

УДК 004.822

А.В. Григорьев, Е.В. ПавловскийДонецкий национальный технический университет, г. Донецк
кафедра прикладной математики и информатикиE-mail: grigorie@pmi.dgtu.donetsk.ua, pavlovskiy.evgeny@mail.ru**АНАЛИЗ МЕТОДОВ ПОСТРОЕНИЯ ОНТОЛОГИЙ ДЛЯ ПОСТРОЕНИЯ
ЭКСПЕРТНЫХ СИСТЕМ ПО СИНТЕЗУ МОДЕЛЕЙ СЛОЖНЫХ СИСТЕМ В САПР****Аннотация**

Григорьев А.В., Павловский Е.В. Анализ методов построения онтологий для построения экспертных систем по синтезу моделей сложных систем в САПР. Онтологии из философского понятия стали передвигаться в базы знаний искусственного интеллекта. Это средство создания концептуальной модели для инструментальных средств создания экспертных систем. В работе рассматриваются существующие онтологии в разных предметных областях, предлагается онтологический подход для создания сайтов.

Ключевые слова: онтология, САПР, средний эксперт, база знаний, и-или дерево.

Общая постановка проблемы.

На сегодняшний день актуальной является задача формирования концептуальных и понятных представлений для слабоструктурированных предметных областей. В некоторых предметных областях информация расплывчата и неоднозначна. С целью «понятного» представления данных используют для представления данных онтологию. Это иерархическая структура, которая связывает между собой понятия предметной области. Информация, представленная в виде онтологии, является концептуальным описанием предметной области. Она легко изменяема, имеет наследственные связи, что характерно для объектно-ориентированного программирования. Для того чтобы внести изменения в один блок данных (понятие), не нужно затрагивать всю иерархию, а следует только изменить понятие. При этом прежние отношения между понятиями будут справедливыми. Вообще, древовидная система данных – это хорошо структурированный механизм для обработки данных. Обычно с узлами такого дерева работают по технологии вложенных множеств. Чтобы иерархия была однозначной и целостной, минимальное требование для узла – это его идентификатор и идентификатор его родителя. В отличие от простого дерева, механизм онтологии сложнее по своей структуре. Каждый блок представляет собой понятие предметной области, которое может иметь различные типы связей между подблоками и блоками различных уровней. Онтология – это спецификация концептуализации по определению Тома Груббера [1], впервые применившего это понятие в области информационных технологий. Термин онтологии – это философское понятие, которое означающий учение о бытии. Со временем это понятие стало использоваться не только в философии, но и области точных наук. Онтология – это точная спецификация некоторой предметной области. Другими словами это модель данных, которая описывает понятия предметной области, что они означают, как они соотносятся друг с другом, и как они могут или не могут быть связаны друг с другом. Эта модель отображает представление о знании в некоторой предметной области и множество связей между ними. Сейчас распространенным явлением стало использование многоагентных интеллектуальных систем, в которых очень активно используются онтологии. Наряду с многоагентными системами, онтологию активно используют системы управления знаниями. Существует ряд различных языков и систем, использующих онтологию. Более понятны для пользователя и востребованы системы,

которые предоставляют визуальный подход, позволяющий специалистам проектировать онтологию в графическом виде, что помогает наглядно сформулировать и объяснить природу и структуру явлений [2]. Достаточно хорошим представлением данным обладают базы данных (БД). Они описывают тип данных записей и отношения между полями и записями, но этого недостаточно для описания предметной области. В БД нет описания семантики данных, что приводит к разработке новых приложений, если технические условия предметной области меняется. Хорошей альтернативой БД является онтология, а точнее представление данных в виде онтологии. Здесь отношение между понятиями определяются новым техническим условием, согласно семантике данных. В последние годы были разработаны инструменты и системы для создания метаданных онтологии, например, Protégé, Ontolingua, OntoSaurus. Эти инструменты кодируют понятия в онтологии, которую могут понять компьютеры. Однако получение понятий и описание связи между этими понятиями все еще остаются трудной проблемой.

