(cc) BY

Ильин Борис Васильевич
канд. техн. наук, доцент
ФГБОУ ВО «Керченский государственный
морской технологический университет»
г. Керчь, Республика Крым

МОДЕЛИРОВАНИЕ ПРЕДМЕТНЫХ ОБЛАСТЕЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Аннотация*: в статье описывается подход к формированию инфологической модели предметной области при проектировании базы данных с использованием средств естественного языка. Полученная в результате модель является однородно структурированным описанием предметной области, объективно отражающим её семантику, независимым от способа последующего отображения на пространство памяти ЭВМ.*

Ключевые слова*: инфологическое моделирование, предметная область, база данных, проектирование, отображение.*

Повышение уровня интеллектуальности автоматизированных информационных систем (АИС) напрямую зависит от наличия в них знаний о соответствующей предметной области. Под предметной областью (ПО) здесь понимается фрагмент реального мира, информация о предметных реалиях которого (объектах, их свойствах, отношениях между ними) подлежит хранению в базе данных (БД).

Поскольку основное назначение БД – быть достоверным отображением ПО, то особую актуальность приобретает проблема разработки подходов, методов и средств формирования описания ПО в АИС различного назначения. Работы в этом направлении интенсивно развиваются в рамках исследований по представлению знаний в системах искусственного интеллекта. Однако в большинстве таких подходов преобладает прагматически-технологический аспект, связанный с использованием для отражения ПО известных моделей данных, зачастую не отвечающих требованиям естественности и навязывающих пользователям не

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

свойственные им представления. Вводятся формальные языки представления знаний, не решающие гносеологических проблем описания ПО, но создающие дополнительные трудности пользователю – непрофессионалу в области программирования.

В то же время, как отмечает ряд исследователей [1; 2], давно известной моделирующей системой, средствами которой можно описать все многообразие мира, является естественный язык (ЕЯ). Поэтому основной целью настоящей работы является изложение подхода к формированию описания ПО, обеспечивающего адекватное ее отображение с использованием средств выражения, естественных для пользователя, технология обработки которых на ЭВМ предложена в [1]. Используя терминологию основоположников теории моделирования предметных областей [4; 5], отнесём формируемое описание к классу инфологических моделей ПО.

Целью инфологического моделирования является создание структурированной информационной модели ПО, отображаемой посредством данных в БД. Инфологическая модель должна соответствовать таким требованиям:

– адекватное отображение моделируемой ПО;

– использование для своего описания языковых средств, понятных будущим пользователям АИС и проектировщикам БД.

В работе [3] показано, что наиболее общей и непротиворечивой логической системой, необходимой и достаточной для описания самых общих структурных особенностей любых предметных областей, является тернарная система, основанная на триаде категорий «объект» – S, «свойство» – C, «отношение» – R, где R определено на S и C, т.е. R = <SRS, SRC, CRC>. Поэтому можно сказать, что суть инфологического моделирования состоит в выделении объектов ПО, установлении их характеристик и взаимосвязей между ними.

Описывая ПО, проектировщик БД имеет дело с сущностями различной природы: а) наблюдаемыми в ПО объектами S, их свойствами C и отношениями R между ними, б) отображающими их понятиями в процессе мыслительной деятельности, в) соответствующими им именами. Поэтому моделирование ПО как

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

процесс, неотделимый от речемыслительной деятельности человека, выполняется в виде последовательности следующих этапов отражения (рис. 1):

– «ПО» → «мысленная модель ПО»;

– «мысленная модель ПО» → «языковая модель ПО»;

– «языковая модель ПО» → «письменная модель ПО».

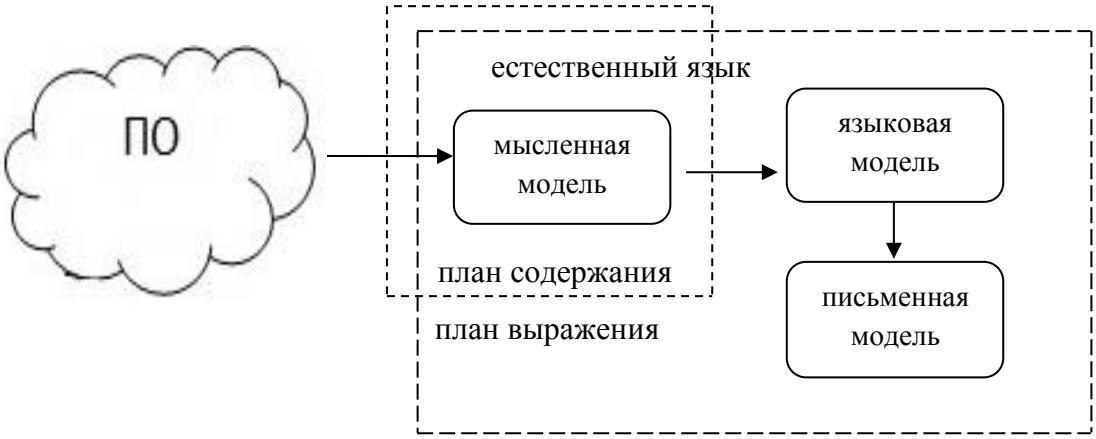

Рис. 1. Этапность отображения ПО

На первом этапе реализуется отображение «ПО» – «мысленная модель ПО», где под мысленной моделью понимается отображение описываемой ПО в мышлении человека с использованием определенной системы понятий (*план содержания*).

