

Н.В. Лукиных, Л.Н. Лядова

Национальный исследовательский университет  
«Высшая школа экономики»  
(Пермский филиал)

LLyadova@hse.ru

## **О ПОДХОДЕ К РЕАЛИЗАЦИИ СРЕДСТВ ВЕДЕНИЯ НОРМАТИВНО-СПРАВОЧНОЙ ИНФОРМАЦИИ**

### **Введение**

Деятельность любого предприятия, вне зависимости от отраслевой специфики, связана с постоянным обращением к нормативно-справочной информации (НСИ). По мере роста и развития бизнес-системы количество такой информации быстро увеличивается. Данная информация поступает как из внешних источников (ГОСТы, справочники и классификаторы отраслевых министерств, ведомств), так и порождается внутри подразделений самой компании (бизнес-системы). У крупной или средней компании, в состав которой входит несколько предприятий, филиалов, подразделений, неизбежно возникает потребность в интеграции данных, которые поступают из разных источников [1, 3]. Эта задача может быть решена только на основе использования единых классификаторов, справочников и стандартизации способов представления и кодирования данных во взаимодействующих системах, обменивающихся информацией [13].

В соответствии с требованиями, предъявляемыми к проектам современных интегрированных автоматизированных информационных систем (ИС), необходимо обеспечить *однократность ввода данных* в рамках отрасли и *возможность интеграции* ИС различных уровней, принадлежащих различным учреждениям и организациям (бизнес-системам), в единую ИС отрасли. Одним из элементов, определяющих свойства системы, является реализация в ней установленных для соот-

ветствующей отрасли стандартов представления и кодирования информации [2, 3].

*Универсальную систему управления НСИ* следует рассматривать как совокупность методик ведения и поддержки НСИ, а также ряда технологических решений, обеспечивающих решение задач создания единого информационного пространства предприятия. Посредством такой системы должны быть обеспечены не только возможность интеграции систем на основе централизованного ведения НСИ, но могут также быть реализованы средства анализа данных, полученных из различных источников.

Средства управления НСИ должны не только обеспечивать работу программных компонентов ИС при выполнении пользователями операций автоматизируемых бизнес-процессов, но и предоставлять в распоряжение пользователей удобный пользовательский интерфейс, позволяющий создавать и модифицировать ресурсы, быстро искать необходимую информацию в соответствии с заданными критериями. При этом должна быть обеспечена возможность работы с классификаторами, имеющими различную структуру, использующими различные алгоритмы обработки информации, импорта классификаторов из внешних источников.

Средства управления нормативно-справочной информацией CASE-системы METAS предназначены для стандартизации представления и кодирования информации в информационных системах, унификации средств работы с нормативно-справочной информацией, в частности, с классификаторами различных видов. В основе реализации этих средств – разработанная авторами математическая модель [6, 16].

Описываемый комплекс программ предназначен в первую очередь для ведения НСИ в информационных системах, созданных на основе CASE-технологии METAS [8-10], для применения их в различных предметных областях без каких бы то ни было ограничений.

Программные компоненты могут быть также интегрированы и в информационные системы сторонних разработчиков для ведения классификаторов различных типов (иерархических, фасетных и пр.).

### **Понятие классификатора и задача управления НСИ**

*Классификатор* – это документ, с помощью которого осуществляется формализованное описание информации в информационных системах. Классификатор содержит наименования объектов, наименования классификационных группировок и их кодовые обозначения.

Основными объектами классификации и кодирования являются

*реквизиты-признаки* (справочные и группировочные). Они описывают процессы, место, время выполнения процессов, субъекты и объекты действия, отражаемые в различных показателях, учитываемых и анализируемых в ИС. К объектам классификации и кодирования относятся также *наименования показателей и документов*, используемых в ИС.

Для кодирования объектов их необходимо упорядочить по некоторым признакам. Результат упорядоченного распределения объектов заданного множества и называется *классификацией*, а совокупность правил распределения объектов множества на подмножества – *системой классификации*. То свойство (или характеристика, атрибут) объекта классификации, которое позволяет установить его сходство или различие с другими объектами, называется *признаком классификации*. *Основанием классификации* называется тот признак, по которому ведется разбиение множества на подмножества на определенной ступени классификации. *Ступень классификации* – это результат очередного распределения объектов одной классификационной группировки. *Уровень классификации* – это совокупность классификационных группировок, расположенных на одних и тех же ступенях классификации. *Глубина системы классификации* – это количество уровней классификации, допустимое в данной системе.

Системы классификации являются основой для построения всех словарей и справочников ИС, обеспечивая единообразие представления информации, сопоставимость показателей при обработке и анализе данных.

## **Модель классификатора**

В основе работы программных компонентов лежит математическая модель классификатора, разработанная авторами [6, 16]. Данная модель должна быть *универсальной* для различных типов классификаторов.

Под *универсальностью математической модели* будем понимать абстракцию, позволяющую описать:

- классификатор любого типа, формализуя общепринятые определения в области классификации и кодирования информации;
- представление классификатора в рамках ИС (на физическом, логическом и презентационном (визуальном) уровнях);
- средства управления классификаторами (создания, поддержки актуального состояния);
- алгоритмы преобразования к универсальной структуре классификаторов, импорта/экспорта классификаторов, приведение

импортируемых классификаторов в соответствие с единой системой управления.

