

**Ильин Борис Васильевич**

канд. техн. наук, доцент

ФГБОУ ВО «Керченский государственный  
морской технологический университет»

г. Керчь, Республика Крым

## **ПОДХОД К МОДЕЛИРОВАНИЮ ПРЕДМЕТНОЙ ОБЛАСТИ В ИНФОРМАЦИОННЫХ СИСТЕМАХ**

*Аннотация:* в статье рассматривается подход к формированию информационной модели предметной области при проектировании информационной системы с применением средств естественного языка. Предложенная модель использует для своего представления логику предикатов. Повышение семантической выразительности информационной модели реализуется за счёт введения в её состав многоуровневых понятийных конструкций.

*Ключевые слова:* предметная область, информационное моделирование, семантика.

Важным фактором успеха в различных профессиональных сферах деятельности является наличие в организации эффективно функционирующей информационной системы (ИС), основанной на применении технологии баз данных (БД). Обязательной компонентой любой ИС является представленная в ней в том или ином виде информационная модель предметной области (ПО). В силу этого современные технологии проектирования ИС всё в большей степени основываются на использовании методологии информационного моделирования ПО [4; 5].

Следует отметить, что общепризнанное формальное определение понятия «предметная область» в настоящее время отсутствует. Принято считать, что оно не может быть формализовано как первичное понятие [6]. Тем не менее, анализ множества существующих трактовок этого понятия делает его содержание интуитивно понятным. В рамках настоящего исследования ПО рассматривается как мысленно выделенный по определённым соображениям фрагмент реально-

го мира, информация о предметных реалиях которого (объектах, их свойствах и отношениях между ними) подлежит хранению в БД информационной системы [3].

Под информационной моделью ПО будем понимать некоторую систему, адекватно отображающую структуру (и/или функционирование) выделенной ПО. Такая модель должна, очевидно, обладать достаточной степенью формализации, обеспечивающей однозначное описание состава и структуры, соответствующей ПО.

Целью настоящей работы является изложение подхода к формированию информационной модели ПО, обеспечивающей адекватное её отображение в виде структурированных текстов на естественном языке, компьютерная технология обработки которых предложена в [2].

В [3] отмечается, что информационное моделирование ПО как процесс, неотделимый от речемыслительной деятельности человека, можно рассматривать в виде последовательности следующих этапов отражения:

1. ПО → «мысленная» модель ПО.
2. «мысленная» модель ПО → «языковая» модель ПО.
3. «языковая» модель ПО → «письменная» модель ПО.

Вследствие этого, моделируя ПО, исследователь имеет дело с реалиями различной природы:

- а) наблюдаемыми в ПО сущностями (объектами, процессами), их свойствами и отношениями между ними;
- б) отображающими их в процессе мыслительной деятельности понятиями;
- в) соответствующими понятиям именами в языковой сфере.

В рамках рассматриваемого подхода к моделированию ПО в качестве категориального базиса используется наиболее общая и непротиворечивая логическая система, необходимая и достаточная для описания любых предметных областей [3], основанная на триаде категорий: «объект»- S, «свойство» – C, «отношение» – R, где R определено на S и C.

Использование абстракции отождествления позволяет множество индивидуальных объектов  $s$ , выделенных в ПО, привести к конечному числу классов объектов  $K$  и сделать тем самым обозримым их многообразие. Под классом здесь понимается абстракция множества однородных объектов ПО, обладающих одинаковым набором характеризующих их свойств. Индивидуальный объект соответствующего класса в этом случае рассматривается как «элемент класса».

Свойство  $C_j^i$ , определенное на классе объектов  $K_i$ , можно рассматривать как класс, элементами которого являются значения, принимаемые свойством  $C_j$  на элементах  $s_p$  класса объектов  $K_i$  ( $s_p \in K_i$ ).

Объекты, выделяемые в ПО, связываются различными отношениями, наиболее значимые из которых должны быть отражены в формируемой модели ПО. Под отношением понимается некоторая ассоциация между двумя или более объектами (или их свойствами). А поскольку отношения, наблюдаемые между объектами и свойствами в ПО, можно свести к совокупности бинарных отношений [6], то в формируемой модели ПО используются отношения именно этого типа. При моделировании ПО осуществляется переход от отношений, наблюдаемых в ПО между индивидуальными объектами и их свойствами, к отношениям между соответствующими классами  $K_i$  и  $C_j^i$ . При этом следует учитывать как прямые, так и обратные (в случае необходимости) отношения, что обеспечивает логическую полноту представления отношений, наблюдаемых в ПО.

Из изложенного выше следует, что если в БД представляются сведения о каждом индивидуальном объекте  $s_p \in K_i$ , то в информационной модели ПО знания о ней представляются на уровне представления сведений о соответствующих классах объектов  $K_i$ .

В качестве средства выражения модели ПО предлагается использовать естественный язык (ЕЯ), поскольку он является универсальной моделирующей системой [2]. Информационная модель ПО в этом случае может быть представлена в текстовом виде множеством предложений ЕЯ, каждое из которых вы-

строено по определённому шаблону. Формальным выражением такого шаблона выбран «отображающий элемент» (ОЭ) [1], имеющий в общем виде следующую структуру:  $b = G_i, r_p, G_j, q, t$ . Здесь  $G_i$  и  $G_j$  – имена классов (объектов или свойств),  $r_p$  – имя связывающего отношения,  $q$  – значение,  $t$  – обозначение момента (или периода времени), в который справедливо высказывание, выраженное соответствующим ОЭ. В результате получим информационную модель ПО, являющуюся, по сути, множеством всех суждений о ней на ЕЯ, представленную как совокупность ОЭ.

