

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
СХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
імені ВОЛОДИМИРА ДАЛЯ**

ВІСНИК

**Східноукраїнського
національного університету
імені ВОЛОДИМИРА ДАЛЯ**

**№ 8 (179)
2012 -Ч2**

НАУКОВИЙ ЖУРНАЛ

Луганськ 2012

УДК 004.2: 004.3

Палагин А.В., Петренко Н.Г., Величко В.Ю.,
Малахов К.С. Семенов В.В., Тихонов Ю.Л.

Программные модели Инструментального комплекса онтологического назначения (ИКОН): библиотека словарей предметной области (ПдО)

В статье описана программная модель цифровой библиотеки справочной информации, включающей энциклопедии, толковые словари и тезаурусы домена близких предметных областей. Определения и толкования библиотеки участвуют в формировании множества понятий ПдО, функций интерпретации, аксиом и ограничений при разработке онтологии ПдО.

Ключевые слова: архитектура, онтология, предметная область, интегрированная информационная технология, предметная дисциплина, электронный курс, SQLite, UML.

Введение

Методология проектирования онтологии ПдО предполагает формирование множеств концептов, отношений, функций интерпретации и аксиом. Построение указанных множеств вручную является трудоемким процессом, как по времени, так и по количеству вовлеченных в процесс проектирования высококвалифицированных специалистов. Ручное проектирование онтологий мало чем отличается от проектирования экспертных систем [1]. Понимание важности создания инструментальных средств поддержки процесса проектирования онтологии заданной ПдО пришло практически одновременно с принятием парадигмы компьютерных онтологий. В настоящее время известно более ста инструментальных программных систем или онторедкторов с широким спектром характеристик и функциональных возможностей.

Постановка задачи

В работе рассматривается оригинальный Инструментальный комплекс онтологического назначения (ИКОН), основным назначением которого является реализация интегрированной информационной технологии автоматизированного построения онтологии в произвольной предметной области. Он выполняет анализ и обработку больших объемов неструктурированных данных, в частности лингвистических корпусов текстов на украинском и/или русском языке, извлечение из них предметных знаний с последующим их представлением в виде системно-онтологической структуры или онтологии предметной области [1]. На рис.1 представлена архитектурно-структурная организация ИКОН, на котором затемненный блок *“Внешний информационный ресурс – вид 1”* (ИР 1) (точнее, его программная модель) представляет предмет исследования в данной работе. Он создается на основе первоисточников, предшествуя процессу автоматизированного построения онтологии ПдО (электронные коллекции (ЭлК) энциклопедических словарей, толковых словарей и тезаурусов).

Решение задачи

Электронные коллекции энциклопедических и толковых словарей, а также тезаурусов являются (по определению) общезначимыми ресурсами знаний в заданном домене прикладных областей. Они содержат, как правило, несколько определений для каждого понятия с учетом смысловых оттенков и ориентированных на широкий круг потребителей информации. Такое представление знаний соответствует основным принципам построения формальной онтологии, а их смысловая интерпретация является основным источником формирования онтологии ПдО для инженера по знаниям.

