

**ПРИМЕНЕНИЕ ОНТОЛОГИЙ ПРИ РАЗРАБОТКЕ
РАСПРЕДЕЛЕННЫХ АВТОМАТИЗИРОВАННЫХ
ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫХ СИСТЕМ**

С. П. Ковалев

Институт вычислительных технологий СО РАН, г. Новосибирск

E-mail: kovalyov@nsc.ru

Дан обзор результатов применения технологий инженерии предметной области к разработке распределенных автоматизированных информационно-измерительных систем (АИИС). Представлена онтология предметной области «сетевые объекты измерений». Описана архитектура ядра типовой распределенной АИИС, построенная с помощью методов аспектно-ориентированного проектирования.

Введение. Согласно современному подходу к организации информационных систем основным предметом обработки в них являются цифровые сущности [1] – комплексные информационные единицы, объединяющие данные, информацию и знания. Под данными в этом контексте понимаются элементарные числовые значения, распознаваемые компьютерными средствами. Они интерпретируются в соответствии с метаданными, образующими информационную составляющую цифровой сущности. Знания задаются в форме структурных отношений между метаданными. Таким образом, архитектура информационной системы представляет собой трехслойную иерархию, где каждый слой реализует функции управления одной из составляющих поддерживаемых цифровых сущностей.

Программное обеспечение цифровых средств измерений традиционно относится в основном к слою управления данными. Действительно, их задача сводится к представлению характеристик объекта измерений в виде числовых значений. Если акт выполнения измерения локализован в пространстве и времени, то структура этих значений тривиальна. Однако ситуация меняется в случае, когда объект измерений представляет собой крупномасштабную распределенную сеть, состоящую из разнородных компонентов. Здесь первичные характеристики образуют массивы, значительный объем которых не позволяет непосредственно увидеть состояние объекта в целом. Требуется вычислять и предоставлять пользователям разнообразные интегральные и статистические показатели, закономерности динамики их изменения во времени, корреляции с другими показателями из смежных областей и т. д. Для удовлетворения этих требований необходим мощный ин-

формационно-аналитический слой, образующий программное ядро распределенной автоматизированной информационно-измерительной системы (АИИС).

Основные трудности, возникающие при разработке информационно-аналитического слоя, связаны с большим объемом и запутанной структурой семантического пространства объекта измерений, из которого извлекается модель метаданных [2]. Поэтому распределенные АИИС, имеющиеся на информационном рынке, представляют собой вертикальные решения для отдельных видов сетевых объектов (учет энергопотребления на предприятиях, биллинг при оказании телекоммуникационных услуг и др.). Их архитектуры обладают ярко выраженной предметной спецификой, а реализации базируются на разнородных технологиях программирования [3]. Сильная зависимость от вида объекта препятствует выработке горизонтальных (типовых) технологических решений, способных упростить и удешевить создание распределенных АИИС.

В данной работе предложен подход к преодолению этой проблемы, основанный на принципах инженерии предметной области [4]. Процесс разработки АИИС разделен на два технологических цикла: цикл инженерии предметной области (domain engineering), предназначенный для анализа и проектирования семейства систем автоматизации выделенной предметной области, и цикл инженерии приложений (application engineering), в рамках которого создаются конкретные приложения. Результатом цикла инженерии предметной области является набор формальных моделей, описывающих архитектуру ядра типовой распределенной АИИС, интерфейсы взаимозаменяемых проблемно-ориентированных компонентов, точки изменчивости и расширения системы. В целях обеспечения высокой гибкости моделей привлекаются технологии Semantic Web и аспектно-ориентированный подход к проектированию.

Структура статьи воспроизводит схему технологического цикла инженерии предметной области. В разд. 1 описываются результаты идентификации предметной области «сетевые объекты измерений». Разд. 2 посвящен моделированию этой предметной области на основе ее онтологии, а разд. 3 – проектированию типового решения по ее автоматизации.

1. Идентификация предметной области. Первым этапом технологического цикла инженерии предметной области является ее идентификация – установление границ, определение круга представляющих ее лиц и их целей, выделение смежных предметных областей. Область применения распределенных АИИС, рассматриваемых в предлагаемой работе, ограничена объектами измерений, допускающими представление в виде сети распространения потоков ресурсов (материальных объектов, энергии, информации и т. д.). Измерению подлежат значения расхода и характеристик ресурсов в определенных точках сети. Результаты измерений применяются в целях оперативного обслуживания потоков и принятия решений об управляющих воздействиях на объект. Поэтому в качестве смежных предметных областей выступают различные инфраструктуры управления, такие как бухгалтерия или контролирующие инстанции.

