

УДК 004.652;510.62

ФОРМАЛЬНЫЕ СВОЙСТВА СОВМЕСТИМОСТИ СПИСКОВ**Гладков Сергей Львович¹,***e-mail: gladkovs@list.ru,*¹*Общество с ограниченной ответственностью «Айгео», г. Красноярск, Россия*

Целью настоящей статьи является построение теоретической модели автоматического соединения управленческих списков различных ведомств. Тема статьи актуальна для организации межведомственного взаимодействия в части обмена данными. На основе остова (общей части) управленческих списков в статье даны формальные определения атомарных и структурированных управленческих списков с атомарными и лингвистическими атрибутами, представляющих собой расширение отношений, введенных Эдгаром Коддом. Рассмотрены признаки совместимости таких списков. Методы исследования: логическая категоризация управленческих списков от простых к сложным.

Рассмотренные в статье формальные определения однородных и структурированных списков, содержащих атомарные и лингвистические атрибуты, а также признаки их совместимости могут лечь в основу разработки стандартов автоматизированного обмена данными между информационными системами. В частности, определения однородных и структурированных списков могут использоваться в процессе разработки обменных форматов, а также сервисов обмена данными между различными информационными системами.

Ключевые слова: управленческие списки, семантика управленческих списков, предикат управленческого списка, совместимость управленческих списков, соединимость управленческих списков, лингвистическая переменная

FORMAL COMPATIBILITY FEATURES OF REGISTERS**Gladkov S.L.¹,***e-mail: gladkovs@list.ru,*¹*Aygeo Limited, Krasnoyarsk, Russia*

The purpose of this article is build a theoretical model of automatic connection of various departmental managerial lists. The topic of the article is relevant for organizing interagency cooperation in terms of data exchange. Basing on the framework (common part) of management lists the article provides formal definitions of atomic and structured management lists with atomic and linguistic attributes. The latter are an extension of the relations introduced by Edgar Codd. The article also views the compatibility signs of such lists. It uses such a research method as logical categorization of management lists from simple to complex ones.

The formal definitions of homogeneous and structured lists considered in the article containing atomic and linguistic attributes as well as signs of their compatibility can form the basis for the development of standards for automated data exchange between information systems. In particular the definitions of homogeneous and structured lists can be used in the development of exchange formats and for data exchange services between various information systems as well.

Keywords: managerial registers, semantics of managerial registers, registers predicate, compatibility registers, connectivity registers, linguistic variable

DOI 10.21777/2500-2112-2021-3-60-71

Введение

Федеральный закон РФ от 27 июля 2010 года № 210-ФЗ «Об организации предоставления государственных и муниципальных услуг» устанавливает требование автоматического межведомственного взаимодействия между базами данных различных государственных и муниципальных органов власти. Но пока эта возможность реализуется только специальными приложениями, организующими обмен данными между двумя (или больше) заранее определенными базами данных.

Начнём с анализа отношения (relation), введенного Эдгаром Коддом [1], как ближайшего «родственника» управленческого списка. Для этого рассмотрим упрощенное определение отношения. (Точное определение отношения приведено в книге Кристофера Дейта «Введение в системы баз данных» [2, с. 207].)

Определение 1. Отношение r состоит из заголовка и тела, которые соответствуют определениям, приведенным ниже.

- Заголовок отношения $h(r)$ представляет собой полное множество атрибутов r .
- Атрибут – упорядоченная пара $\langle A_i, T_i \rangle$, где A_i – имя атрибута, T_i – имя типа атрибута;
- Тело отношения $b(r)$ представляет собой множество кортежей, имеющих один и тот же заголовок.

– Кортеж t отношения r – множество упорядоченных троек в форме $\langle A_i, T_i, v_i \rangle$, где A_i – имя атрибута, T_i – имя типа и v_i – значение атрибута A_i с типом T_i из коллекции типов T_i ($i = 1, 2, \dots, n$), которые не обязательно все должны быть разными.

Здесь обозначение r имеет два значения. С одной стороны, r – это отношение в целом. С другой стороны, r – это название отношения. Пока такая двойственность обозначения не приводит к противоречивости изложения, r будет пониматься как пара: отношение и его наименование. В особых случаях, когда будет обсуждаться именно название отношения, будет использоваться обозначение $N(r)$.