Постановка задач исследования.

Сейчас тенденции в программировании ведут к созданию генераторов программ, решающих ту или иную проблему в виде базы знаний (БЗ). Имеется два подхода к построению генератора программ: 1) генерация предметно независимого языка с кодированием в проблемно зависимый язык; 2) генерация на проблемно зависимом языке.

При создании генератора нужно ориентироваться на соответствующие уровни квалификации эксперта предметной области, который обеспечивает созданием генератора как БЗ. Фактически эта проблема передачи знаний от эксперта к БЗ. Авторами рассматривается три подхода [3]: 1) эксперт не в состоянии сформировать БЗ и инструментальная система должна извлекать знания из прецедентов; 2) пользователь может описать множество структур как И-ИЛИ дерево и описать на нем отношения совместности или несовместности вариантов; 3) пользователь в состоянии явно сформировать набор продуктов, составляющих БЗ. Средний эксперт достаточно хорошо знает предметную область, может описать связи между объектами и представить их в виде И-ИЛИ дерева. Именно на среднего эксперта предметной области будет ориентирован генератор программ. Нашей задачей является исследование технологии создания генератора сайтов на базе онтологий в режиме среднего эксперта предметной области. Сам по себе язык html уже является онтологией. Теги языка подчинены друг другу, имеют четкие связи между собой, атрибуты и представляют собой иерархию. Задача автоматического проектирования сайтов является на сегодняшний день актуальной, так как растет количество пользователей интернета и потребность размещения информации в глобальной сети.

Целью работы является: 1) выполнить анализ имеющихся подходов и методов для создания онтологии; 2) предложить порядок построения онтологий САПР в режиме среднего эксперта; 3) предложить начальный вариант программного комплекса, для генерации текста на языке HTML.

Решение задач и результаты исследований.

1. Определение онтологии

На сегодняшний день существует множество определений понятия онтологии. Многие понятия дополняют предыдущие, а некоторые даже противоречат друг другу. Википедия [4] определяет онтологию (в информатике) как попытку всеобъемлющей и детальной формализации некоторой области знаний с помощью концептуальной схемы. Под концептуальной схемой подразумевается набор понятий + информация о понятиях (свойства, отношения, ограничения, аксиомы и утверждения о понятиях, необходимых для описания процессов решения задач в избранной предметной области). С помощью онтологии описывают знания о предметной области. Эти знания включают понятия предметной области, из взаимоотношений между ними, значимость каждого понятия и их свойства. Можно понимать онтологию как средство создания концептуальной модели для инструментальных

оболочек для создания экспертных систем. Т.е., назначением онтологий является представление знаний о предметной области, которые могут быть впоследствии обработаны программой (в нашем случае экспертной системой). Обычно [4] под онтологией подразумевается явная спецификация концептуализации, где в качестве концептуализации выступает описание множества объектов и связей между ними. Формально онтология состоит из понятий терминов, организованных в таксономию, их описаний и правил вывода. Формально онтология определяется как (базовое формальное определение):

$$O = \langle X, R, F \rangle,$$

где: X — конечное множество понятий предметной области; R — конечное множество отношений между понятиями; F — конечное множество функций интерпретации.

Обычно выделяют такие отношения между понятиями: класс – подкласс; часть – целое; экземпляр – класс. Нариньяни А.С. в работе [5] понимает под онтологией набор определений (на формальном языке) фрагмента декларативных знаний, ориентированный на совместное многократное использование различными пользователями в своих приложениях. Автор рассматривает онтологию и тезаурус, как ядро с активным использованием компьютерной лингвистики интеллектуальной системы. Тезаурус закрепляется за лексикой в проекции на семантику, а онтология употребляется как семантика и прагматика. Вводится понятие идеального определения (при разное в терминологии идеальное вряд ли возможно). Под Тезаурусом понимается словарь, в котором максимально полно представлены все слова языка с исчерпывающим перечнем примеров их употребления. Понятие онтологии близко с Тезаурусом, но, по мнению автора статьи, оно более сложное.