Основным критерием адекватности такой измерительной системы является полнота охвата потоков точками измерения: точки должны образовывать замкнутые контуры, суммарный расход в которых должен быть равен объему потребления или производства ресурса внутри (теорема Остроградского). Проверка этого условия для каждого временного среза результатов из-

мерений является одним из важнейших методов контроля их достоверности, а суммарные значения – важнейшими параметрами состояния сети. Для того чтобы проводить требуемые вычисления, необходимо непосредственно по мере проведения измерений приводить результаты к общему формату, снабжать их нужными метаданными и комплексировать их в единый массив. Поэтому недостаточно оснастить каждую точку измерения локальным измерительным комплексом – требуется создать единую интегрированную распределенную АИИС, охватывающую весь объект.

В качестве примера приведем систему учета энергопотребления крупного промышленного объединения, состоящего из множества территориально распределенных производственных единиц. К таким объединениям относятся федеральные транспортные инфраструктуры (железные дороги, магистральные нефте- и газопроводы), металлургические холдинги полного цикла и т. п. В целях оптимизации затрат объединение стремится выступать на энергетическом рынке как единый субъект, приобретающий энергоносители оптовыми партиями. Для этого оно должно внедрить систему измерения расхода энергоносителей, охватывающую все его энергетические сети [5]. Результаты измерений предоставляются энергоснабжающим организациям для выполнения финансовых расчетов и оперативно-диспетчерским службам для регулирования энергопотребления предприятия.

Потребность в распределенной АИИС возникает также при мониторинге окружающей среды [6]. Здесь измерению подлежат характеристики потоков химических веществ, распространяющихся в атмосфере и гидросфере. Точки измерения представляют собой стационарные станции наблюдения либо передвижные лаборатории, оснащенные метеорологическим и гидрологическим оборудованием. Результаты измерений используются для экологического контроля работы предприятий, предсказания погоды, оценки состояния биосферы. Примером такой системы регионального масштаба служит электронный атлас «Атмосферные аэрозоли Сибири» [7].

Приведенные примеры относятся к традиционным сферам применения АИИС в промышленности и экспериментальной науке. Современные информационные технологии позволяют выйти за их пределы, если рассматривать в качестве распределенных объектов измерений информационные порталы – единые интегрированные точки доступа к распределенным совокупностям разнородных информационных ресурсов [8]. Действительно, интеграцию ресурсов можно свести к их связыванию посредством каналов распространения информации, образующих виртуальную сеть [9]. Измерение потока информационных блоков, проходящих по каналам, заключается в автоматическом извлечении их метаданных для анонсирования (публикации) на портале. Единый формат анонсов повышает удобство восприятия и поиска тематической информации. Примером применения такого подхода служит портал математических ресурсов MathTree [10].

Эти примеры позволяют составить количественный «портрет» сетевого объекта измерений, определяющий требования к качеству распределенной АИИС. Объект содержит 10^2 – 10^5 точек измерения, измерительный цикл в них выполняется с периодом 10^1 – 10^5 с. Количество потребителей результатов измерений в режиме постоянного доступа достигает 10^2 – 10^3 . Географическая протяженность объекта составляет 10^2 – 10^4 км, обычно объект располагается в нескольких часовых поясах. Объекту присуща техническая разнородность (в нем применяются десятки наименований оборудования), административная разобщенность (его обслуживают десятки организаций), вы-

сокая степень структурной изменчивости (характерный период стабильности состояния объекта составляет 10^5 – 10^6 с).

2. Моделирование предметной области. После того как предметная область идентифицирована, необходимо произвести ее моделирование. В ходе моделирования выполняется сбор, систематизация и формализация информации о предметной области, поступающей из различных источников. Существует ряд технологий моделирования, нацеленных на построение доменных моделей различных видов. Так, FODA (Feature-Oriented Domain Analysis) и родственные методы позволяют представить предметную область в виде диаграммы характеристик – элементарных свойств, различные сочетания которых отвечают конкретным объектам автоматизации [4]. Метод RSEB (Reuse-driven Software Engineering Business) [11] задает правила объектно-ориентированной декомпозиции предметной области с фиксацией результатов на языке UML. Существуют методы, ориентированные на поддержку имеющихся во многих предметных областях специфических формализованных языков представления знаний, например метод моделирования трансформаций Draco [12].