Определение 2. Подмножество атрибутов k отношения r является потенциальным ключом этого отношения, тогда и только тогда, когда соответствует следующим условиям¹

– а) *Уникальность.* Ни одно допустимое значение r никогда не содержит два разных кортежа с одним и тем же значением k .

– б) *Несократимость.* Никакое строгое подмножество k не обладает свойством уникальности.

Теперь оценим, насколько определение отношения годится для представления управленческого списка, попытавшись на его основе сформулировать условия совместимости списков.

Допустим, что управленческие списки L и M – это синонимы отношений r_1 и r_2 , для которых при знаком совместимости по атрибутам $\langle A_i(r_1), T_i(r_1) \rangle$ и $\langle A_j(r_2), T_j(r_2) \rangle$ следует считать:

– совпадение названий типов этих отношений $T_i(r_1) = T_j(r_2)$;

– домен, т.е. множество значений атрибута $A_i(r_1)$ является подмножеством домена $A_j(r_2)$ или $A_j(r_2)$ является подмножеством домена $A_i(r_1)$. Другими словами

$$DOMAIN(A_i(r_1)) \subseteq DOMAIN(A_j(r_2)) \vee DOMAIN(A_j(r_2)) \subseteq DOMAIN(A_i(r_1)).$$

Если считать списки L и M тождественными отношениям, то окажется, что их возможно соединить, например, по атрибутам РОСТ(L) и ВЕС(M) на том основании, что их значения относятся к одному и тому же числовому типу. Но здесь исследуется возможность совместимости списков по данным об одних и тех же объектах.

Если же дополнить приведенный выше признак совместимости требованием совпадения названий атрибута $\langle A_i(L), T_i(L) \rangle$ списка L с названием атрибута $\langle A_j(M), T_j(M) \rangle$ списка M , то окажется что списки L и M окажутся формально несовместимыми по атрибутам ИМЯ СТУДЕНТА(L) и ИМЯ ШКОЛЬНИКА(M). А значит из-за несовпадения названий родственных атрибутов в списках студентов и школьников мы можем не найти данные об учёбе в школе нынешнего студента.

Определение отношения включает сведения о синтаксисе и структуре, соответствующей списку (таблице), что обеспечивает широкие возможности и гибкость при его практическом использовании

¹ Подробнее о точном определении ключей отношений [2, с. 356].

как единицы хранения данных. Но в нашем случае это определение оказалось слишком широким, чтобы стать фундаментом для определения совместимости управленческих списков.

Управленческие списки

Одним из препятствий на пути совместимости списков является разнообразие наименований, как самих списков, так и их атрибутов. Так в разных организациях один и тот же по смыслу список может называться по-разному, например, «Граждане», «Люди», «Население». Более подробный анализ проблем работы с управленческими списками приведен в статье «Списки» на сайте автора. Поэтому дальнейшее изложение будет вестись в предположении истинности следующего утверждения, принятого без доказательства.

Аксиома о смысле списка. Для каждого списка L с одним (или более) потенциальными ключами существует по крайней мере одно родовое понятие C , обладающее следующими свойствами:

- каждому идентифицирующему понятие набору характеристик $\{F_k(C)\}$ соответствует один потенциальный ключ списка $\{K_i(L)\}$, где $K_i(L)$ – атрибут списка L в ключе;
- каждой идентифицирующей характеристике (свойству) понятия $F_k(C)$ соответствует, по крайней мере, один атрибут, входящий в состав соответствующего потенциального ключа $\{K_i(L)\}$;
- каждый набор значений такого потенциального ключа позволяет однозначно идентифицировать любой чувственно воспринимаемый объект (предмет, вещь и т.д.) из полного множества объектов, объединенных этим понятием.

Заметим, что в аксиоме нет требования взаимной однозначности набора идентифицирующих понятие характеристик и атрибутов соответствующего потенциального ключа списка, т.к. список и соответствующее ему понятие формализованы (нормализованы [3]) обычно в разной степени.