Выбор в пользу *теоретико-графовой модели* аргументируется следующими фактами:

- Внутренняя структура классификатора изначально представляет собой граф, где вершины – это объекты классификации, а дуги – информация о соподчиненности объектов.
- Любая информационная система хранит данные в базе данных (БД). Структуру БД легко можно представить в виде графа, тогда представление классификатора в ИС сведется к отображению одного графа на другой.
- Пользователям удобно воспринимать классификатор в виде иерархии объектов – дерева. Дерево – это построение иерархии на множестве объектов классификации и отношений между ними. При этом часто деревом описываемая структура является только на экране пользователя (отсюда и название). С точки зрения структуры это ориентированный граф, возможно, с циклами.

В описании модели [6, 16] последовательно определяются все ее элементы – объекты классификации, классификаторы, отношения соподчиненности, операции над объектами классификации, отношениями соподчиненности и классификаторами.

Пусть имеется конечное линейно упорядоченное множество  $A$  предметных переменных  $a_1, \dots, a_n$ , называемых *атрибутами*, с областями возможных значений  $D_1, \dots, D_n$  соответственно, называемыми *их доменами* и содержащими символ  $\Lambda$  – *неопределённое значение*.

Домены различных атрибутов могут совпадать. В этом случае атрибуты называются *однотипными*. Для любого подмножества  $M \subseteq A$  декартово произведение  $D_M$  доменов всех атрибутов в  $M$  называется *доменом подмножества  $M$* , а его элементы – *значениями* этого подмножества.

*Объект классификации  $O_i$*  – элемент классифицируемого множества, характеризующийся своими атрибутами  $O_i = f(A_i) = A'$ , где  $f$  – характеристическая функция, определяющая множество  $A'$  из  $A$ .

*Структура классификатора*, как правило, должна иметь три блока:

- *блок идентификации*, включающий коды объектов классификации и классификационных группировок;

- блок наименований объектов и классификационных группировок на естественном языке;
- блок дополнительных признаков объектов, включающий наименования и коды дополнительных признаков объектов классификации.

В соответствии со структурой классификатора *разделим атрибуты на три группы:*

- $A_{iden}$  – *идентификационные атрибуты*, подмножество элементов из множества  $A$  (подмножество множества  $A_{iden}$  или все множество  $A_{iden}$  однозначно определяет объект классификации  $O_i$ ):
 
$$A_{iden} \subseteq A, A^*_{iden} \subseteq A, \exists! h = H(O_i(A^*_{iden}))$$
- $A_{name}$  – *атрибуты наименования*, подмножество элементов из множества  $A$ ;
- $A_{add}$  – *дополнительные атрибуты*, подмножество элементов из множества  $A$ .

*Признак классификации*  $p_{ij}$  – свойство или характеристика объекта, по которому производится классификация, т.е. существует такая характеристическая функция  $f_i$ , аргументом которой является объект классификации, а результатом – значение признака классификации. Признак классификации иногда называют основанием деления:

$$\exists n \in N, j = \overline{1, n}, \forall p_{ij} = f_j [O_i(A_{iden})], P_i = \{p_{i1}, \dots, p_{in}\}.$$

У объекта классификации может быть несколько признаков классификации.  $P_i$  – множество значений признаков классификации объекта  $O_i$ .

*Классификационная группировка*  $G_i$  – подмножество объектов, полученное в результате классификации. Это подмножество обладает совокупностью одних и тех же значений признаков классификации для каждого объекта классификации, и называется *множеством признаков группировки*  $G_i$ . Обозначим его  $P_i^G$ .

Классификационная группировка имеет *код* и *наименование* в системе классификации и может представлять собой объект классификации в классификаторе:

$$G_i = \{O_{i1}, \dots, O_{in}\}, \forall j, k \subseteq n: P_{ij} = P_{ik} = P_i^G.$$

Всегда существует некоторое *правило*, по которому определяется, принадлежит ли объект классификации данной классификационной группировке.

В качестве правила будем использовать предикат  $Rule_{G,O}(A_{iden})$ . Если значение предиката – истина, то объект  $O$  принадлежит группировке  $G$ , если ложно, то нет:

$$Rule_{G,O}(A_{iden}) = \text{true} \Leftrightarrow (\exists A^O_{iden} \in A^O): f_{rule}(A^O_{iden}) = f_{rule}(A^G_{iden}).$$

Как уже говорилось выше, классификационные группировки могут отражаться или не отражаться в кодах классификаторов. Например, в классификаторе ОКАТО (общероссийский классификатор административно-территориальных образований), в коде отражается группировка, то есть любой код ОКАТО несет смысловую нагрузку, некоторые позиции символов в коде отвечают за уровень группировки. Группировки могут задаваться независимо от кода, например в ОКФС (общероссийский классификатор форм собственности) используется сквозная нумерация кода, а группировки задаются специальным алгоритмом.

