C, R, F>$, где C – конечное множество концептов; R – конечное множество отношений между концептами; F – конечное множество функций интерпретации, заданных на концептах и/или отношениях.

Суть онтологического подхода заключается в создании универсальных, кроссплатформенных информационных структур, использующих семантические метаописания данных как собственных, так и надстроек к уже существующим разнородным информационным массивам, в том числе к базам данных различных архитектур.

В качестве базовой информационной единицы онтологии выступает так называемый «*RDF*-триплет» (от англ. *Resource Description Framework* – «среда описания ресурсов») [6]. Под триплетом при этом понимается набор информационных сущностей «объект-предикат-субъект». Предметная область, описанная таким способом, обладает компьютерной семантикой, т.е. возможностью устанавливать и обрабатывать смысловые отношения между понятиями с помощью вычислительных машин. Компьютерная семантика раскрывается математическим аппаратом дескрипционных логик [3]. Достоинствами дескрипционных логик являются достаточная выразительность и высокая сходимость вычислительных алгоритмов (обычно – полиномиальная). Этим обеспечивается возможность формирования и обработки запросов к онтологиям на основе программного механизма логического вывода.

Структурно онтология состоит из двух частей. Первой является набор понятий предметных областей, называемый «*A-Box*» и являющийся аналогом таблиц данных в базах данных. Вторая часть называется «*T-Box*» и представляет собой описание взаимосвязей между понятиями. Именно *T-Box* содержит логику отношений объектов предметной области. Кроме того, онтология обладает несвойственным для баз данных элементом искусственного интеллекта, называемым механизмом логического вывода, который позволяет получать новые знания при анализе существующих понятий и логики их отношений [7]. Это позволяет автоматизировать процедуры семантического (смыслового) анализа имеющейся информации, что дает вычислительной машине возможность выработки предвари-

тельных рекомендаций на формирование управляющих воздействий.

О перспективности и практической реализуемости онтологического подхода для решения проблемы информационного взаимодействия разнородных АС свидетельствует также тот факт, что на сегодняшний день стандартизован ряд языков разработки онтологий – *OWL (Web Ontology Language)* и *SWRL (Semantic Web Rule Language)* [8, 9], а также спецификации языков отображения *DM (Direct Mapping)* и *R2RML (Relational DataBase to Resource Description Framework Mapping Language)* [10].

Реализация онтологического подхода для организации информационного взаимодействия

Возможности реализации онтологического подхода для организации информационного взаимодействия были исследованы на примере задачи интеграции баз данных, содержащихся в разнородных АС. Интеграции подвергались базы данных, удаленные друг от друга и работающие под управлением разных СУБД (использовались СУБД MySQL под ОС MS Windows-7 и СУБД Линтер-ВС под ОС МСВС). Несмотря на то, что создание запросов к разным реляционным базам данных само по себе не является технически сложной задачей, практическая реализация механизма их интеграции посредством отображения в онтологические конструкции служит ярким подтверждением жизнеспособности онтологического подхода. Кроме того, применение онтологий открывает перспективы интеграции не только реляционных, но и нереляционных баз данных.

Обычно для интеграции разнородных баз данных используется метод прямого связывания, при котором из внешней оболочки осуществляется доступ к обеим базам данных (например, через сервисы *ODBC, JDBC* и им подобные) с использованием *SQL*-конструкций. Однако здесь имеются существенные трудности. Оказывается, что даже при наличии программного доступа к базам данных использовать часть языковых конструкций языка *SQL* в привычном виде не представляется возможным. Причиной этого является тот факт, что разработчики СУБД используют в качестве средства манипуляции данными *SQL*-диалекты – упрощенные версии языка *SQL*. Это ограничение лишает конечного разработчика возможности в полной мере использовать выразительность и мощь классического языка *SQL* при создании внешних запросов к базам данных. По этой причине в ряде случаев для получения желаемых результатов приходится пользоваться приемом дробления исходных запросов на более простые, производя впоследствии дополнительные действия по агрегированию и анализу полученных промежуточных результатов. Это приводит к увеличению суммарного времени выполнения запросов, а также к усложнению самого процесса разработки интегрированной АС.

В отличие от описанного выше метода прямого связывания, объединение реляционных баз данных посредством онтологий не вызывает существенных проблем и осуществляется в два этапа. Вначале происходит накопление *RDF*-триплетов путем обработки массивов данных, полученных из различных источников, с помощью языков отображения *DM* или *R2RML*. Затем выполняется объединение полученных массивов за счет обычной вставки между соответствующими информационными сущностями связей вида «`rdfs:SubClassOf`» или «`owl:EquivalentClass`».

Онтологический подход в применении к интеграции разнородных баз данных имеет свои достоинства и недостатки. Несомненным достоинством семантического подхода является возможность использования онтологии как инструмента интеграции разнородных АС. Это означает, что имеется возможность надстраивать онтологию над уже разработанными информационными комплексами. В результате по сравнению с затратами на создание единой корпоративной АСУ отрасли, онтологический подход позволяет в значительной степени удешевить процесс разработки программно-информационных систем поддержки принятия решений.