Аксиома о смысле списка позволяет представить признак совместимости двух списков через ссылку на общую характеристику общего родового понятия этих списков. Аксиома также утверждает наличие у каждого списка потенциального ключа соответствующего набору идентифицирующих характеристик родового понятия. На практике далеко не каждый список в силу принципа минимальности [3, с. 9] содержит такой ключ. Но такой ключ всегда может быть добавлен, поэтому в этой работе считается, что каждый список им обладает.

Реальные управленческие списки часто используются не только по основному назначению, но, например, как место для записок на память. Поэтому вместо универсального определения управленческого списка будет дана серия определений списков различной формы по мере увеличения уровня их сложности. Под сложностью формы списка понимается неоднородность записей списка, сложность значений атрибутов списка (наличие структуры значения), а также сложность разделения «не атомарных» атрибутов, т.е. атрибутов, значения которых разделяются на отдельные осмысленные части, к атомарным [1, с. 150].

Вместе с каждым уровнем списка будут рассматриваться вопросы совместимости таких списков, а также вопросы их преобразования к спискам более простой формы.

Предварительные определения

Начнем с базового определения остова [4, с. 108] списка, которое в дальнейшем будет уточняться.

Определение 3. (Базовое). Список L состоит из названия списка $N(L)$, заголовка, тела, а также родового понятия C с набором его характеристик, которые соответствуют определениям, приведенным ниже.

– *Заголовок списка $H(L)$ представляет собой полное множество атрибутов L , включающее характеристический ключ списка.*

– *Атрибут – упорядоченная пара (A_i, D_i) , где A_i – имя атрибута, D_i – имя домена (множества допустимых значений атрибута), i – целочисленный индекс из интервала $[1, n]$, n – количество атрибутов списка.*

– *Характеристический ключ списка – подмножество полного множества атрибутов $\{(A_i, D_i) | i_k \in [1, k] \& k \leq n\} \subseteq \{(A_i, D_i) | i \in [1, n]\}$, обладающее свойствами уникальности и несократимости. (В дальнейшем для удобства атрибуты, входящие в состав ключа, будут обозначаться K_i)*

- Тело списка $\mathbf{B}(L)$ представляет собой множество записей, имеющих один и тот же заголовок.
- Запись \mathbf{R} списка – множество упорядоченных троек в форме (A_i, D_i, v_i) , где A_i – имя атрибута, D_i – имя домена и v_i – значение атрибута A_i из множества D_i .
- Родовое понятие \mathbf{C} списка – пара, состоящая из имени понятия $\mathbf{N}(\mathbf{C})$ и множества характеристик понятия $\mathbf{F}_k(\mathbf{C})$, где $j \in [1, p]$.

Дополнительно приведем полезное определение общего атрибута двух списков.

Определение 4. Списки L и M обладают общим атрибутом $A(L, M)$, если

- Совпадают их родовые понятия $\mathbf{C} \equiv \mathbf{C}(L) = \mathbf{C}(M)$.
- Существует характеристика $\mathbf{F}(C)$ общего понятия \mathbf{C} , которая соответствует как атрибуту $(A(L), D(L))$ списка L , так и атрибуту $(A(M), D(M))$ списка M , т.е.

$$(\exists (A(L), D(L)) \in \{(A_i(L), D_i(L))\}) \wedge (\mathbf{F}(C) \Rightarrow A(L)) \wedge$$

$$(\exists (A(M), D(M)) \in \{(A_j(M), D_j(M))\}) \wedge (\mathbf{F}(C) \Rightarrow A(M)).$$

- Список L принадлежит списку M по общему атрибуту $A(L, M)$, если домен, т.е. множество значений атрибута, $A(M)$ является подмножеством домена $A(L)$ или домен $A(L)$ является подмножеством домена $A(M)$, т.е. $DOMAIN(A(M)) \subseteq DOMAIN(A(L))$

Предикат списка

Понятие \mathbf{C} обладает логической формой, которая называется предикатом [5, с. 103]. Следовательно, можно определить логическую форму каждого списка L , для которого \mathbf{C} – родовое понятие.