В результате объект классификации можно представить как тройку:

$$O_i = (A', P_i, G_i),$$

где  $A'$  – множество атрибутов объекта  $O_i$ ;

$P_i$  – множество признаков классификации объекта  $O_i$ ;

$G_i$  – классификационная группировка, к которой относится объект классификации  $O_i$ .

Рассмотрим пример представления объектов классификации в соответствии с описанными определениями. Пусть есть текстовое описание классификации (пример взят из ОКАТО):

---

Классификатор включает объекты административно-территориального деления, кроме сельских населенных пунктов, которые в настоящее типографское издание не входят из-за большого объёма этой информации.

Каждая позиция классификатора структурно состоит из 3-х блоков:

- 1) блок идентификации объекта;
- 2) блок наименования объекта;
- 3) блок дополнительных данных.

Блок идентификации объекта включает идентификационный код и контрольное число. Идентификационный код строится с использованием серийно-порядкового, последовательного и параллельного методов кодирования.

Длина кода – от 2-х до 8-ми разрядов в зависимости от уровня классификации, на котором находится объект.

Структура кодового обозначения в блоке идентификации:

XX XXX XXX КЧ,

где

- 1,2 знаки – объекты первого уровня классификации;
- 3,4,5 знаки – объекты второго уровня классификации;
- 6,7,8 знаки – объекты третьего уровня классификации;
- КЧ – контрольное число.

Контрольное число в кодовых обозначениях рассчитывается по действующей в ЕСКК методике расчета и применения контрольных чисел.

Для кодирования большинства объектов ОКАТО используется следующая структура кода. Разряды 1 и 2 предназначены для кодирования объектов федерального значения, расположенных на первом уровне классификации. Система кодирования этих объектов – серийно-порядковая, обеспечивающая преемственность с ранее действующим классификатором СОАТО. Разряды 3, 4, 5 используются для кодирования объектов второго уровня классификации, разряды 6, 7, 8 – для кодирования объектов третьего уровня. В этих случаях применяется последовательный метод кодирования. При этом разряды 3 и 6 отведены под признаки соответственно Р1 и Р2, указывающие уровень классификации и вид кодируемого объекта. В этом случае применяется параллельный метод кодирования.

Признак второго уровня классификации – Р1 (разряд 3) имеет значение:

- 1 – автономный округ;
- 2 – район (в том числе внутригородской), округ;
- 4 – город, посёлок городского типа.

Признак третьего уровня классификации – Р2 (разряд 6) имеет значение:

- 3 – внутригородской район, округ города;
- 5 – город, посёлок городского типа;
- 8 – сельсовет.

Из описанного выше текста можно выделить 5 признаков классификации. Запишем признаки в формате

«знак в блоке идентификации» = «какие значения может принимать»:

$p_1 : A_{iden}[1,2] = [0..9][1..9]$  принадлежность 1 уровню вложенности

$p_2 : p_1 \& A_{iden}[3,4,5] = [1..9][1..9][0..9]$  принадлежность 2 уровню вложенности

$p_3 : p_2 \& A_{iden}[6,7,8] = [1..9][1..9][0..9]$  принадлежность 3 уровню вложенности

$p_4 : A_{iden}[3] = [1,2,4]$  соответствие признаку  $P_1$

$p_5 : A_{iden}[6] = [3,5,8]$  соответствие признаку  $P_2$

Таким образом, в классификационную группировку первого уровня  $G_1$  попадают все объекты классификации  $O$ , у которых выполняется признак классификации  $p_1$  (первые две позиции в блоке идентификации – цифры) и остальные позиции – нули:

$$G_1 = \{O_1 \dots O_n\}, P_{G_1} = \{\forall i = \overline{0,9}, \forall j = \overline{1,9} : p_1(i, j) \& A_{iden}[3-8] = ij000000\}.$$

Аналогичного для группировок  $G_{xy}$  второго уровня вложенности должен выполняться признак  $p_1$ ,  $p_2$  и остальные позиции – нули:

$$G_{xy} = \{O_{11} \dots O_{nn}\},$$

$$P_{G_{xy}} = \{\forall i = \overline{1,9}, \forall j = \overline{1,9}, \forall k = \overline{0,9} : p_1(x, y) \& p_2(i, j, k) \& A_{iden}[678] = xyijk\ 000\}$$

*Классификатор* – это документ, с помощью которого осуществляется формализованное описание информации в информационных системах, содержащий наименования объектов, наименования классификационных группировок и их кодовые обозначения.

Пару  $K = (V, E)$ , где  $V$  – некоторое множество классификационных объектов с различными атрибутами и  $E$  – отношение соподчиненности на нем, назовём *классификатором*.

Введем понятие *корневой вершины* [12, 18] графа  $(V, E)$ , обозначим ее  $O_0, O_0 \in V$ .

Корневая вершина, как и любой объект классификации, является тройкой:

$$O_0 = (A^0, P_0, G_0),$$

где  $A^0$  – множество наименований атрибутов классификатора  $K$ .

Пример объекта:  $O_{14} : (A^4 = \{4, \text{высшее}\}, P_1, G_1 = \{O_{11}, \dots, O_{1n}\})$ .