В качестве недостатков онтологического подхода следует отметить более низкую производительность вычислений над онтологиями по сравнению с вычислениями в реляционных СУБД. Также необходимо иметь в виду, что ограничения на установление связей в онтологиях существенно более мягкие, чем в реляционных базах данных. Если объединение онтологий произведено в условиях недостаточно полного анализа их структуры, то тогда увеличивается вероятность ошибочных связей. Это обстоятельство также способно привести к нежелательным вычислительным издержкам, что может выражаться в нарушении семантической связности понятий и классов и снижении степени адекватности онтологической модели предметной области.

Для устранения этих недостатков предлагается следующий гибридный способ взаимодействия онтологий.

Пусть имеется m различных АСУ, которым соответствуют m онтологических моделей описания их предметных областей. При достаточно большом количестве связываемых систем необходимо избегать создания внутри каждой частной онтологии описаний для остальных взаимодействующих систем. В пределе их количество для каждой частной онтологии равно $(m-1)$. Если это допустить, то в результате получится $m\cdot(m-1)$ описаний взаимодействия, что будет нерациональным с позиций разработки и использования вычислительных ресурсов в ходе эксплуатации. Для того чтобы избежать подобной ситуации, предлагается создать $(m+1)$-ю онтологическую модель, которая будет объединять частные модели и описывать в терминах дескриптивной логики механизмы логического вывода (взаимодействия АС). Для объединения онтологий в данном случае используется способ включения, представляющий собой указание URI соответствующей

онтологии в качестве параметра «rdf:resource» тэга <owl:>, например:

```
<owl:Ontology rdf:about="АСУ-1">
  ...
 <owl:imports ="http://www.ASU-1.ru/root"/>
  ...
</owl:Ontology>
```

Очевидно, что подобный способ не ведет к неоправданному дублированию информации, позволяя разработчикам структурировать области описания предметных областей и механизмов логического вывода, сегментируя вычислительную нагрузку при эксплуатации системы в целом.

Для практической реализации предложенного подхода необходимо построить онтологии предметных областей интегрируемых АС, ориентируясь на решаемые в них практические управленческие задачи. Рассматривая информационные возможности каждой из интегрируемых АС, необходимо установить, что кроме решения определенного круга задач по предназначению, каждая система обладает определенным потенциалом для своего развития. Это дает возможность классифицировать программно-информационные средства, имеющиеся в распоряжении пользователя, в соответствии со следующими признаками, касающимися их программной реализации:

- уровень задач;
- уровень сервисов;
- уровень данных.

Задачей T_i предлагается считать законченный программный блок АС, позволяющий получить в качестве ответа на запрос (оператора или системы) формализованный документ установленного образца.

Сервисом S_j будем считать набор информационных услуг, а также наборы данных, формируемые АС на основании запросов пользователя или системы, если эти запросы не определены заранее руководящими документами в качестве стандартных задач, или если выдаваемый ответ представлен в форме, отличной от установленной.

Представителями уровня данных D_k будем считать исходные данные для решения задач в АС (значения полей классификаторов, цифровые изображения, наборы координат и прочее).

Таким образом, можно формально определить:

– *множество задач* T_M (M – количество АС), где для m-ой АС имеется $i(m)$ реализованных в ней задач. Множество задач определяется как $T_{m.I(m)}$, где $I(m) = \{\{1, ..., i(1)\}; ...; \{1, ..., i(m)\}\}$, $m = 1, ..., M$;

– *множество сервисов* S_M, где для m-ой АС будет описано $j(m)$ предоставляемых системой сервисов, определяемое как $S_{m.J(m)}$, где $J(m) = \{\{1, ..., j(1)\}; ...; \{1, ..., j(m)\}\}$, $m = 1, ..., M$;

– *множество данных* D_M, где для m-ой АС будет описано $k(m)$ обрабатываемых системой элементарных видов данных, определяемое как $D_{m.K(m)}$, где $K(m) = \{\{1, ..., k(1)\}; ...; \{1, ..., k(m)\}\}$, $m = 1, ..., M$.

Для формирования семантической модели предметной области КССА необходимо ввести метаописания сущностей уровней программной реализации, а также ввести понятие отношений семантической модели.

Отношения объектов RM семантической модели служат в онтологиях для формирования базиса логического вывода. Логический вывод позволяет программному ядру онтологии, или ризонеру (от англ. *reasoner*), производить автоматизированные рассуждения. Объекты онтологии, имеющие отношения между собой, количество отношений между объектами, свойства отношений (например, «являетсяДочерним», «поставляетсяС», «имеетШифр», «граничитСАкваторией» и