В последнее время стали появляться технологии, рассматривающие в качестве модели предметной области ее онтологию – формализованный свод основных понятий, отношений между ними и правил логического вывода предметных знаний. Примером служит метод ODE (Ontology Domain Engineering) [13], который обладает такими преимуществами, как независимость от конкретных видов предметных областей и технологий программирования, наличие формальных аксиоматических определений понятий. Для записи онтологии предметной области «сетевые объекты измерений» мы используем язык OWL (Ontology Web Language), разработанный и стандартизованный консорциумом W3C. К его достоинствам относятся: возможность машинной обработки онтологии с привлечением технологий Semantic Web, наличие механизма логического вывода утверждений с помощью аппарата дескриптивной логики, легкость расширения онтологии новыми фрагментами (принцип «открытого мира»). Существуют средства автоматизации процесса разработки онтологии, например графический редактор Protégé [14], система поддержки распределенной работы экспертных групп ONTOGRID [15]. Отметим, что с этой точки зрения подход в данной работе отличается от метода ODE, в рамках которого предлагается специальный язык записи онтологии LINGO.

Вместе с тем мы следуем основным этапам моделирования предметной области, указанным в ODE, посредством онтологии. Прежде всего строится и формализуется базовая онтология, позволяющая зафиксировать основные задачи системы. В нашем случае источником терминов для нее естественным образом служит метрология – наука о выполнении измерений. Ее положения стандартизованы и изложены в нормативных документах, послуживших основными источниками систематизированных знаний о предметной области. В частности, ключевые метрологические термины и определения зафиксированы в стандарте ГОСТ Р 8.596–2002 «Метрологическое обеспечение измерительных систем. Основные положения» [16]. Он определяет измерительную систему (Measuring_system) как совокупность компонентов (Component): измерительных (Measuring_component), вычислительных (Computing_component), связующих (Connecting_component), вспомогательных (Auxiliary_component), комплексных (Complex_component), образующих измерительные каналы. Измерительный канал (Measuring_channel) –

это часть измерительной системы, выполняющая законченную функцию от восприятия измеряемой физической величины (Physical_quantity) до получения результата ее измерения. Точка измерения в сетевом объекте – это место подключения измерительного канала-источника первичных данных, так что измерительные каналы выступают в роли контейнеров метаданных результатов измерения. Физически компоненты измерительных каналов реализуются средствами измерений (Measuring_instrument), причем измерительная система как целое также является сложным средством измерений.

Общая схема получившейся базовой онтологии представлена на рис. 1: понятия обозначены овалами, структурные (эпистемологические) отношения частное/общее (is-a) – сплошными линиями, прочие отношения – пунктирными линиями.

Базовая онтология служит ядром для интеграции необходимых знаний о конкретном сетевом объекте. Эти знания подключаются по возможности в виде готовых онтологий. Например, понятие физической величины конкретизируется понятиями, зафиксированными в онтологии физических величин и единиц их измерения согласно Международной системе единиц. Общее понятие средства измерений конкретизируется типами и техническими характеристиками используемых устройств. Большой набор подходящих для этого онтологий разработан в рамках модели CIM [2]. Еще одним важным расширением служат понятия и отношения, описывающие жизненный цикл измерительных каналов: виды и характеристики мероприятий по их техническому обслуживанию (ввод в эксплуатацию, ремонт, метрологическая поверка и др.), номенклатура сопутствующей нормативно-справочной информации, организационная и технологическая структура объекта измерений. Фиксация всех этапов жизненного цикла измерительных каналов в АИИС наделяет ее функциями системы класса PLM (Product Lifecycle Management).

3. Проектирование предметной области. Модели предметной области служат источником информации для проектирования семейства систем, обеспечивающих автоматизацию предметных задач. В ходе проектирования формируется архитектура ядра, фиксируются интерфейсы взаимозаменяемых компонентов, принимаются ключевые решения по реализации системы.