Определение 5. Предикат списка L – логическая функция $\mathbf{P}(L) \equiv \mathbf{P}(C)$, определяющая принадлежность набора значений $\{v_j\}$ характеристик $\{F_k(C)\}$ понятию \mathbf{C} , имеющая следующую структуру:

$\mathbf{P}(C) \equiv P_1(F_1(C)) \wedge P_2(F_2(C)) \wedge \dots \wedge P_j(F_j(C)) \wedge \dots \wedge P_m(F_m(C))$, где $P_j(F_j(C))$ – предикат над одним атрибутом $F_j(C)$ (одноместный предикат [5, с. 92]);

$\{F_j(C)\}$ – набор характеристик, идентифицирующий понятие \mathbf{C} .

Таким образом, предикат $\mathbf{P}(C)$ списка L является средством верификации (подтверждения) и фальсификации (опровержения) принадлежности набора значений соответствующих характеристикам атрибутов списка L понятию \mathbf{C} , т.к. каждому набору значений списка атрибутов ставит в соответствие одно из логических значений «истина» или «ложь».

Характеристики $\{F_j(C)\}$, идентифицирующие понятие \mathbf{C} , подразделяются на родовые и видовые отличия понятий [5, с. 92]. При этом видовые характеристики преобладают при идентификации понятий, соответствующих спискам.

Набор значений идентифицирующих характеристик неявно задает последовательность сужающихся понятий, и тем самым может быть представлен как путь образования понятия \mathbf{C} .

На рисунке 1 для наглядности использованы атомарные предикаты вида $P_j(F_j(C)) \equiv F_j(C) = a_j$, где a_j – константа, которые представляют собой звенья пути образования (формирования) понятия. Например, в составе предиката понятия «Улицы Красноярск» присутствует атомарный предикат F_1 («Город») = «Красноярск». Этот атомарный предикат соответствует понятию «населённые пункты» и указывает на то, что среди его объектов должен быть выбран «город Красноярск».

Определение 6. Пусть понятие \mathbf{C} занимает место n на пути его образования, тогда последовательность понятий $\{C_1; C_2; \dots; C_{n-1}\}$, называется контекстом понятия \mathbf{C} . Контекст родового понятия \mathbf{C} списка L будем называть **контекстом списка** [6]. Контекст списка L , дополненный родовым понятием \mathbf{C} списка будем называть **путем списка**

$$\{C_1; C_2; \dots; C_{n-1}\} \cup C \equiv \{C_1; C_2; \dots; C_{n-1}, C\} \equiv \{C_1; C_2; \dots; C_{n-1}, C_n\}.$$

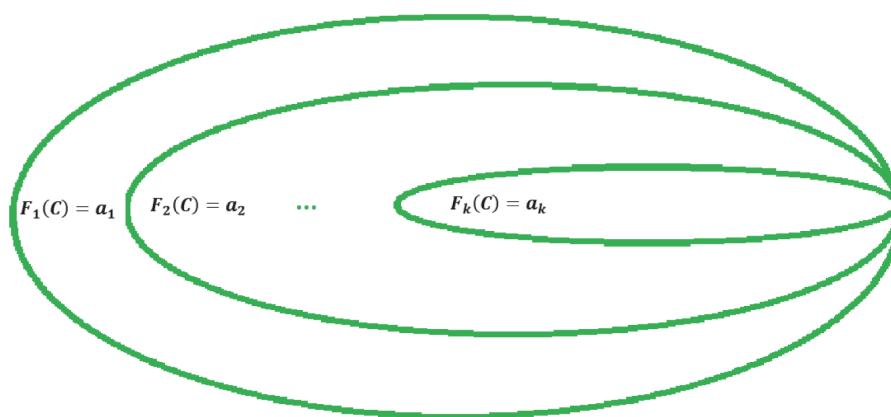


Рисунок 1 – Путь образования понятия

Из предыдущих рассуждений следует, что каждому понятию пути списка соответствует один или несколько атомарных предикатов. Число атомарных предикатов списка должно быть больше или равно числу понятий в пути списка, т.к. каждый атомарный предикат соответствует одному из понятий контекста или родовому понятию списка.

Контекст списка L будет использован в дальнейшем для сведения (преобразования) к списку, соответствующему одному из понятий контекста C (путем сокращения числа атрибутов списка, а также удаления одинаковых записей, возникших из-за удаления атрибутов). Наличие такого преобразования позволит совмещать списки, пути которых содержат общие понятия. Например, для выделения адресов из списков объектов недвижимости.