2. Классификация онтологий.

Выделяют два типа классификации онтологий – семантическая и прагматическая. Дадим краткую характеристику классификации.

1) *Семантическая классификация*: по уровню выразительности; по степени формальности; по уровню детальности представления. По уровню выразительности онтологии бывают тяжеловесные и легковесные. Тяжеловесные онтологии в отличие от легковесных, более однозначны, так как правила в них указаны явно. Аксиоматизация позволяет избежать терминологической неоднозначности из-за неправильной интерпретации. Т.о., тяжеловесные онтологии, хотя и более трудоемки в процессе построения, можно считать более предпочтительными на практике. Онтологии, которые описываются на произвольном языке, носят название неформальных, а онтологии, которые строго типизированы, с заданной семантикой термов – формальная онтология. Бывают онтологии, которые более детализированы. В них может содержать много деталей, включая правила, по которым термины могут быть связаны между собой. Такие онтологии имеют высокий уровень детализации.

2) *Прагматическая классификация*: по степени зависимости от конкретной задачи или прикладной области; по языку представления онтологических знаний; по предметной области; по цели создания; по наполнению (содержимому). По степени зависимости от конкретной задачи или прикладной области можно выделить онтологии верхнего уровня, ориентированные на предметную область и ориентированные на задачу. Онтологии верхнего уровня описывают онтологию общими понятиями и связями. Такие онтологии описывают наиболее общие концепты (пространство, время, материя, объект, событие, действие и т.д.), которые не зависят от конкретной проблемы или области. На сегодняшний день существует множество онтологий, которые применяются в конкретной предметной области. В области медицины созданы большие стандартные, структурированные словари, такие как SNOMED. Когда онтология используется для решения конкретной задачи, то говорят, что она ориентирована на задачу. Такие онтологии используются программным обеспечением, ориентированным на решение конкретной задачи в конкретной предметной области. Классификация онтологии по предметной области является важнейшим отличием одной

онтологий от другой. Онтология отражает общие знания о предметной области, такие, как иерархия классов понятий и семантические отношения на этих классах. Для каждой предметной области онтологии создаются экспертами своей области, которые проводят формализацию знаний, определений и правил получения новых знаний.

3. Построения онтологий в различных предметных и проблемных областях.

Актуальной является проблема разработки методологии проектирования формальной онтологии для заданной предметной области. На сегодняшний день нет общепринятой методологии, которую бы все придерживались. Проектировщики выбирают для себя модель представления знаний и средства для их извлечения произвольно, согласно своим предпочтениям и видения решения задачи. Проектирования формальной онтологии подразумевает под собой сложный программный комплекс, состоящий из инструментальных средств и алгоритмов, которые в сумме обеспечивают разработчику онтологии автоматизированный и интерактивный процесс проектирования. По своей функциональной полноте различают три вида онтологий [6]: полная (или строгая); простая; промежуточная онтология, представляющая собой один из вариантов множества неполных онтологий, т.е. - промежуточных между полной онтологией и простой. Строгая или полная онтология характеризуется максимально полными множествами концептов и концептуальными отношениями. При этом аксиомы, определения и ограничения представлены на некотором формальном языке, доступном для их интерпретации компьютером [7]. Из базового формального определения онтологии следует, что есть понятия, отношения между ними и функции. Но ничего не сказано об ограничениях на значения атрибутов понятий. Без ограничений модель была бы не полна, даже в некоторых случаях – бесполезна. Ведь в этом случае не будет проверки корректности информации для вывода. Автор [8] приводит следующее расширенное определение онтологической системы:

$$\Sigma^o = \langle O^{meta}, \{O^{d\&t}\}, \Xi^{inf} \rangle, \quad (1)$$

где: O^{meta} - онтология верхнего уровня (метаонтология), которая не зависит от предметной области; $\{O^{d\&t}\}$ - множество предметных онтологий и онтологий задач предметной области; Ξ^{inf} - модель машины вывода, ассоциированной с онтологической системой.