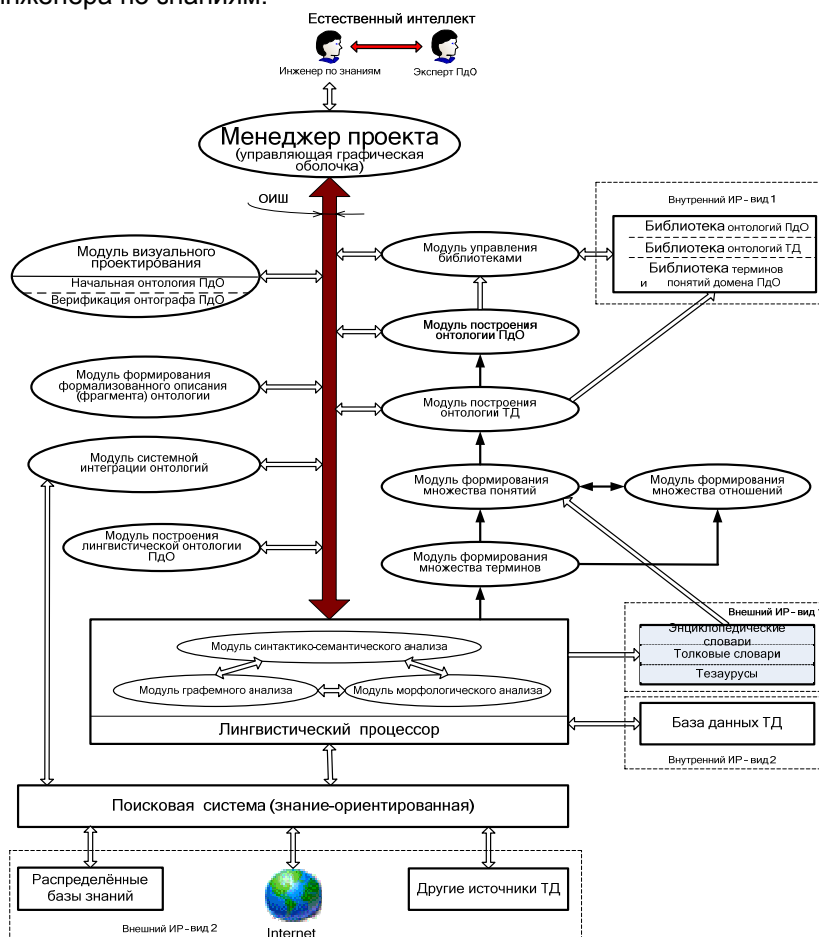


Рис. 1. Архитектурно-структурная организация ИКОН

Онтология ПдО разрабатывается безотносительно к конкретной прикладной задаче и, поэтому, представляет собой некоторую обобщенную (с точки зрения иерархии понятий) и усредненную (с точки зрения множеств функций интерпретации понятий) понятийную структуру предметных знаний. С одной стороны (преимущество) онтология ПдО является открытой, общезначимой системно-онтологической структурой, доступной к пополнению новыми знаниями и/или адаптации к конкретному кругу задач. С другой стороны (недостаток) общезначимые структуры не являются оптимальными для конкретных прикладных задач и требуют приложения серьезных усилий разработчиков, чтобы стать таковыми. Следствием указанного выше преимущества формальной онтологии ПдО является ее многократное использование для разных наборов типовых задач. Другими словами, один раз построенная формальная онтология ПдО пригодна для решения произвольного набора задач. Когда же речь идет о поиске оптимальной онтологической системы, необходимо адаптировать онтологическую базу знаний к целевому применению (исключить не используемые цепочки понятий в онтографе, конкретизировать множества функций интерпретации и ограничить значения признаков). Например, при адаптации онтологической базы знаний к задаче автоматизации разработки электронных курсов (ЭК) по определенной предметной дисциплине (ПдД), в связи с ограничением времени освоения ЭК учебным планом, часть цепочек понятий будет опущена.

При этом справочная информация в виде энциклопедий и словарей играет важную роль при выполнении указанной оптимизации.

Предлагаемая программная модель ИР 1 выполняет следующие функции.

1. При запуске модуля цифровой библиотеки справочной информации (модуль цифровой БСИ) на экране появляется окно программы «Dictionary» (рис.2), в котором находится: строка меню

(MenuBar) – содержит подменю Файл, Правка, Вид, Помощь, выпадающее меню «Выбор информационного ресурса», выпадающее меню снабжено всплывающей подсказкой «Выбор информационного ресурса» (ComboBox – строка выбора словаря), из которого можно выбрать необходимый тематический словарь, области для отображения списка ключевых терминов (List) – окно «общий список». Также есть кнопка Поиск и к ней строка набора искомого текста. В дальнейшем для удобства навигации в окне программы будут добавлены пиктограммы основных команд и будет добавлено второе окно «описание термина».