В [3] автором был предложен следующий подход к формализации представления информационной модели ПО. Пусть  $G = \langle KUC \rangle$  – множество выделенных в ПО классов (объектов  $K$ , свойств  $C$ ) и задано отображение  $D : G \rightarrow I_G$ , где  $I_G$  – множество имён классов  $G$ , а  $I_R = \langle r_k \rangle$  – множество имён бинарных отношений, определённых на  $G \times G$ . Каждому из имён  $r_k \in I_R$  ставится в соответствие двухместный предикат  $Pr_k(G_i, G_j) \rightarrow \{И, Л\}$  так, что для любой пары  $\langle G_i, G_j \rangle$  предикат  $Pr_k(G_i, G_j)$  имеет значение «И» (истина), если справедливо высказывание  $G_i r_k G_j$ , и значение «Л» (ложь) в противном случае. Задав таким образом на множестве классов  $G$  предикаты, соответствующие всем  $r_k \in I_R$ , получим представление модели ПО в виде знаковой модели

$$M_{по} = \langle I_G, Pr_1, Pr_2, Pr_3, \dots, Pr_m \rangle,$$

где  $Pr_k$  – истинные предикаты, а  $m$  – количество элементов множества  $I_R$ .

Использование логики предикатов для информационного моделирования ПО позволяет применить хорошо развитый и понятный математический аппарат. Само же представление модели при этом является однородно структурированным, достаточно формализованным, объективно отражающим семантику описываемой ПО.

Однако следует признать, что представление модели ПО в таком виде является в ряде случаев излишне дробным и, в то же время, семантически «обеднённым», поскольку представляет единственную точку зрения на состав и структуру ПО без учета представлений о ней потенциальных различных групп пользователей с разными информационными интересами.

С целью повышения семантической выразительности моделей ПО в настоящее время значительное внимание уделяется разработке методов и средств формирования многоуровневых понятийных структур, включаемых в состав моделей. Основным методом, применяемым для этого, является метод обобщающей абстракции, позволяющий формировать понятийные конструкции, производные по отношению к классам конкретных объектов. Получаемая в результате абстракция рассматривается как класс, обобщающий в себе общие признаки всех элементов выделенных классов.

Так, например, обобщением понятий *студент*, *курсант*, *магистрант*, *аспирант* по признаку участия в учебном процессе может быть понятие *обучающийся*. Соответственно, в информационную модель ПО вводится класс *обучающийся*, связываемый отношением «род-вид» с классами, имена которых совпадают с именами выше названных понятий. При этом одноимённые свойства классов «видового» уровня становятся свойствами класса «родового» уровня.

Выстраивание понятий в иерархию «род-вид» и поддержка наследования свойств по этой иерархии «сверху-вниз» позволяют повысить семантическую выразительность модели, отображаемой ПО и обеспечить, в то же время, уменьшение объёма её представления.

Необходимо отметить, что один и тот же класс объектов  $K_i$  может связываться отношением «вид-род» с несколькими классами при использовании различных аспектов обобщения. Так, например, класс *аспирант* может быть связан отношением «вид-род» с классом *преподаватель* по признаку выполнения преподавательской нагрузки (поскольку аспирант при определённых условиях может привлекаться к проведению занятий в качестве преподавателя).

Поэтому, в целях сохранения семантики, моделируемой ПО необходимо в имени «видородового» отношения указывать имя признака, послужившего аспектом обобщения. В этом случае можно считать, что класс *аспирант* связывается с классом *обучающийся* отношением «вид-род в аспекте быть обучаемым», а с классом *преподаватель* – отношением «вид-род в аспекте быть обучающим». Соответственно сформированные в результате применения операции

обобщения надклассы будут связываться с классами нижележащего уровня отношениями типа «род-вид» с указанием признака разбиения на подклассы, что позволяет реализовать возможности фасетной классификации.

Известно, что удобным средством, обеспечивающим механизм для обобщения и реализующим один из традиционных методов фиксации знания, является структурирование этого знания в иерархическую классификационную систему. А поскольку одна и та же ПО может рассматриваться разными группами пользователей информационной системы с разных точек зрения, это должно найти свое отражение в разработке различных схем классификации, построенных над выделенными в ПО подмножествами классов.

Отсюда следует, что повышение семантической выразительности информационной модели ПО может быть обеспечено путем включения в её состав классификационных схем, задающих иерархии классов различных уровней разбиения. А поскольку любая классификационная схема представляет собой совокупность классов, связанных родовидовыми отношениями, она естественным образом описывается с помощью ОЭ и реализуется в предложенной выше знаковой модели ПО, использующей для своего представления логику предикатов.

### *Список литературы*

1. Зайцев Н.Г. Принципы информационного обеспечения в системах переработки информации управления / Н.Г. Зайцев. – Киев: Наук. думка., 1976. – 182 с.
2. Зайцев Н.Г. Технология обработки данных в языковой форме / Н.Г. Зайцев. – Киев : Техника, 1989. – 180 с.
3. Ильин Б.В. Моделирование предметных областей в информационных системах / Б.В. Ильин //Актуальные направления научных исследований: перспективы развития: материалы XIII Междунар. науч.-практ. конф. – Чебоксары: ЦНС «Интерактив плюс», 2020. – С. 125–128.
4. Никаноров С.П. Концептуализация предметных областей. Серия «Концептуальный анализ и проектирование» Методология и технология / С.П. Никаноров. – М.: Концепт, 2009. – 268 с.

5. Соснин П.И. Концептуальное проектирование систем: учеб. пособие / П. И. Соснин. – Ульяновск: УлГТУ, 2017. – 198 с.

6. Цаленко М.Ш. Моделирование семантики в базах данных / М.Ш. Цаленко. – М.: Наука. – 1989 – 287 с.