Формальное описание онтологии существенно отличается от базового понятия. Этот факт очередной раз подтверждает, что онтология - понятие обширное. Можно одну задачу решить несколькими способами. Формальное описание онтологии зависит не только от предметной области, но и от видения автором решения задачи. В данной модели $\{O^{d\&t}\}$ и Ξ^{inf} должны быть не пустыми. Вообще онтология задач – эта онтология, которая служит для решения какой-то задачи (подзадачи). На основании подзадачи можно решать более сложные задачи. Система разбивается на множество задач, которые направлены на решение глобальной задачи. Имеются также и другие работы, направленные на расширение формальной модели онтологии. Так, в [9] авторы рассматривают подход к разработке решателей задач на основе моделей онтологий. Они задают связи между входными и выходными параметрами и терминами предметной области.

Модель системы [9] состоит из четырех компонентов:

$$T = \langle O, IN, Out, C \rangle, \quad (2)$$

где: O - небогатая система логических компонентов; IN, OUT - множество определений входных и выходных параметров; C - ограничения на значения неизвестных системы O .

Сама по себе предметная область хорошо структурирована, четко прослеживаются уровни онтологии и отношения между ними. Как и в базовом определении онтологии, концепты характеризуются названием, атрибутами и экземплярами концептов. В [10]

рассматривается построение онтологии интеллектуальной системы мониторинга учебного процесса дистанционного образования (ДО). Приводятся различные модели онтологий, аргументируется их использование в виде онтологической системы. Дается формальное описание онтологии интеллектуальной системы предметной области ДО на основе кортежей множеств. В качестве концептуального моделирования используются диаграммы UML, расширенные дополнительными типами отношений. Формальное определение онтологии (ДО) [10]:

$$S = \langle O^{\text{meta}}, \{O^{\text{domain}}\}, I^f \rangle, \quad (3)$$

где O^{meta} – метаонтология $\langle X, R, \emptyset \rangle$; $\{O^{\text{domain}}\}$ – множество предметных онтологий $\langle X, R, F \rangle$, I^f – машина вывода, ассоциированная с S .

Автор для проектирования системы мониторинга учебного процесса использует стандартное расширенное определение онтологии, в которой присутствует система вывода. Концепты разделяют на уровни (домены), которые дают четкое структурирование и более ясное представление о предметной области. Система расширена вводом связей между концептами, которая разделяет их на синонимы, антонимы и другие классы дескриптивной логики [10].

Из обзора имеющихся онтологий в различных предметных областях, можно сделать вывод о том, что для проектировщика онтологий есть большой простор для создания своих методов представления концептов предметной области, связей между ними, а также представления онтологий в виде множества отдельных областей данных. Ведь онтология изначально служит для более ясного и понятного представления данных для экспертов предметных областей. Это способствует автоматизации процесса доработки и адаптации данных под конкретную предметную область.

4. Преимущества предлагаемого подхода в области создания программного комплекса для генерации web-приложений

Покажем, что предлагаемый подход может быть с успехом применен в области создания программного комплекса для генерации web-приложений. Инструментальная система для создания генератора web-приложений рассчитана на среднего эксперта в предметной области средней квалификации (далее – средний эксперт). Средний эксперт достаточно хорошо знает предметную область, т.е. модель объекта проектирования, но затрудняется описать методику проектирования. Такой эксперт может описать связи между объектами и представить варианты структур отношений в форме И-ИЛИ дерева, а так же задать отношения совместности или несовместности между различными комбинациями значений ИЛИ-узлов, вычлняя тем самым реальные прототипы.