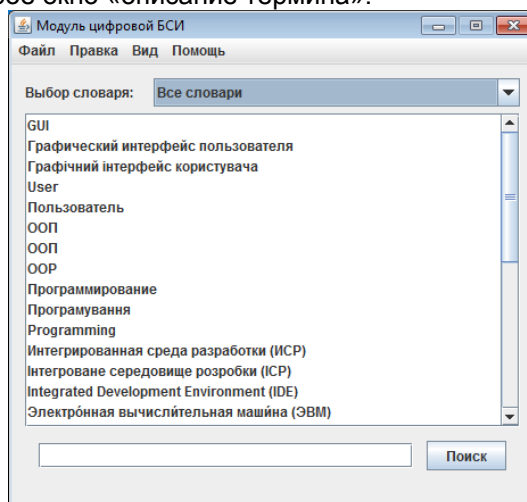


Рис. 2. Элементы графического интерфейса окна программного модуля «Цифровой БСИ»

2. После ввода термина в строку поиска и нажатия кнопки «Поиск» в окне «общий список» первой строкой будет искомое понятие.

3. Для выбора тематического словаря, необходимо нажать на выпадающее меню «Выбор словаря», откроется весь список имеющихся словарей, после это необходимо выбрать нужный словарь – в результате в окне «Общий список» отобразится содержимое выбранного словаря в алфавитном порядке.

4. Для получения описания термина нужно сделать двойной клик на нем, в результате откроется новое окно «Описание термина» в котором отобразиться словарное описание термина на русском и украинском языках.

5. Для реализации хранения и обработки информации со словарей используется SQLite – легковесная встраиваемая реляционная база данных. SQLite хранит всю базу данных (включая определения, таблицы, индексы и данные) в единственном стандартном файле с расширением *.sqlite. В программе используется бесплатная библиотека SQLiteJDBC-v056, которая позволяет работать с файлами баз данных данного типа.

Функциональная модель внешнего информационного ресурса.

Для проектирования функциональных моделей (функционального моделирования) программного обеспечения и различного рода информационных технологий известно ряд общепринятых стандартных методологий и языков функционального моделирования, таких как IDEF, DFD, UML.

UML (англ. Unified Modeling Language – унифицированный язык моделирования) – язык графического описания для объектного моделирования в области разработки программного обеспечения [2]. UML является широко используемым стандартом де-факто объектно-ориентированного визуального языка моделирования. Он разрабатывался с учетом преимуществ и недостатков SADT (IDEF) и DFD методов. В UML есть три основных вида моделей [3,4]:

- статическая модель (static model);
- динамическая модель (dynamic model).
- физическая модель (physical model)

Статическая модель описывает элементы системы и их взаимоотношения (классы, атрибуты, операторы). Одной из реализаций статической модели является диаграмма классов (class diagram). Динамическая модель описывает поведение системы, например, изменение программных сущностей (software entities) во время выполнения приложения. К динамическим моделям относятся: диаграмма прецедентов, также: вариантов использования, сценариев использования (use case diagram); диаграмма активности, также деятельности (activity diagram). Физическая модель отображает неизменяемую структуру программных сущностей, в частности файлов исходного кода, библиотек, исполняемых файлов, и отношений между ними.

В настоящее время разработаны средства визуального программирования на основе UML, обеспечивающие интеграцию, включая прямую и обратную генерацию кода программ, с наиболее распространенными языками и средами программирования, такими как Java, C++, C#, Visual Basic, Object Pascal/Delphi. Поскольку при разработке языка UML были приняты во внимание многие передовые идеи и методы, можно ожидать, что на очередные версии языка UML также окажут влияние и другие перспективные технологии и концепции. Кроме того, на основе языка UML могут быть определены многие новые перспективные методы. Язык UML может быть расширен без переопределения его ядра.