Каждый объект классификации  $O_i$  обладает множеством атрибутов  $A^i$ , таким что, существует однозначное отображение  $f$  множества  $A^0$  на множество  $A^i$ :

$$f : A^0 \rightarrow A^i, \forall j = \overline{1, n}, n = |A^0| : a_j^i = f(a_j^0).$$



Под отображением  $f$  будем понимать следующее: каждому элементу множества  $A^0$  ставится в соответствие элемент множества  $A^1$  с тем же порядковым номером элемента:

$$A^0 = \{\text{код, наименование}\}$$

$$A^1 = \{1, \text{мужской}\}$$

$$A^2 = \{2, \text{женский}\}$$

$$\forall j = \overline{1, n}, n = |A^0| = 2:$$

$$a_1^1 = 1 \rightarrow a_1^0 = \text{код}, a_2^1 = \text{мужской} \rightarrow a_2^0 = \text{наименование}$$

$$a_1^2 = 2 \rightarrow a_1^0 = \text{код}, a_2^2 = \text{женский} \rightarrow a_2^0 = \text{наименование}$$

$P_0$  – множество признаков классификации, которое определяет первый уровень вложенности классификатора  $K$ .

$G_0$  – классификационная группировка нулевого уровня, к которой относятся объекты классификации первого уровня.

Два классификатора  $K_1$  и  $K_2$  *не пересекаются*, если не пересекаются множества их вершин. Таким образом, каждый классификатор  $K$  представляет собой множество классификационных объектов, отвечающих различным значениям классификации  $G$ .

Как уже говорилось выше, классификаторы могут быть линейными, иерархическими, фасетными. Рассмотрим представление данных типов классификаторов на описанной модели.

*Линейный классификатор*, или просто справочник, представляет собой множество объектов классификации, которые находятся на первом и единственном уровне вложенности (рис. 1).

*Фасетный классификатор* представляет собой множество объектов классификации, разбитых на *фасеты*, которые представляют собой первый уровень вложенности; на втором уровне вложенности находятся объекты классификации, принадлежащие тому или иному фасету. Для наглядности переобозначим объекты классификации первого уровня вложенности  $O_i = F_i$  (рис. 2). Каждый фасет представляет собой линейный или иерархический классификатор (пример – ОКИН, общероссийский классификатор информации о населении).

*Иерархический классификатор* представляет собой множество объектов классификации разбитых на классификационные группировки, которые имеют строгую иерархию подчиненности. На первом уровне вложенности находятся вершины иерархии; в классическом варианте вершина должна быть одна, в реальности их может быть больше (рис. 3).