Создание генератора программ, ориентированного на среднего эксперта, является актуальной задачей, так как большинство потенциальных создателей и пользователей таких генераторов являются чаще всего именно средними специалистами в своей предметной области. Такие специалисты чаще всего ориентированы на представление информации в виде графических моделей изображения.

Как правило, такие специалисты прекрасно знают, как устроена система, какие в ней используются классы или термины, знают, как взаимосвязаны классы между собой, какие между ними зависимости, но при этом затрудняются: описать систему на формально зависимом языке; задать систему правил вывода, обеспечивающих выбор требуемого решения из ряда возможных по их характерным особенностям. Предполагается, что упомянутая самая трудная часть задачи будет решаться инструментальной системой автоматически.

Готовый генератор на основе задания ряда отличий требуемого решения, будет генерировать на выходе формально независимый текст, описывающий полученный прототип. Этот текст в последствии будет обработан транслятором с последующим

переводом в проблемно-ориентированный язык. В нашем случае это будет браузер интернета. На выходе мы получим текст, на языке разметки HTML.

Наша задача – представить сайт в виде онтологии. Чтобы это сделать, нужно понимать, что такое онтология. Нам нужно описать понятия, связи между ними, атрибуты понятий и ограничения. Ограничения задает эксперт предметной области. Ему нужно прописать взаимоисключающие элементы.

К примеру, если одно семантическое понятие присутствует в структуре страницы сайта, то другое не может присутствовать на данной странице. Или же, в случае более глобальных ограничений, если на всем сайте используется табличная верстка, то нельзя одновременно использовать деление страницы на участки фреймами. Минимум информации, которая нужна для представления сайта в виде онтологии: это совокупность понятий (форма регистрации, логотип, главное меню и т.д.); связи между понятиями (в большинстве случаев включение одного понятия в другой); ограничения на размещения тех или иных понятий. Чтобы автоматически создавать страницы сайтов на основании онтологий, нужно разобраться в следующих деталях: из каких компонентов может состоять страница сайта в принципе; конкретный состав компонентов; взаиморасположение компонентов; атрибуты компонентов и его составляющих.

На сегодняшний день создание сайтов – очень актуальная задача. Для этой задачи, на сегодняшний день, разработана масса программных комплексов, которые создают сайты не с нуля, а на основании шаблонов. Такие системы получили название CMS, что расшифровывается как «Content Management Software» («программное обеспечение для управления содержимым»). У этой системы есть программная база, которая расширяется в зависимости от требований разработчика готовыми компонентами. Такие системы имеют свои преимущества и недостатки.

К преимуществам можно отнести тот факт, что пользователю достаточно владеть базовыми знаниями HTML и CSS, а так же то, что в этом случае достигается высокая скорость создания сайта.

Недостатком таких систем является низкая скорость обработки на сервере. Т.е. такой сайт более «тяжелый», чем тот, который может быть создан с нуля опытным разработчиком. При этом, каким бы инструментом не пользовался для создания сайтом пользователь, ему все равно приходится изучать язык разметки и таблицу стилей. В предлагаемом подходе эксперт предметной области строит архитектуру сайта в виде онтологии, задает связи и ограничения. Вариант такого подхода изображен на рисунке 1.

Здесь в левой части приводится архитектура сайта, т.е. взаимосвязи между его страницами. В правой части приводится архитектура главной страницы сайта. Таблица включает в себя логотип, главное меню, вертикальное меню, текст и низ сайта. Все эти блоки разбиты на уровни. Такой элемент как таблица, должна включать в себя элементы определенных уровней, чтобы отображение элементов в таблице было корректным. Затем, в свою очередь, нужно разбить каждый подблок главной страницы на подблоки. Ведь главное меню, скорее всего, будет также отображено в таблице, в которой в виде ссылок будут отображены пункты меню. Требуется ввод нового понятия «ссылка».