UML технологии стали основой для разработки и реализации во многих инструментальных средствах: в средствах визуального и имитационного моделирования, а также в CASE-средствах самого различного целевого назначения. Более того, заложенные в языке UML потенциальные возможности могут быть использованы не только для объектно-ориентированного моделирования систем, но и для представления знаний в интеллектуальных системах, которыми, по существу, являются перспективные сложные программно-технологические комплексы. Принимая во внимание все преимущества языка UML, на его основе была спроектирована функциональная модель ИНИТ.

Функциональная модель ИНИТ представляет собой набор диаграмм трёх видов:

- диаграмма вариантов использования (рис. 3);
- диаграмма активности (рис. 4);
- диаграмма классов (рис. 5).

Целями разработки диаграммы вариантов использования являются:

1. Определить общие границы и контекст моделируемой предметной области на начальных этапах проектирования модуля цифровой БСИ.
2. Сформулировать общие требования к функциональному поведению проектируемого модуля цифровой БСИ.
3. Разработать исходную концептуальную модель модуля цифровой БСИ для её последующей детализации в форме логических и физических моделей.

Суть данной диаграммы состоит в следующем: проектируемая система представляется в виде множества сущностей или акторов (actor), взаимодействующих с системой с помощью так называемых вариантов использования. При этом актором или действующим лицом называется любая сущность, взаимодействующая с системой извне. Это может быть человек, программа или любая другая система, которая может служить источником воздействия на моделируемую систему. В свою очередь, вариант использования служит для описания сервисов, которые система предоставляет актору.

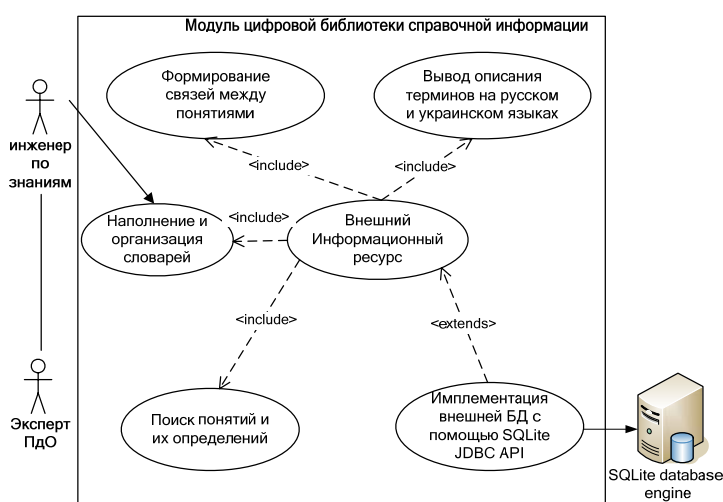


Рис. 3 Диаграмма вариантов использования внешнего ИР

С помощью диаграммы активности (рис. 4) можно изучать поведение модуля цифровой БСИ с использованием моделей потока данных и потока управления. Диаграмма активности отображает некоторый алгоритм, описывающий жизненный цикл объекта, состояния которого могут меняться.

Диаграмма активности в отличие от блок-схемы, имеет более широкую нотацию. Например, на ней можно указывать состояния объектов.