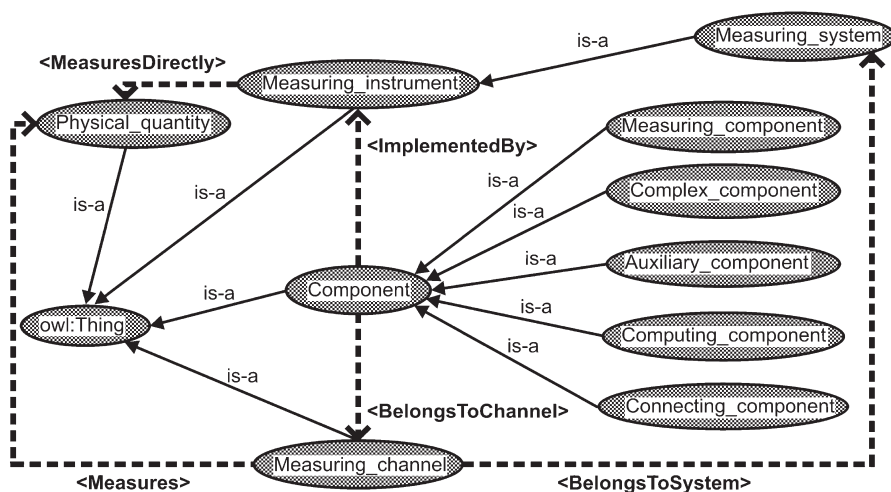


Рис. 1

Выбор представления результатов определяется стилем проектирования предметной области. В [4, с. 80] перечислены следующие стили:

- стиль взаимодействия, выражающий взаимодействие в виде диаграмм вариантов использования, исчислений процессов и т. п.;
- алгоритмический стиль, задающий алгоритмы вычислений с помощью различных псевдокодов;
- информационно-ориентированный стиль, описывающий модели данных в виде диаграмм «сущность–связь»;
- потоковый стиль, фиксирующий потоковую структуру системы в диаграммах потоков данных.

Традиционно рекомендуется выбрать какой-либо один стиль, наилучшим образом отражающий задачу автоматизации. Однако при создании распределенной АИИС такой выбор затруднен вследствие того, что в ней представлены все перечисленные стили, тесно переплетенные между собой в рамках измерительного цикла. Для того чтобы выполнять проектирование в этих условиях, мы привлекаем аспектно-ориентированный подход [17]. В рамках этого подхода единицами декомпозиции системы выступают аспекты – логически замкнутые сквозные потоки функциональности, подлежащие отдельной реализации и последующему связыванию в явно заданных точках. Многие технологии разработки программного обеспечения специально ориентированы на реализацию отдельных аспектов, поэтому аспектная декомпозиция облегчает организацию деятельности коллектива разработчиков, являющихся узкими специалистами. Аспектная архитектура располагает к привлечению гибких инкрементных технологий реализации, поскольку модификации аспектов, не затрагивающие интерфейсов точек связывания, локализуются внутри их кода и не требуют модификации остальных аспектов. Характерным примером аспекта является защита информации: меры защиты должны срабатывать при каждом акте доступа к хранилищу информации независимо от решаемой при этом предметной задачи, предоставляя субъекту доступа внутренне согласованный набор разрешений.

Аспекты взаимодействия, представленные в распределенной АИИС, обслуживают три вида контрагентов: инструментальные средства измерения, пользователей и смежные автоматизированные системы. Аспект взаимодействия со средствами измерения должен отвечать требованиям реального времени, поэтому он реализуется с применением технологий критического характера, таких как операционная система реального времени QNX. Аспект взаимодействия с пользователями отвечает за требования эргономики (удобства), поэтому имеет вид портала, реализуемого на базе web-технологий. Аспект взаимодействия с внешними системами отвечает за требования совместимости с ними, поэтому для его реализации привлекаются интеграционные платформы на базе технологий Semantic Web. Основным источником терминов и схем протоколов взаимодействия и форм пользовательского интерфейса служит онтология предметной области.

Алгоритмические аспекты выполняют функции хранения и анализа результатов измерений. В рамках этих аспектов формируются косвенные (расчетные) результаты измерений, проводится контроль достоверности значений, выполняется прогнозирование состояния объекта, определяются управляющие воздействия на средства измерений. Структура потока выполнения алгоритмов задается правилами вывода предметных знаний, содержащими

ся в онтологии. Алгоритмы реализуются на базе современных технологий анализа данных класса OLAP (Online Analytical Processing).