При таком подходе создается система, которая способна ориентироваться в множестве возможных структур, и в случае необходимости - решать задачу наследования свойств страниц. Так, выбор структуры главной страницы будет влиять на структуру других страниц сайта, учитывая тот факт, что главная страница находится на вершине иерархии. Возможно прервать процесс определения структур на текущей странице, с тем, что бы следующие страницы по иерархии наследовали структуру ближайшего своего родителя. На рисунках 1 и 2 представлены в виде онтологии структура страниц сайта, а также связи между ними.

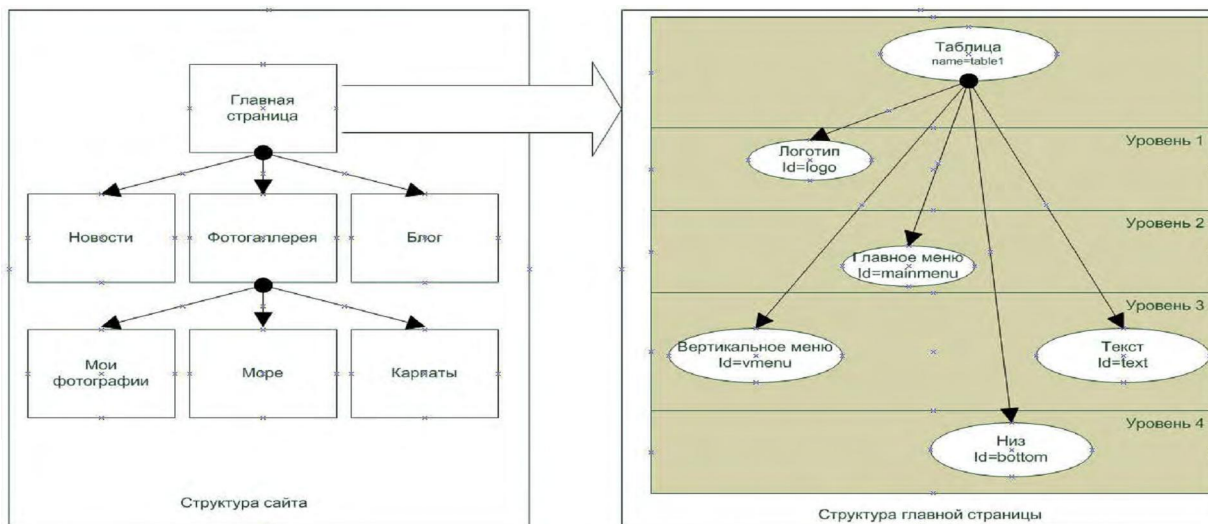


Рисунок 1 – Представление сайта в виде онтологии

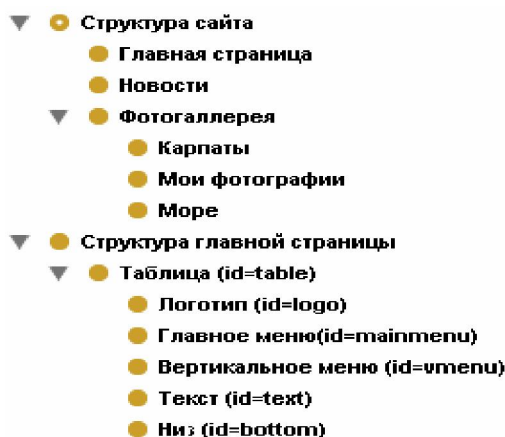


Рисунок 2 – Представление страниц сайта в виде дерева

Выводы. В работе решены все задачи, поставленные целью работы, т.е.: выполнен анализ имеющихся подходов и методов для создания онтологии; предложен порядок построения онтологий САПР в режиме среднего эксперта; предложен начальный вариант программного комплекса, для генерации текста на языке HTML.