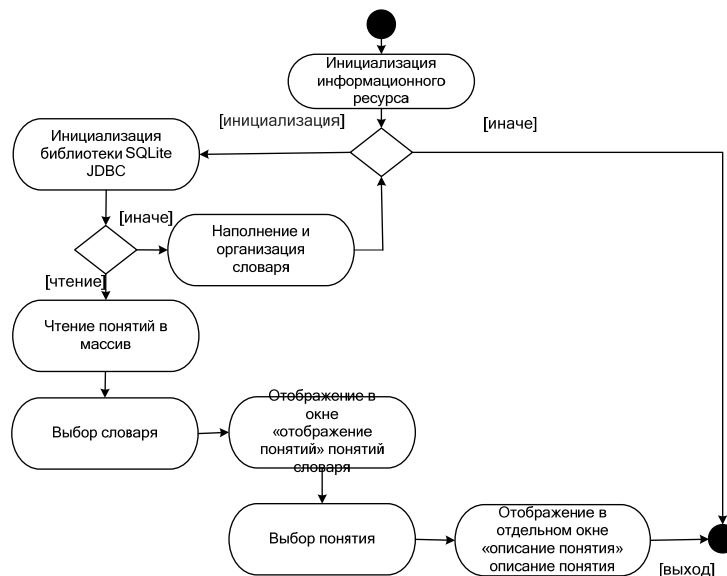


Рис. 4. UML-диаграмма активности программной модели ИР

Диаграмма классов описывает структуру объектов программной модели информационного ресурса: их индивидуальность, отношения с другими объектами, атрибуты, функции и процедуры. Модель классов создает контекст для диаграмм состояний и взаимодействия. На рис. 5 представлен фрагмент UML-диаграмма классов программной модели ИР:

- класс **ArchiveSearch** реализует технологию обработки и управления хранилищами данных и знаний;
- класс **GraphEditor** реализует технологию формально-логического представления и интеграции онтологических структур;
- класс **ControlGraphicShell** реализует управляющую графическую оболочку, которая осуществляет общее управление процессом реализации связанных информационных технологий;
- интерфейс для доступа к внешней библиотеке **KonspektLib**, реализующей функции лингвистического анализа.

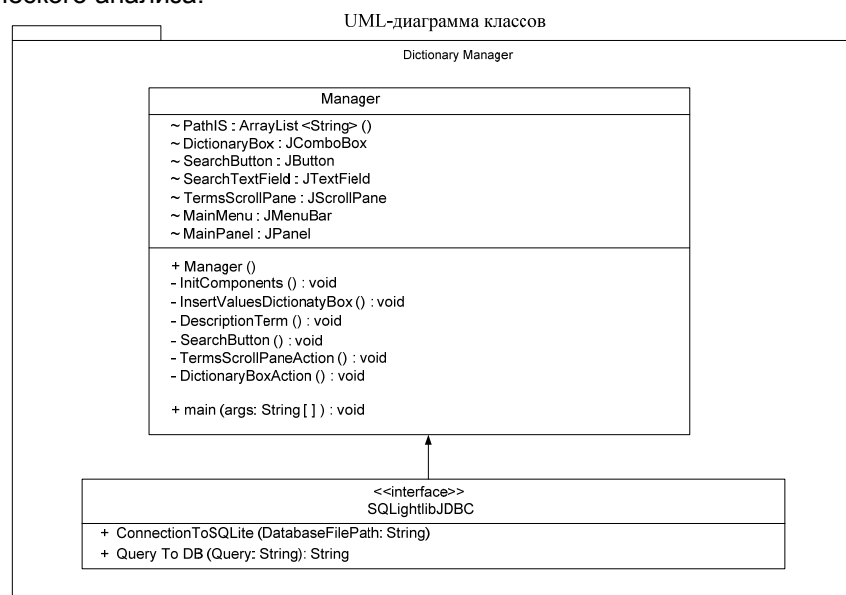


Рис. 5 Фрагмент UML-диаграммы классов программной модели ИР.

Программная реализация программного модуля «Библиотека словарей» ИКОН выполнена с использованием платформы Java Swing Framework. В отличие от других платформ, она не только предоставляет интерфейс разработки на основе шаблона MVC, но и сама реализована на его основе. Представлением является класс – наследник класса Frame. Вследствие организации

событийной модели Java [6] на интерфейсах, контроллер представляет собой набор анонимных классов обработки соответствующих событий. Как и остальные платформы, Swing предоставляет разработку модели программисту.

Для хранения и работы с базой терминов и их описанием использовалась SQLite [7] – легковесная встраиваемая реляционная база данных. Исходный код библиотеки является открытым, то есть она является бесплатной. SQLite не использует парадигму клиент-сервер, то есть движок SQLite не является отдельно работающим процессом, с которым взаимодействует программа, а предоставляет библиотеку, с которой программа компонуется и движок становится составной частью программы. Таким образом, в качестве протокола обмена используются вызовы функций (API) библиотеки SQLite. Такой подход уменьшает накладные расходы, время отклика и упрощает программу. SQLite хранит всю базу данных (включая определения, таблицы, индексы и данные) в единственном стандартном файле на том компьютере, на котором исполняется программа. Именно в этом файле и хранятся все термины и их описание.