Записи списка

С точки зрения теории множеств список – это множество записей, которые могут быть представлены точками этого множества. Но с логической и смысловой точки зрения, записи – это единичные списки, т.е. списки, состоящие из одной записи. Поэтому к ним применимы все предыдущие рассуждения.

Родовое понятие C списка L всегда определяет класс объектов. При этом каждая запись списка соответствует понятию c . Поэтому в определении однородности понятия введено требование неразделимости его на подклассы.

Родовое понятие C списка L может быть представлено как синтез понятий об объектах, соответствующих записям списка:

$C \equiv c_1 + c_2 + \dots + c_i + \dots + c_w$, где c_i – понятие записи списка, w – мощность (объём) C , а «+» операция «обобщения» [7] или «синтеза» понятий [6].

Для того чтобы преодолеть возможную неопределённость в дальнейшем прописными буквами (C) будут обозначаться понятия о классах объектов, а строчными буквами единичные понятия.

Каждое понятие записи списка c_i , как и всякое понятие, обладает набором идентифицирующих характеристик, подразделяющихся на две части: идентифицирующие характеристики понятия C и характеристики, определяющие видовые особенности c_i , как подвида. Характеристикам видовых особенностей понятия c_i соответствует ключ записи l_i , значения которого уникальны в пределах списка L .

Каждому понятию записи списка c_i соответствует предикат

$$P(c_i) \equiv P(C) \wedge P_1(f_1(c_i)) \wedge P_2(f_2(c_i)) \wedge \dots \wedge P_j(f_j(c_i)) \wedge \dots \wedge P_m(f_m(c_i)),$$

где $P(C)$ – предикат списка; $P_j(f_j(c_i))$ – одноместный предикат над одной характеристикой записи атрибутом $f_j(c_i)$; $\{f_j(c_i)\}$ – набор видовых особенностей понятия c_i , соответствующего записи l_i .

Ну и наконец, каждое понятие c_i записи обладает контекстом и путем образования. Контекст понятия записи начинается с пути образования родового понятия списка C и дополняется подвидами класса C , которым принадлежит единичное понятие c_i . Путь записи списка получается объединением ее контекста с характеристиками, идентифицирующими запись в пределах списка.

Каждый список может содержать 2 типа записей: обычные и итоговые записи.

Независимо от уровня списка контекст и путь образования итоговой записи t_i списка L совпадают с контекстом и путём образования самого понятия. Т.е. итоговая запись списка L содержит характеристики списка как объекта, порожденного совокупностью обычных записей. Значения характеристик итоговой записи получаются путем применения к соответствующим характеристикам обычных записей агрегатных функций, например, SUM(), MAX(), MIN(), AVG(). На этом основании в [6] операция $C \equiv c_1 + c_2 + \dots + c_i + \dots + c_n$, называется операцией синтеза или порождения понятий.

Таким образом, итоговая запись списка может быть создана из набора обычных записей списка L , поэтому при изучении возможности совместимости списков, наличие в них итоговых записей будет игнорироваться.

Замечание. В определение атрибута списка можно добавить соответствующую ему агрегатную функцию, тогда операция соединения списков может порождать и одну итоговую запись.

Однородные списки

Однородное понятие C – это неразделенное в пределах списка на подклассы понятие [5, с. 203], т.е.

$\nexists m > 1 \mid C \equiv C_1 + C_2 + \dots + C_m$, где «+» операция «обобщения» [7] или «синтеза» понятий [6].

Рисунок 2 демонстрирует различные примеры операций «обобщения». В пределах этого рисунка однородными понятиями являются: «Вагончик», «Земельный участок», «Жилое здание», «Гараж».