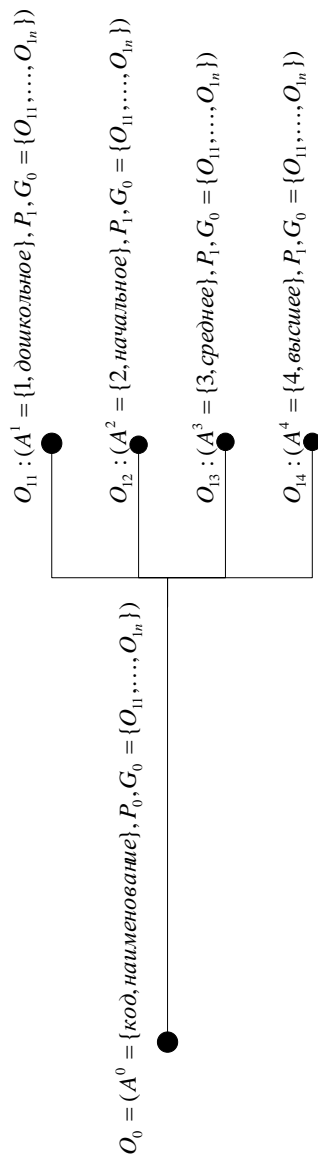


Рис. 1. Представление линейного классификатора

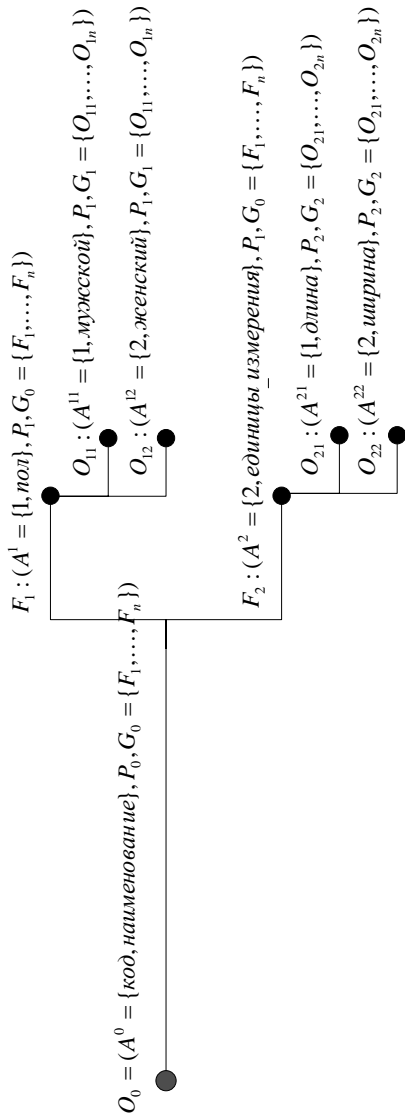


Рис. 2. Представление фасетного классификатора

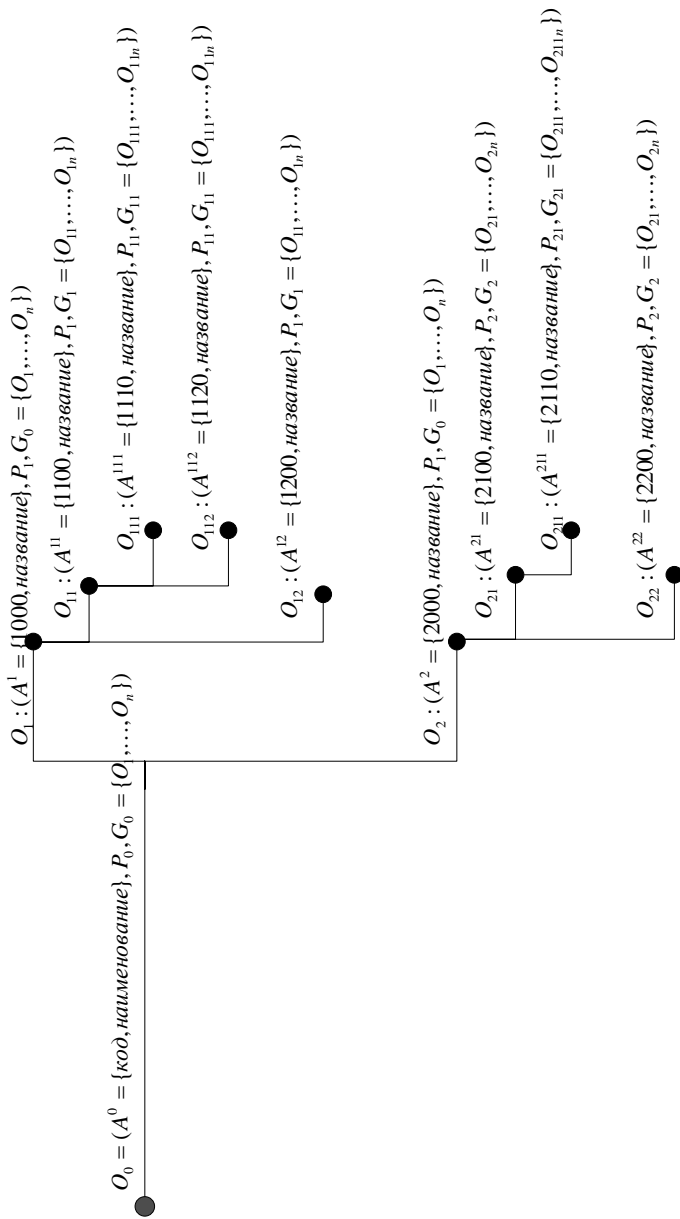


Рис. 3. Представление иерархического классификатора

Для любых двух классификационных объектов  $O_1 \in K$  и  $O_2 \in K$ ,  $K = (V, E)$ , говорят, что  $O_2$  *ссылается на*  $O_1$ , или  $O_2$  *подчинен* объекту  $O_1$ , и пишут  $O_2EO_1$ , если множество признаков классификации  $X_i$  объекта  $O_1$  включает в себя множество признаков классификации  $X_{ij}$  объекта  $O_2$ , т.е. если в графе  $K = (V, E)$  есть дуга, соединяющая вершину в  $O_1$  с вершиной  $O_2$ . Для каждого классификатора множество отношений соподчинённости представляет множество дуг  $E$  в графе  $K = (V, E)$ .

На рис. 4 представлен фрагмент графа классификатора  $K = (V, E)$ , где  $V = \{O_i, O_{i1}, O_{i2}\}$ ,  $E = \{(O_i, O_{i1}), (O_i, O_{i2})\}$ .

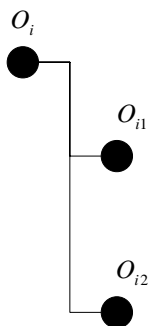


Рис. 4. Фрагмент графа классификатора

### Операции над классификационными объектами

Для начала рассмотрим операции, которые можно выполнять над атрибутами классификационного объекта.