Основным информационным аспектом системы является модель организации измерений – хранилище метаданных, полностью описывающих условия формирования каждого результата измерений. Она представляет собой результат трансформации онтологии предметной области в диаграмму «сущность–связь», реализуемую в виде реляционной базы данных. В состав модели организации измерений входят следующие массивы данных:

- 1) модель объектов измерений, содержащая реестры их структурного состава;
- 2) модель средств измерений, содержащая реестры их состава и технических характеристик;
- 3) модель компонентов, фиксирующая компонентный состав измерительных каналов;
- 4) модель расчета результатов косвенных измерений, определяющая алгоритмы получения результатов косвенных и производных измерений из результатов прямых;
- 5) модель процесса измерения, включающая состав и расписания выполнения этапов измерительного цикла;
- 6) модель нормативно-справочной информации, представляющая собой хранилище документального обеспечения системы.

Базовая онтология (см. рис. 1) порождает каркас модели компонентов. Ее диаграмма «сущность–связь» представлена на рис. 2.

Сравнение рис. 1 и 2 позволяет увидеть «смещение центра тяжести», происходящее при переходе от анализа к проектированию. Если онтология нацелена на представление предметной области в наиболее понятном виде, то структура базы данных должна в первую очередь обеспечивать эффектив-

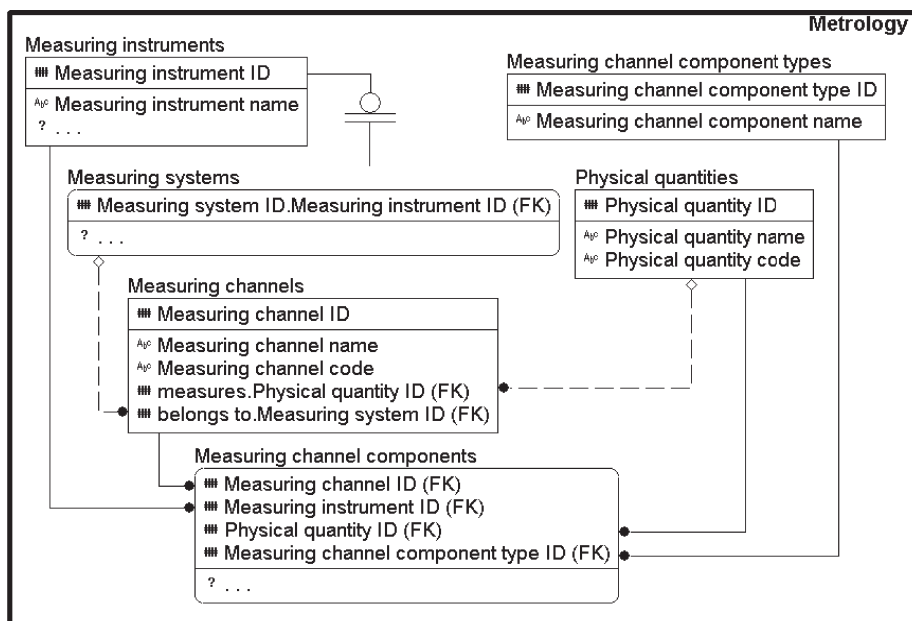


Рис. 2

ность выполнения запросов на выборку и запись данных. В соответствии с этим разбиение онтологического класса «Компонент измерительного канала» (Component) на подклассы реализовано при помощи справочника типов компонентов. Кроме того, онтологические отношения «Компонент реализуется средством измерения» (ImplementedBy) и «Средство измерения измеряет физическую величину» (MeasuresDirectly) свернуты путем композиции в одну реляционную ссылку, направленную из реестра компонентов в справочник физических величин. В результате компонентный состав измерительных каналов определяется четырехместным отношением

$$MCC \subseteq MC \times MI \times PQ \times MCCT,$$

где MCC (Measuring channel components) – множество компонентов; MC (Measuring channels) – множество измерительных каналов; MI (Measuring instruments) – множество средств измерения; PQ (Physical quantities) – множество физических величин; MCCT (Measuring channel component types) – множество типов компонентов.

Коротко рассмотрим применение потокового стиля при проектировании распределенной АИИС. Он удобен для реализации координации функционирования аспектов. Применяется реактивная модель координации, согласно которой система снабжается аспектом мониторинга состояния и реагирования на события. Этот аспект включает ведение единого журнала событий, отражающих изменения состояния системы, и оперативное оповещение аспектов обработки событий о фактах их возникновения. Таким образом, в системе постоянно поддерживаются потоки событий, поступающих в журнал, и сигналов оповещения, выходящих из него. Богатый набор правил оповещения позволяет придать системе высокий уровень автономности – способности автоматически оптимизировать свою конфигурацию при изменениях во внешнем окружении. Для реализации необходимых функций можно применить технологию автономных вычислений [18].