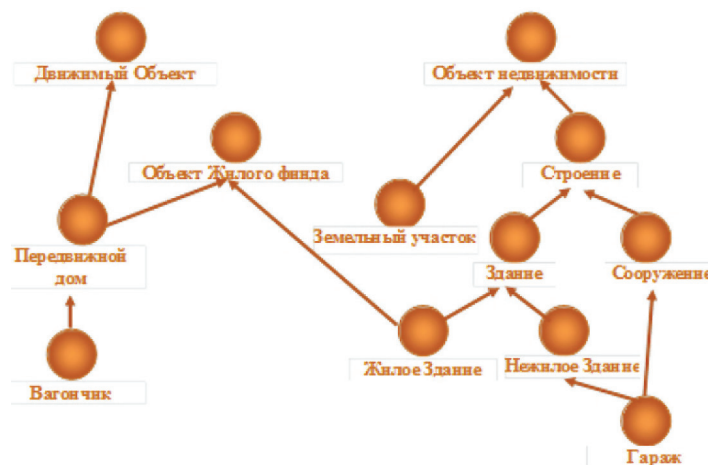


Рисунок 2 – Структура понятий (рисунок из [6])

Однородность класса C , может быть определена через понятия этого класса. Понятие C однородно, если не существует понятия c_i , принадлежащего C , которое принадлежит и подвиду $C_j \subset C$.

$$\nexists c_i \in C \mid \exists C_j \subset C \wedge c_i \in C_j.$$

Определение 7. Список L называется однородным, если он соответствует однородному понятию C .

Следствие 1. Контекст каждой записи однородного списка совпадает с путем образования этого списка.

Однородные атомарные списки

Атомарный атрибут (A, D_i) – это атрибут, значения которого не разделяются на отдельные осмысленные части, т.е.

$$\nexists m > 1 \mid D_i \equiv D_{i_1} \times D_{i_2} \times \dots \times D_{i_m}.$$

Атомарному атрибуту списка L соответствует атомарный одноместный предикат.

Поясним на примере. Пусть A_i – дата в формате «ДД.ММ.ГГГГ». Этот атрибут в общем случае не атомарный, т.к. может быть разделен на три: день, месяц и год. Предикат в накладывающий условие правильности значения этого атрибута не будет одноместным, т.к. в качестве переменных такого предиката должны быть как месяц, так и год.

А теперь разделим атрибут A_i на три: A_{id} – «день», A_{im} – «месяц», A_{iy} – «год». В этом случае общий предикат $A_{iy} = 2021 \wedge A_{im} = 2 \wedge A_{id} \leq 28$ представляется тремя одноместными предикатами, соединенными логической операцией «и» (см. рисунок 1).

Определение 8. Список L – атомарный список, если содержит только атомарные атрибуты. Если однородный список L , обладающий свойством атомарности, называется однородным атомарным списком.

Из аксиомы о смысле списка следует, что каждому набору характеристик, идентифицирующему понятие C списка L , соответствует потенциальный ключ $\{K_i(L)\}$ этого списка.

Следствие 1. Если L – однородный атомарный список, то каждой характеристике $F_k(C)$ соответствует только один атрибут списка $\{A_i(L)\}$.

Если отношение между списком характеристик C и соответствующим подмножеством атрибутов L не взаимно-однозначно, то либо список содержит не атомарный атрибут, либо понятие C не однородно.

Записи атомарного списка

Все обычные записи I_i однородного атомарного списка обладают одним и тем же контекстом, который соответствует пути образования списка. Что следует из определения контекста и однородности списка L .

Итоговые записи однородного атомарного списка могут содержать характеристики родового понятия $F_j(C) = a_j$, со значениями пути образования понятия, а также атрибуты, значения которых получены путем применения агрегатных функций, например, SUM(), MAX(), MIN(), AVG().

Заметим, что перечисленные выше функции могут быть применены только к атомарным атрибутам.

Совместимость однородных атомарных списков

Признак совместимости по атрибуту. Пусть даны однородные атомарные списки L и M . Будем говорить, что список M совместим со списком L по атрибуту, если

- списки L и M обладают общим атрибутом (см. данное выше определение общего атрибута списков);
- домены атрибутов, соответствующих общему атрибуту, $A(L, M)$ обладают следующим свойством: $DOMAIN(A(M)) \subseteq DOMAIN(A(L))$.

Например, пусть M – список объектов недвижимости города Красноярска, а L – список объектов недвижимости Красноярского края. Тогда список M совместим со списком L по атрибуту «Адрес объекта недвижимости», если список адресов объектов недвижимости города Красноярска полностью содержится в списке адресов объектов недвижимости Красноярского края. Подчеркнем, что здесь «Адрес объекта недвижимости» – это характеристика общего родового понятия списков, на которую ссылаются их атрибуты с возможно различными названиями.