- *Добавление атрибута.* Добавление атрибута рассматривается как теоретико-множественная операция объединения двух множеств:

$$Add(O_i, a): A^{i'} = A^i \cup \{a\}.$$

- *Удаление атрибута.* Удаление атрибута рассматривается как теоретико-множественная операция вычитания двух множеств:

$$Del(O_i, a): A^{i'} = A^i \setminus \{a\}.$$

- *Обновление атрибута.* Обновление атрибута рассматривается как две операции, удаление старого атрибута и добавление нового:

$$Update(O_i, a, a'): Del(O_i, a) \& Add(O_i, a').$$

Над однотипными классификационными объектами можно выполнять любые теоретико-множественные операции, в том числе:

1. *Объединение:*

$$V_1 \cup V_2 = \{(\{O_{1i}\} \cup \{O_{2j}\}) : (\exists O_{1i} : \{O_{1i}\} \in V_1) \& (\exists O_{2i} : \{O_{2i}\} \in V_2)\}$$

2. *Вычитание:*

$$V_1 - V_2 = \{(\{O_{1i}\} - \{O_{2j}\}) : (\exists O_{2i} : \{O_{2i}\} \in V_1) \& (\exists O_{2i} : \{O_{2i}\} \in V_2)\}$$

Введем операцию *манипулирования* классификационными объектами, записываемую в форме  $name(O_i, O'_i)$ , где  $name$  – имя операции,  $O_i, O'_i$  – объекты классификации. Результатом операции является объект классификации. Все объекты классификации, участвующие в операции, однотипные.

3. *Обновление объекта классификации.* Обновление объекта классификации – это замена атрибута  $O_i$  на соответствующий атрибут  $O'_i$ :

$$Update\_O(O_i, O'_i) : (\forall a_i \in A, \exists! a'_i \in A') \rightarrow (Update(O_i, a_i, a'_i)).$$

Над однотипными отношениями соподчиненности можно выполнять любые теоретико-множественные операции, в том числе:

4. *Объединение:*

$$E_1 \cup E_2 = \{(\{(O_{1ik}; O_{1il})\} \cup \{(O_{2jn}; O_{2jm})\}) : (\exists(O_{1ik}; O_{1il}) : (O_{1ik}; O_{1il}) \in \{(O_{1ik}; O_{1il})\}) \& (\exists(O_{2jn}; O_{2jm}) : (O_{2jn}; O_{2jm}) \in \{(O_{2jn}; O_{2jm})\})\}$$

5. *Вычитание:*

$$E_1 - E_2 = \{(\{(O_{1ik}; O_{1il})\} - \{(O_{2jn}; O_{2jm})\}) : (\exists(O_{2jn}; O_{2jm}) : (O_{2jn}; O_{2jm}) \in \{(O_{1ik}; O_{1il})\}) \& (\exists(O_{2jn}; O_{2jm}) : (O_{2jn}; O_{2jm}) \in \{(O_{2jn}; O_{2jm})\})\}$$

6. *Добавление отношения соподчиненности.* Добавление отношения соподчиненности рассматривается как теоретико-множественная операция объединения двух множеств:

$$Add(E, (O_i, O_j) : E' = E \cup \{(O_i, O_j)\}.$$

7. *Удаление отношения соподчиненности.* Удаление отношения соподчиненности рассматривается как теоретико-множественная операция вычитания двух множеств:

$$Del(E, (O_i, O_j) : E' = E \setminus \{(O_i, O_j)\}.$$

Над однотипными классификаторами можно выполнять любые теоретико-множественные операции, в том числе:

8. *Объединение:*

$$(V_1, E_1) \cup (V_2, E_2) = \{(V, E), \text{ где } V = (V_1 \cup V_2) \text{ и } E = (E_1 \cup E_2)\}.$$

9. *Вычитание:*

$$(V_1, E_1) \setminus (V_2, E_2) = \{(V, E), \text{ где } V = (V_1 \setminus V_2) \text{ и } E = (E_1 \setminus E_2)\}.$$

Введем операции *манипулирования* классификаторами, записываемые в форме  $name(K, O_i)$ , где  $name$  – имя операции,  $K$  – классификатор,  $O_i$  – объект классификации. Результатом операции является классификатор, а в операции Search результатом является объект классификации. Все объекты классификации, участвующие в операции, однотипные.

10. *Добавление объекта классификации.* Добавление объекта классификации рассматривается как теоретико-множественная операция объединения двух множеств. Операция возвращает сам классификатор, если такой объект классификации в нем уже есть, или объединение уже существующих объектов классификации с добавляемым:

$$Insert(K(V, E), O) = \begin{cases} K, \text{ если } O \in V \\ V \cup \{O\} \text{ and } E \cup \{O\} \text{ иначе} \end{cases}.$$

Добавление объекта в классификатор может быть выполнено по-разному. Например, если классификатор является ОК, то для того, чтобы поддерживать интегрируемость, при передаче данных нужно отслеживать, какие объекты классификации являются первоначальными, а какие по мере необходимости были дополнены пользователями. В противном случае может получиться, что один и тот же объект классификации определяет разные информационные объекты.

11. *Удаление объекта классификации.* Удаление объекта классификации рассматривается как теоретико-множественная операция разности двух множеств. В результате выполнения операции возвращается сам классификатор, если такого объекта классификации в нем нет, или из существующих в классификаторе объектов удаляется указанный объект классификации:

$$Delete(K(V, E), O) = \begin{cases} K, \text{ если } O \notin V \\ V - \{O\} \text{ and } E - \{O\} \text{ иначе} \end{cases}.$$

Операция удаления может также выдавать разные результаты. Например, если классификатор является ОК, то удалить можно только те объекты классификации, которые ввёл пользователь. Иначе может возникнуть противоречивость классификаторов, находящихся на разных рабочих местах.

