

Д.К.Пархоменко, О.С.Мезенцева (Ставрополь, Сев-КавГТУ).
ORM-подход для обработки полуструктурированных данных.

Современные требования к уровню информатизации делают актуальной задачу эффективной обработки полуструктурированных данных в составе реляционных хранилищ корпоративных приложений, при разработке которых используют методы ООП и реляционной теории. Возможность прозрачной совместной работы этих технологий отсутствует ввиду объектно-реляционного рассогласования интерфейсов. Для преодоления этой проблемы существует ряд методов и решений, объединенных аббревиатурой ORM (Object-Relational Mapping). В рамках настоящего доклада предлагается формальное описание методов ORM для обработки полуструктурированных данных в реляционном виде.

Для формализации процесса обработки полуструктурированных данных автором была предложена алгебра ациклических ориентированных графов, состоящая из 5 операций: выборки ($\prec_{\nu_B, F} A$), объединения ($A \cup B$), слияния (A в $T_{a,b} B$), модификации ($A \mapsto B$), удаления ($A - B$) [1, 2]. Согласно определению алгебры, данные представляются в виде ациклических ориентированных графов; такое представление будем называть *концептуальным*. Представление данных в виде графа объектов абстрактных типов данных (АТД) будем называть *программно-зависимым*.

Таблица. Определения используемых множеств

Множество	Описание
T_{class}	множество АТД сущностей предметной области (ПО)
P_{name}	множество имен свойств (properties) АТД
$R_{name} = \{n_1, n_2, \dots, n_m\}$	множество имен ссылок на объекты АТД в памяти
P_{val}	множество значений свойств АТД
T_{sim}	множество простых типов данных языка
$P = \{(n, v) n \in P_{name}, v \in P_{val}\}$	множество свойств объектов
$C = \{(e, c) c \in T_{class}, e \in E\}$	множество связей АТД с сущностями ПО, где $c = (name, pnames, rnames)$
O	множество объектов АТД в памяти
E, A	множества сущностей и атрибутов сущностей ПО

Определим следующие функции: $Map_e: T_{class} \rightarrow E$, $Map_a: P_{name} \rightarrow A$, $Map_p: A \rightarrow P_{name}$, $Map_c: E \rightarrow T_{class}$, $GetClass: O \rightarrow T_{class}$, $Createobect: T_{class} \rightarrow O$, $Link: O \times O \rightarrow R_{obj}$, где $R \subseteq O \times O$.

Метод 1. Пусть граф $K'(O, R)$ — программно-зависимое представление данных. Метод преобразования его в граф концептуального представления $K(V, L)$ определяется следующей последовательностью операций.

1. Сформировать множество вершин $V_i := \{v_i, \dots\}$, $\forall o_i \in O$, $1 \leq i \leq |O|$; $v_i.key := o_i.key$, где $v_i.e := Map_e(GetClass(o_i))$, $v_i.values := \{(a, value) | (n, v) \in o_i.values \wedge a = Map_a(n) \wedge value = v\}$.

2. Сформировать подграфы $K_i = (\{v\}, \emptyset)$ в $\{v, v'\}$ $(\{v'\}, \emptyset)$, где $v \in V_i$ и $v' \in V_i$ такие, что $o.key = v.key$ и $o'.key = v'.key \forall r = (o, o') \in R_{obj}$; $1 \leq i \leq |R_{obj}|$.

3. Объединить все получившиеся подграфы $K = \cup K_i$.

Метод 2. Пусть граф $K(V, L)$ — концептуальное представление данных. Метод преобразования его в программно-зависимое представление $K'(O, R)$ определяется следующей последовательностью операций.

1. Для каждой вершины $(\forall)v_i \in V$:

— если $\exists(c, e) \in C$; $v_i.e = e$, то $class = C$ иначе $class = Map_c(v_i.e)$;

— создать $o_i = CreateObject(class)$; $O := O \cup \{o_i\}$;

— инициализировать множество значений $o_i.values := \{(n, v) | (a, value) := v_i.values \wedge n := Map_p(a) \wedge v := value\}$, $o_i.key = v_i.key$.

2. Для каждой дуги $(\forall)(v, v') \in L$ создать $o \in O \wedge o.key = v.key$, $o' \in O \wedge o'.key = v'.key$, $R := R \cup \{Link(o, o')\}$.

Программная реализация представленных методов позволит получить универсальный программный компонент взаимного преобразования полуструктурированных данных системы, представленных в виде ациклических орграфов, в объектно-ориентированные структуры.

Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы по проблеме «Разработка теоретических основ функционирования систем управления полуструктурированными данными».

СПИСОК ЛИТЕРАТУРЫ

1. *Пархоменко Д. К.* Предпосылки введения алгебры для формализации операций над иерархическими данными. — В сб.: Международная научная конференция «Актуальные проблемы и инновации в экономике, управлении, образовании и информационных технологиях». Ставрополь–Кисловодск: 2009.
2. *Пархоменко Д. К.* Алгебра орграфов для моделирования иерархических данных. Ставрополь: Сев-КавГТУ, 2009, 21 с. Деп. в ВИНТИ 15.10.09, № 627-В 2009.