Признак совместимости списков. Список M совместим по ключу со списком L , если M совместим с L по каждому атрибуту $(K(M), D)$, входящему в потенциальный ключ списка M .

Признак совместимости списков неявно содержит требование эквивалентности потенциальных ключей $K(L)$, $K(M)$ на основании эквивалентности каждого из них одному и тому же набору характеристик $\{F_j(C)\}$, идентифицирующих понятие C . Поэтому совмещение списков можно представлять как взаимную замену ключей в описаниях совместимых списков M и L . Тогда совместимость списков может

быть определена как L-взаимозаменяемость по ключам в терминах Рудольфа Карнапа [8, с. 90]. При этом взаимозаменяемыми по ключам можно считать списки при полном совпадении доменов их ключей.

Лингвистические списки

На практике полностью атомарные списки скорее исключение, чем правило. Поэтому рассмотрим атрибуты списков с неатомарными значениями, т.е. значениями, обладающими некоторой формальной структурой, описанной предложениями на некотором формальном языке. Для этого воспользуемся определением лингвистической переменной, введенной Лотфи Заде [9, с. 71], приспособив его к излагаемому материалу.

Определение 9. Лингвистический атрибут (A_i, G_i, D_i) – это атрибут списка, которому в соответствие поставлена порождающая грамматика Хомского [10, с. 169].

$$G_i = (V_N, V_T, P, S) \quad [11, с. 36],$$

где V_N и V_T – основной и вспомогательный словари (или множество переменных) грамматики G_i соответственно. Объединение V_N и V_T составляет полный словарь $V \equiv V_N \cup V_T$ грамматики G_i , $V_N \cap V_T \equiv \emptyset$;

P – конечное множество правил вывода или правил подстановки, обозначаемых как $\alpha \rightarrow \beta$, где α и β цепочки символов из V , причем α содержит по крайней мере один символ из V_N ;

$S \in V_N$ – начальный символ, к которому начинают применяться правила вывода.

Особенность лингвистического атрибута в том, что он не обязан быть атомарным, а значит для него может выполняться условие $\exists m > 1 \mid D_i \equiv D_{i_1} \times D_{i_2} \times \dots \times D_{i_m}$. Отсюда следует, что предложения языка, порождаемого грамматикой G_i , – это списки строк следующего формата $\{строка_1 \mid строка_2 \mid \dots \mid строка_m\} \in V_T$. В этом случае словарь V_N должен содержать названия $\{A_{i_j} \mid j \leq m\}$ для строк из порожденных грамматикой списков.

Значит, грамматика нашего лингвистического атрибута в результате преобразования его значения порождает набор дополнительных атомарных атрибутов списка $\{(A_{i_j}, D_{i_j}, v_{i_j}) \mid j \leq m\}$ вместе с их значениями.

Определение 10. Производные атрибуты $\{(A_{i_j}, D_{i_j}) \mid j \leq m\}$ лингвистического атрибута (A_i, G_i, D_i) – это атрибуты, порожденные грамматикой G_i . Полный набор атрибутов, порожденных грамматикой G_i , будем называть **расширением лингвистического атрибута**.

Определение 11. Список L – лингвистический список, если наряду с атомарными содержит лингвистические атрибуты. Однородный и лингвистический список L , называется однородным лингвистическим списком.

Совместимость однородных лингвистических списков

Пусть даны два однородных лингвистических списка M и L . Каждый из них может содержать как лингвистические, так и атомарные атрибуты. При этом список M может содержать лингвистический атрибут, значения которого в списке L соответствуют набору значений нескольких атрибутов или наоборот.

Для большей общности дальнейших рассуждений будем считать, во-первых, что расширение атомарного атрибута совпадает с самим этим атрибутом.

Это позволит в дальнейшем рассуждении пользоваться термином *расширение атрибута лингвистического списка* или, если это не приводит к неопределенности, то *расширение атрибута списка*.

Во-вторых, определим эквивалентность списков по расширению атрибута.