Продланное исследование позволяет сделать следующие выводы: 1) Онтологии на сегодня представляют единственный реальный путь создания концептуальной модели предметной области; 2) Полноценная онтология, чтобы быть востребованной практикой и иметь высокую эффективность, должна иметь агрегированный вид и объединять фактически весь комплекс рассмотренных определений.

Т.о., перспективными задачами является: 1) разработка формального определения онтологий для построения САПР в режиме среднего эксперта; 2) реализация инструментальной оболочки для построения интеллектуальных систем на базе данного подхода; 3) применение данного подхода для автоматизации сайтов требуемого назначения.

Литература

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

2. Гаврилова Т.А. Использование онтологий в системах управления знаниями / Т.А. Гаврилова // Труды международного конгресса «Искусственный интеллект в XXI веке», Дивноморское, Россия. - М.: Физматлит, 2001. - С. 21-33.

3. Григорьев А.В. Пути создания интеллектуальных САПР при различных уровнях квалификации экспертов / А.В. Григорьев // Научно-теоретический журнал «Искусственный интеллект». – 2005. – №3. - С. 758–763.

4. Онтология [Электронный ресурс]. Энциклопедия Википедия. – Режим доступа: <http://ru.science.wikia.com/wiki/Онтология>.

5. Нариньяни А.С. Кентавр по имени теон: тезаурус+онтология [Электронный ресурс] / А.С. Нариньяни. - Режим доступа: <http://www.dialog-21.ru/materials/archive.asp?id=6771&y=2001&vol=6077>.

6. Палагин О.В. Архитектурно-онтологічні принципи розбудови інтелектуальних інформаційних систем / О.В. Палагин, М.Г. Петренко // Математичні машини і системи. – 2006. - №4. – С. 15-20.

7. Палагин А.В. К проектированию онтологоуправляемой информационной системы с обработкой естественно-языковых объектов [Электронный ресурс] / А.В. Палагин, Н.Г. Петренко. – Режим доступа: <http://www.aduis.com.ua/books/4.pdf>.

8. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – С-Пб.: Питер, 2000. – 384 с.

9. Артемьева И.Л., Крылов Д.А. Концепция оболочки для разработки решателей задач на основе моделей онтологий [Электронный ресурс] / И.Л. Артемьева, Д.А. Крылов – Режим доступа: http://iai.dn.ua/public/JournalAI_2005_3/Razdel2/02_Artem'eva_Krylov_prav.pdf.

10. Некрашевич С.П. Построение модели онтологии интеллектуальной системы мониторинга учебного процесса дистанционного образования [Электронный ресурс] / С.П. Некрашевич – Режим доступа: http://www.nbu.gov.ua/portal/natural/ii/2009_2/3%5C00_Nekrashevich.pdf.

Надійшла до редакції:
26.02.2011

Рекомендовано до друку:
д-р техн. наук, проф. Скобцов Ю.О.

Abstract

Grigor'ev A.V., Pavlovskiy E.V. Analysis of methods of construction of ontologies for the construction of consulting models on the synthesis of models of the difficult systems in CAD. From a philosophical concept it was begun ontology to move in the bases of knowledges of artificial intelligence. It is a mean of creation of conceptual model for the tools of creation of consulting models. In this work, existing is examined ontology in разных subject domains, and ontological approach is offered for creation of sites.

Keywords: ontology, CAD, middle expert, base of knowledges, and-or tree.

Анотація

Григор'єв О.В., Павловський Е.В. Аналіз методів побудови онтологій для побудови експертних систем по синтезу моделей складних систем в САПР. Онтології з філософського поняття почали пересуватися в бази знань штучного інтелекту. Це засіб створення концептуальної моделі для інструментальних засобів створення експертних систем. У роботі розглядаються ті, що існують онтології в різних наочних областях, і пропонується онтологічний підхід для створення сайтів.

Ключові слова: онтологія, САПР, середній експерт, база знань, та-або дерево.

© Григорьев А.В., Павловский Е.В., 2011