Следующим этапом является отображение «мысленная модель» – «языковая модель», поскольку материальное воплощение первой требует словесного выражения. На данном этапе решается проблема выбора языка, в качестве которого предлагается использовать грамматические и семантические средства естественного языка, необходимые для описания рассматриваемой ПО.

В качестве предметной основы «языковой модели» используется привычная форма в виде текста на естественном языке, т.е. на данном этапе осуществляется отображение «языковая модель» – «письменная модель» (различаясь с операционной точки зрения, эти модели адекватны по смысловой сути).

Использование абстракции отождествления позволяет практически неисчислимое множество индивидуальных объектов S, выделенных в ПО, привести к конечному числу классов объектов K и сделать тем самым обозримым их

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

многообразие, Заметим, что задание класса, вообще говоря, может осуществляться двояко: перечислением или указанием эталона.

Свойство C^i, определенное на классе объектов K_j, можно рассматривать как класс с именем IC^i_j, членами которого являются значения, принимаемые свойством C^i на членах s_p класса объектов K_j ($s_p \in K_j$).

Под отношением здесь понимается форма связи, выявленная на основе сопоставления заданным образом нескольких сущностей. Поскольку наблюдаемые в ПО отношения между классами объектов и свойств можно свести к совокупности бинарных отношений, в формируемой модели ПО используются именно эти отношения.

Таким образом, если в БД представляются сведения о каждом конкретном объекте $s_K \in S$ (значения свойств и взаимосвязи с конкретными объектами), то в инфологической модели знания о ПО представляются на уровне отображения сведений о классах.

Результат конкретного акта отображения ПО на уровне *выражения* в языковой сфере фиксируется в виде элементарного высказывания, формальным выражением которого выбрана известная структура «отображающего элемента» (ОЭ): $b=Ig_i$, r, Ig_j, q, t. Здесь Ig_i и Ig_j – имена классов (объектов, свойств), r – имя отношения, q – значение, t – обозначение момента времени, в который справедливо высказывание, выраженное ОЭ.

Использование ОЭ позволяет, в общем случае, идентифицировать связи между классами объектов, обозначенными данными, самими данными и между элементами памяти, в связи с чем устраняется необходимость разработки отдельных языков для описания логических и физических структур БД, а также получить простое, однородно структурированное описание ПО, полностью сохраняющее ее семантику.

Основываясь на том положении, что описание ПО представляет собой совокупность всех суждений о ней, это описание строится как совокупность ОЭ, содержательно состоящая из двух частей: объектной и связной. В объектной части перечисляются все выделенные в ПО классы (объекты и свойства) и задаются

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

составы классов, а в связной части указываются взаимосвязи классов, реализуемые бинарными отношениями. Причём фиксированию подлежат как прямые, так и обратные (в случае необходимости) отношения, что обеспечивает логическую полноту представления всех отношений, наблюдаемых в ПО, а также симметричность БД для всех возможных приложений.

Сформированное таким образом описание ПО может быть формализовано следующим образом. Пусть G = <K, C> – множество выделенных в ПО классов объектов K и свойств C, и на нём задано отображение D: $G \rightarrow I_G$, где I_G – множество имен классов (объектов и свойств). $I_R = \langle r_k \rangle$ – множество имен бинарных отношений, определенных на $G_i * G_j$. Каждому из имен $r_k \in I_R$ ставится в соответствие двухместный предикат P_{rk}: $G_i * G_j$, имеющий значение «истина», если справедливо высказывание $G_i\ r_k\ G_j$, и «ложь» – в противном случае. Задав таким образом на множестве классов G предикаты, соответствующие всем именам $r_k \in I_R$, получим описание ПО в виде знаковой модели

$$M_{по} = \langle I_G; P_{r1}, P_{r2}, P_{r3}, \ldots, P_{rm} \rangle,$$

где P_{rk} – истинные предикаты, а m – число элементов множества.

Полученное в результате инфологического моделирования описание ПО является однородно структурированным, объективно отражает его семантику, использует для своего представления средства естественного языка и независимо от способа последующего отображения на пространство памяти ЭВМ.

Список литературы

1. Зайцев Н.Г. Технология обработки данных в языковой форме / Н.Г. Зайцев. – Киев: Техника, 1989. – 180 с.

2. Попов Э.В. Общение с ЭВМ на естественном языке / Э.В. Попов. – М.: Наука, 1982. – 360 с.

3. Уёмов А.И. Вещи, свойства и отношения / А.И. Уёмов. – М.: Изд-во АН СССР, 1963. – 230 с.

4. Langefors B. Infological model and information user views / B. Langefors // Inform. Systems. – 1980. – №5. – P. 17–32.

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

5. Sundgren B. An infological approach to data bases / B. Sundgren. – Stockholm: National Central Bureau of Statistics. 1973. – 294 p.

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)