12. *Обновление объекта классификации.* Операция *Update* обновляет объект классификации из классификатора, а именно: обновляет значения атрибутов объекта классификации. Операция возвращает  $\Lambda$  – *неопределенное значение*, если объект классификации обновлять не надо, или обновляет значения атрибутов объекта классификации, используя операцию *Update\_O*.

$$Update(K(V, E), O_i, O_j) = \begin{cases} \Lambda, & \text{если } O_i(A) \notin O_j(A) \\ \text{иначе } update\_O(O_i, O_j) \end{cases}.$$

Операция обновления тоже имеет свои особенности, так как обновление классификатора может идти полностью, т.е. одна версия классификатора заменяется другой, или же может быть выполнено частичное обновление каких-то конкретных объектов классификации.

13. *Операция поиска:*

$$Search(K(V, E), O) = \begin{cases} O, & \text{если } O \in V \\ \Lambda & \text{иначе} \end{cases}.$$

Операция *Search* позволяет найти объект классификации в классификаторе. Операция возвращает сам объект классификации, если он есть в классификаторе, или, если объекта классификации нет, то  $\Lambda$  – *неопределённое значение*.

## **Операции над классификаторами**

Рассмотрим основные алгоритмы работы с классификаторами. Можно выделить *четыре основных операции по манипулированию классификаторами* в информационных системах: *создание, удаление, обновление, импорт/экспорт* классификаторов. Опишем алгоритмы выполнения этих операций на основе предложенной модели.

1. *Создание классификатора.* Под созданием классификатора будем понимать последовательное создание множества объектов классификации, в том числе признаков классификации, классификационных группировок, и отношений соподчиненности между объектами классификации. В данной интерпретации не рассматриваются задачи



выявления закономерностей на основе массивов данных, классификации исходных данных (разработка кодов классификации). Будем считать, что этот этап уже выполнен экспертами или автоматизированными системами, то есть в нашем случае создание классификатора сводится к двум этапам: созданию множества объектов классификации и созданию множества отношений соподчинённости.

*Создание множества объектов классификации* может быть выполнено

- экспертом вручную;
- импортом уже созданных объектов классификации из внешних источников.

*Создание множества отношений соподчинённости* может быть выполнено

- экспертом вручную;
- импортом уже созданных отношений соподчинённости из внешних источников;
- на основе существующих в модели признаков классификации и классификационных группировок.

2. *Удаление классификатора.* Под удалением классификатора будем понимать последовательное удаление множества отношений соподчинённости между объектами классификации и множества объектов классификации. Данная операция не является тривиальной, поэтому на ней акцентироваться внимание не будет.

3. *Обновление классификатора.* Под обновлением классификатора будем понимать последовательное изменение множества объектов классификации, в том числе признаков классификации, классификационных группировок, и множества отношений соподчинённости между объектами классификации. Как уже говорилось выше, будем рассматривать три операции:

- Добавление во множество объектов классификации нового элемента.
- Изменение атрибутов у элемента множества объектов классификации:
  - изменение значений атрибутов;
  - удаление атрибутов;
  - добавление атрибутов.
- Удаление элемента из множества объектов классификации.

Введем обозначения:  $K_{update}$  – изменения в классификаторе, которые нужно произвести;  $K_{del}$  – подмножество объектов классификации и отношений соподчинённости, которые нужно удалить;  $K_{add}$  –

подмножество объектов классификации и отношений соподчиненности, которые нужно добавить в классификатор;  $K_{upd}$  – подмножество объектов классификации, которые нужно обновить;  $K_{out}$  – классификатор из внешних источников;  $K$  – классификатор соответствующий универсальной модели;  $a_j \sim a_i$  – элемент  $a_j$  эквивалентен элементу  $a_i$ ;  $a_j \rightarrow a_i$  – элемент  $a_j$  ставится в соответствие элементу  $a_i$ ;  $a_i^*$  – элемент является обязательным элементом множества  $A^0$ .

Следовательно, обновление классификатора можно разделить на следующие этапы:

- 1) удалить подмножество  $K_{del}$  из классификатора  $K$ ;
- 2) если необходимо изменить множество атрибутов  $A^0$  классификатора  $K$ ;
- 3) добавить подмножество  $K_{add}$  в классификатор  $K$ ;
- 4) обновить объекты классификации из подмножества  $K_{upd}$ .

4. *Импорт классификатора.* Под импортом классификатора будем понимать создание множества объектов классификации и множества отношений соподчиненности на основе сторонней модели. Под сторонней моделью будем понимать модель, описывающую данные из внешних источников. Перед созданием множеств возможно приведение (отображение) множества атрибутов сторонней модели на множество атрибутов предложенной модели.

5. *Экспорт классификатора.* Под экспортом классификатора будем понимать создание отдельных множеств объектов классификации и отношений соподчиненности в соответствии со сторонней моделью. Перед созданием множеств возможно приведение (отображение) множества атрибутов предложенной модели на множество атрибутов сторонней модели. Экспортироваться может не весь классификатор, а только его часть, следовательно, при экспорте необходимо учитывать условия экспорта данных.