Заключение. Автоматизация крупномасштабных объектов часто бывает затруднена вследствие большого размера и сложности моделей данных, вынужденных отражать гигантские объемы предметных знаний. В представленной работе предложен подход к преодолению этой проблемы путем выполнения технологического цикла инженерии предметной области, направленного на построение компактной базовой модели с явно выделенными точками гибкого инкрементного расширения. В качестве такой модели предлагается выбрать онтологию предметной области, и на этапе проектирования трансформировать ее в формальные модели различных аспектов системы. Приложением данного подхода является задача создания семейства распределенных АИИС крупных сетевых объектов. Решение других задач на его основе является перспективным направлением дальнейших исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Moore R. W., Baru C. Virtualization services for Data Grids // Grid Computing: Making the Global Infrastructure a Reality. N. Y.: Wiley & Sons, 2003. P. 409.
2. Appelrath H.-J., Uslar M., Lucks A. et al. Interaction of EMS related systems by using the CIM standard // Proc. Second Intl. ICSC Symp. on Information Technologies in Environmental Engineering. Magdeburg: Otto-von-Guericke-Universität, 2005. P. 596.

3. **Grimaldi D., Marinov M.** Distributed measurement systems // Measurement: Journ. Intern. Measurement Confederation. 2001. N 30. P. 287.
4. **Чарнецки К., Айзенекер У.** Порождающее программирование: методы, инструменты, применение. С.-Пб.: Питер, 2005.
5. **СТО АТС 02.07.3-2003.** Коммерческий учет на оптовом рынке электроэнергии. Автоматизированные информационно-измерительные системы. Общие требования. М.: АТС, 2003.
6. **Алексеев В. В.** Проблемы получения информации и оценки экологической обстановки в ИИС мониторинга окружающей природной среды // Мониторинг и прогнозирование чрезвычайных ситуаций. Тез. докл. НТК. С.-Пб.: СПГЭТУ, 1998.
7. **Kovalyov S. P., Molorodov Yu. I.** Data integration methodology for atlas "Atmospheric aerosols of Siberia" // Proc. SPIE. 2005. 6160. Pt. I. P. 9.
8. **Надеин А., Кузнецов В.** Корпоративные Интранет-порталы. М.: TopS Business Integrator, 2003.
9. **Ковалев С. П., Яковченко К. Н.** Организация информационных порталов на основе канальной интеграции // Тр. Междунар. конф. по вычислительной математике «МКВМ-2004». Новосибирск: ИВМиМГ СО РАН, 2004. С. 66.
10. **Баряхнин В. Б., Клименко О. А., Ковалев С. П.** Сбор и систематизация информации для портала математических ресурсов MATHTREE // Тр. Междунар. конф. «Вычислительные и информационные технологии в науке, технике и образовании». Павлодар: ТОО НПФ «ЭКО», 2006. Том II. С. 381.
11. **Jacobson I., Griss M., Jonsson P.** Software Reuse: Architecture, Process, and Organization for Business Success. Reading: Addison-Wesley, 1997.
12. **Neighbors J. M.** Draco: a method for engineering reusable software systems // Software Reusability. Reading: Addison-Wesley, 1989. Vol. I. P. 295.
13. **Falbo R. A., Guizzardi G., Duarte K. C.** An ontological approach to domain engineering // Proc. XIV Intl. Conf. Software Engineering and Knowledge Engineering (SEKE-2002). Ischia, Italy, 2002. P. 351.
14. **KSL Protege Project.** Stanford University // <http://protege.stanford.edu>
15. **Загоруйко Н. Г., Гусев В. Д., Завертайлов А. В. и др.** Система OntoGrid для автоматизации процессов построения онтологий предметных областей // Автометрия. 2005. 41, № 5. С. 13.
16. **ГОСТ Р 8.596-2002.** Метрологическое обеспечение измерительных систем. Основные положения. М.: Госстандарт России, 2002.
17. **Aspect-Oriented Software Development** /Eds. R. E. Filman, T. Elrad, S. Clarke, M. Aksit. Wokingham: Addison Wesley, 2004.
18. **Murch R.** Autonomic Computing. N. Y.: IBM Software press, 2004.

Поступила в редакцию 11 мая 2007 г.