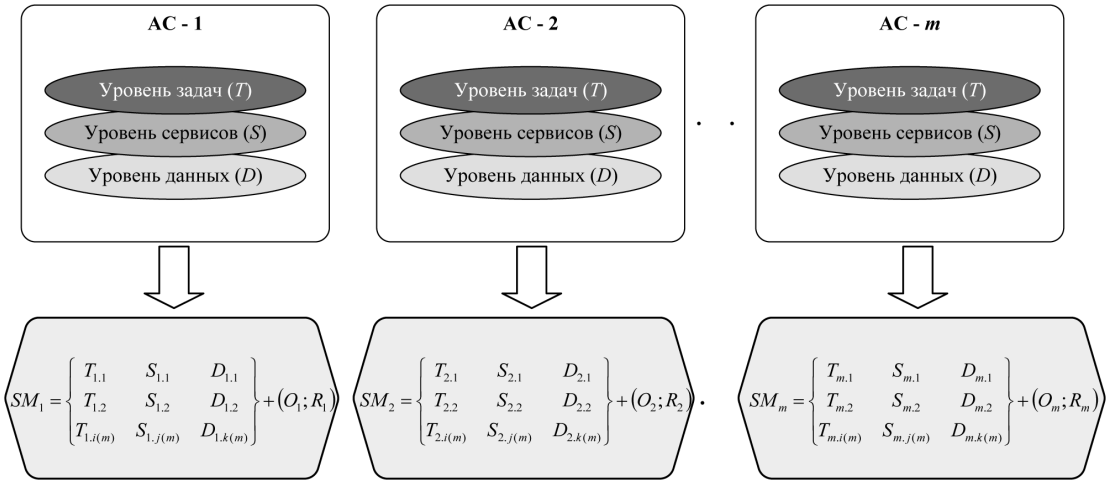

Рис. 1. Семантическая модель взаимодействия разнородных АС

т.д.) и особенности («кардинальность», «количество», «исключительность», «разделенность» и т.д.) определяются в процессе моделирования предметной области исходя из наличия смысловых связей между объектами в контексте рассматриваемых задач.

Метаописания O_M являются атрибутами всех информационных сущностей (объекты, субъекты, экземпляры, классы, подклассы, связи, свойства) семантической модели. Они служат для определения и расширения базы логического вывода онтологии. Как и в случае с отношениями, затруднительно говорить заранее о требуемом количестве экземпляров данной категории.

В общем виде семантическую модель взаимодействия разнородных АС схематично можно представить в виде, показанном на рис. 1.

Формально в общем виде представленную семантическую модель описания предметных областей разнородных АС с учетом введенных выше обозначений можно записать следующим образом: $SM_M = \langle T_M, S_M, D_M, O_M, R_M \rangle$.

Заключение

Предлагаемый подход к организации информационного взаимодействия разнородных АС основан на использовании онтологий, представляющих собой структуры данных, отражающие связи и правила, существующие между концептами предметной области. Основным достоинством онтологий и, следовательно, предлагаемого подхода является возможность реализации логического вывода. Реализуемость предлагаемого подхода была успешно проверена на задаче интеграции разнородных баз данных. Для реализации онтологического подхода при большом количестве взаимодействующих АС предложен гибридный способ взаимодействия онтологий.

Дальнейшие направления исследований связываются с разработкой способов интеграции нереляционных баз данных и созданием глобальных отраслевых онтологий и их использованием в качестве опорных универсальных глоссариев для различных предметных областей.

Литература

1. Котенко, И. В. Архитектура системы интеллектуальных сервисов защиты информации в критически важных инфраструктурах / И.В. Котенко, И.Б. Саенко // Труды СПИИРАН. – 2013. – Вып. 1 (24). – С. 21–40.

2. Котенко, И. В. Перспективные системы хранения данных для мониторинга и управления безопасностью информации / И.В. Котенко, И.Б. Саенко, О.В. Полубелова // Труды СПИИРАН. – 2013. – Вып. 2 (25). – С. 113–134.

3. Лапшин, В. А. Онтологии в компьютерных системах / В.А. Лапшин. – М.: Научный мир, 2010. – 224 с.

4. Maedche, A. Clustering Ontology-based Metadata in the Semantic Web / A. Maedche, V. Zacharias // Principles of Data Mining and Knowledge Discovery. Lecture Notes in Computer Science. – 2002. – Vol. 2431. – P. 348–360.

5. Гаврилова, Т. А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб.: Издательский дом "Питер", 2001. – 384 с.

6. Среда Описания Ресурса (RDF): Понятия и Абстрактный Синтаксис. W3C Рекомендация от 10 Февраля 2004. – Режим доступа: http://www.w3.org/2007/03/rdf_concepts_ru, свободный. – Загл. с экрана.

7. Тузовский, А. Ф. Системы управления знаниями (методы и технологии) / А.Ф. Тузовский, С.В. Чириков, В.З. Ямпольский. – Томск: Изд-во НТЛ, 2005. – 260 с.

8. OWL 2 Web Ontology Language Document Overview (Second Edition). W3C Recommendation 11 December 2012, viewed 30 September 2014, – Режим доступа: http://www.w3.org/TR/owl2-overview, свободный. – Загл. с экрана.

9. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission 21 May 2004, viewed 30 September 2014, – Режим доступа: http://www.w3.org/Submission/SWRL, свободный. – Загл. с экрана.

10. R2RML: RDB to RDF Mapping Language. W3C Proposed Recommendation 14 August 2012, viewed 30 September 2014, http://www.w3.org/TR/2012/PR-r2rml-20120814, свободный. – Загл. с экрана.