Выводы

В работе предложена программная модель цифровой библиотеки справочной информации, входящей в оригинальный инструментальный комплекс онтологического назначения (ИКОН). Определения и толкования библиотеки участвуют в формировании онтологии ПДО. Приведены UML-диаграммы, описывающие функциональную модель программного модуля «Библиотека словарей» ИКОН.

Электронные библиотеки являются общезначимыми ресурсами знаний в заданном домене близких прикладных областей и в этом смысле они инвариантны при адаптации и оптимизации онтологической системы к целевому применению, в частности, при адаптации к задаче автоматизации разработки электронных курсов (ЭК) по определенной предметной дисциплине (ПДД).

Электронные библиотеки играют важную роль при выполнении указанной адаптации.

Предлагаемая программная модель выполняет и обработку больших объемов неструктурированных данных, в частности лингвистических корпусов текстов на украинском и/или русском языке, являющихся внешним информационным ресурсом, взаимодействующим с лингвистическим процессором в архитектурно-структурной организации ИКОН.

Литература

1. *Палагин А.В.* Онтологические методы и средства обработки предметных знаний / А.В. Палагин, С.Л. Крытый, Н.Г. Петренко. – [монография] – Луганск: изд-во ВНУ им. В. Даля, 2012. – 323 с.
2. *Палагин А.В.* К вопросу разработки онтолого-управляемой архитектуры интеллектуальной программной системы / А.В. Палагин, Н.Г. Петренко, В.Ю. Величко К.С. Малахов – Научный журнал «Вісник Східноукраїнського національного університету імені Володимира Даля» – №13 (167) с. 179-184 – Луганск: изд-во ВНУ им. В. Даля, 2011. – 343 с.
3. *C. Marshall, Enterprise Modelling with UML, ISBN 0-201- 43313-3, Addison-Wesley, Reading, MA, 2000.*
4. *G. Booch, J. Rumbaugh, I. Jacobson, The Unified Modeling Language User Guide, Addison-Wesley, Reading, MA, 1999.*
5. *T. Quatrani, Visual Modeling with Rational Rose and UML, Addison-Wesley, Reading, MA, 1998.*
6. *Java Standard Edition* Режим доступа: <http://www.oracle.com/technetwork/java/javase/downloads/index.html> – Дата доступа: 11.05.2012
7. *SQLite database engine.* Режим доступа: <http://www.sqlite.org/> . – Дата доступа: 11.05.2012

О.В. Палагін, М.Г. Петренко, В.Ю. Величко, К.С. Малахов, В.В. Семенков, Ю.Л. Тихонов Програмні моделі ІКОН: бібліотека словників ПДО.

У статті описана програмна модель цифрової бібліотеки довідкової інформації, що включає енциклопедії, тлумачні словники і тезауруси домену близьких предметних областей. Визначення і тлумачення бібліотеки беруть участь у формуванні безлічі понять ПДО, функцій інтерпретації, аксіом і обмежень при розробці онтології ПДО.

Ключові слова: онтологія, предметна область, предметна дисципліна, електронний курс.

**A.V.Palagin, N.G. Petrenko, V.Yu. Velichko, K.S. Malakhov, Yu.L. Tikhonov
Programming model of Instrumental complex of ontological purposes: the library of
dictionaries domains.**

The article describes the programming model of digital library reference, including encyclopedias, dictionaries and thesauruses domain similar domains. Definitions and interpretations of the library involved in the

formation of multiple concepts of a PDO, the interpretation of the functions, axioms and constraints in the development of ontologies PDO.

Keywords: ontology, subject area, subject discipline, the electronic course.