## Заключение

Разработанный на основе описанной модели комплекс программ включает средства «ручного» ведения классификаторов с возможностью задания их структуры, описаний элементов, их ввода и редактирования. Кроме того, реализованы средства импорта классификаторов, представленных в формате таблиц MS Excel или реляционных баз данных – источников данных, доступных через ODBC.

Для конечного пользователя реализованы удобные средства навигации, визуализации НСИ, просмотра и поиска информации.

Разработанная модель может также стать основой для реализации средств автоматизации разработки онтологий предметных областей на основе НСИ, классификаторов (международных, общероссийских, отраслевых).

Комплекс программ ведения НСИ [4, 5, 15, 16, 17] позволяет настраиваться на различные программные платформы, работать под управлением различных операционных систем Microsoft, использовать для работы различные реляционные СУБД и источники данных, для которых существуют драйверы ODBC.

Для функционирования runtime-компонентов необходимо установить .NET Framework (распространяется бесплатно), драйверы ODBC (при установке операционной системы) и СУБД (можно использовать, в частности, Microsoft SQL Server Express, которая распространяется бесплатно). Данные НСИ могут быть импортированы, в частности, из электронных таблиц Excel.

Ограничения использования системы определяются только требованиями лицензионной чистоты и требованиями, предъявляемыми перечисленными выше программными средствами.

### **Библиографический список**

1. *Вендров А.М.* Проектирование программного обеспечения экономических информационных систем. М.: Финансы и статистика, 2000.
2. ГОСТ 50.1.021-2000 Единая система классификации и кодирования технико-экономической информации. М.: Изд. стандартов, 2000.
3. ГОСТ 34.601-90. Автоматизированные Системы Стадии создания. Комплекс стандартов на автоматизированные системы. М.: Изд. стандартов, 1997.
4. *Лукиных Н.В.* Реализация унифицированных средств для работы с классификаторами в информационных системах // Технологии Microsoft в информатике и программировании: тезисы докладов конференции-конкурса работ студентов, аспирантов и молодых ученых. Новосибирск, 2005. С. 39-40.
5. *Лукиных Н.В.* Универсальный компонент управления для представления и администрирования классификаторов // Технологии Microsoft в информатике и программировании:

- тезисы докладов конференции-конкурса работ студентов, аспирантов и молодых ученых. Новосибирск, 2006. С. 109-111.
6. *Лядова Л.Н., Скрябина Н.В.* Математическая модель для предоставления классификаторов // Математика программных систем: межвуз. сб. науч. тр. / Перм. ун-т. Пермь, 2003. С. 56-60.
  7. *Лядова Л.Н.* Архитектура информационной системы «Образование Пермской области» // Математика программных систем: межвуз. сб. науч. тр. / Перм. ун-т. Пермь, 2002. С. 25-35.
  8. *Лядова Л.Н., Рыжков С.А.* CASE-технология METAS // Математика программных систем: межвуз. сб. науч. тр. / Перм. ун-т. Пермь, 2003. С. 4-18.
  9. *Лядова Л.Н., Рыжков С.А.* METAS – технология создания информационных систем сферы образования // XIV Международная конференция-выставка «Информационные технологии в образовании»: сборник трудов участников конференции. Ч. IV. М.: МИФИ, 2004. С. 38-40.
  10. *Лядова Л.Н., Рыжков С.А.* Технология создания информационных систем, управляемых метаданными // Математические методы и информационные технологии в экономике, социологии и образовании: сборник статей XIII Международной научно-технической конференции / Пенза: Пензенская государственная технологическая академия, Приволжский дом знаний, 2004. С. 316-319.
  11. *Рыжков С.А.* Концепция метаданных в разработке информационных систем // Математика программных систем: Межвуз. сб. науч. тр. / Перм. ун-т. Пермь, 2002. С. 36-44.
  12. *Оре О.* Теория графов. М.: Наука, 1980.
  13. *Помазков Я.* Системы НСИ: мировой опыт и тенденции развития // Компьютерный журнал PCWeek/RE. №12/2006. С. 57-61.
  14. *Свами М., Тхуласироман К.* Графы, сети и алгоритмы. М.: Мир, 1984.
  15. *Скрябина Н.В.* Автоматизация построения классификаторов на основе анализа документов // Материалы XLIII Международной научной студенческой конференции «Студент и научно-технический прогресс»: Информационные технологии / Новосибирск: Новосиб. гос. ун-т, 2005. С. 210-211.

16. *Скрябина Н.В.* Математическая модель для предоставления классификаторов // Современные проблемы механики и прикладной математики: сборник тр. междунар. школы-семинара / Воронеж: ВГУ, 2004. С. 230-235.
17. *Скрябина Н.В.* Унифицированные средства управления классификаторами в информационных системах // Молодежь. Образование. Экономика: сборник научных статей участников конференции. Часть 4. – Ярославль: Изд-во «Ремдер», 2004. С. 170-174.
18. *Яблонский С.В.* Введение в дискретную математику. 2-е изд. М.: Наука, 1986.