Определение 12. Список L эквивалентен списку M по расширению атрибута $\{(A_j, D_j) \mid j \leq m\}$, если совпадают их родовые понятия $C \equiv C(L) = C(M)$ и существует набор характеристик

$\{F_k(C)\}$, который соответствует одному или нескольким базовым или расширенным атрибутам каждого из списков

$$\begin{aligned} & - (\exists \{(A_{i_j}(L), D_{i_j}(L))\} \subseteq \{(A_i(L), D_i(L))\} \wedge \\ & - \{F_k(C)\} \Rightarrow \{(A_{i_j}(L), D_{i_j}(L))\}) \wedge \\ & - (\exists \{(A_{i_j}(M), D_{i_j}(M))\} \subseteq \{(A_j(M), D_j(M))\} \wedge \\ & - \{F_k(C)\} \Rightarrow \{(A_{i_j}(M), D_{i_j}(M))\}). \end{aligned}$$

Признак совместимости по атрибуту. Пусть даны два однородных лингвистических списка M и L . Лингвистический список M совместим с L по расширению атрибута, если

– списки L и M эквивалентны по расширению атрибута;

– домены общего расширения атрибута, $\{(A_j, D_j) \mid j \leq m\}$ обладают следующим свойством: $\forall j \in [1, m] \mid DOMAIN(A_j(M)) \subseteq DOMAIN(A_j(L))$.

Как уже отмечалось, совместимые по расширению атрибута списки L и M не обязаны иметь по лингвистическому атрибуту, которые должны совпадать. Но они должны иметь по совпадающему набору базовых или расширенных атрибутов, соответствующему набору характеристик их общего родового понятия.

Например, пусть M – список объектов недвижимости города Красноярска, а L – список объектов недвижимости Красноярского края. Тогда список M совместим с L по атрибуту «Адрес объекта недвижимости», если

– атрибут или группа атрибутов с общим названием «Адрес объекта недвижимости» обоих списков преобразуется к расширению одной и той же форме, например, $\{\text{район или городской округ края} \mid \text{Населенный пункт} \mid \text{улица} \mid \text{дом}\}$;

– справочники районов (городских округов), населённых пунктов, улиц, номеров домов города Красноярска содержатся в соответствующих справочниках Красноярского края.

Признак совместимости лингвистических списков. Лингвистический список M совместим по ключу со списком L , если M совместим с L по каждому расширению атрибута $(K_{i_j}(M), D_{i_j}(M))$, входящему в потенциальный ключ списка M .

Структурированные списки

В этом разделе будут рассматриваться списки, соответствующие неоднородным понятиям, т.е. понятиям, разделенным в пределах своих списков на дочерние понятия, т.е. виды и подвиды. Примером такого списка может служить список субъектов права, состоящий из записей как о юридических, так и физических лицах.

Как следует из предыдущих рассуждений, неоднородное понятие C – это понятие, полученное путем обобщения нескольких понятий, т.е.

$$\exists m > 1 \mid C \equiv C_1 + C_2 + \dots + C_m, \text{ где «+» операция «обобщения» понятий [6].}$$

Структурированные атомарные списки

Определение 13. Неоднородный атомарный список L – это объединение нескольких как однородных, так и неоднородных атомарных списков.

Лемма 1. Путь образования понятия C неоднородного списка L есть пересечение путей образования понятий, соответствующих подклассам L .

Доказательство.

Утверждение справедливо для однородного списка L , т.к. контексты его записей образуют путь образования родового понятия C , а значит самого списка. Напомним также, что путь образования итоговой записи t_i однородного списка L , совпадает с путем образования самого списка.

Пусть L – объединение только однородных списков, т.е. все подвиды C_i понятия C – однородные. Преобразуем L в L' , каждую группу обыкновенных записей, соответствующую виду C_i итоговой записью. Путь образования каждой итоговой записи равен пересечению путей образования обыкновенных записей, из которых эта запись получена.

По условию построения списка L , вновь образованный список L' окажется однородным, для которого утверждение леммы верно. Но оно верно и для L , т.к. результат применения операций пересечения множеств, составляющих пути образования записей списка, не зависит от порядка выполнения этих операций.

Пусть виды и подвиды родового понятия C списка L образуют дерево высотой $h | h > 2$. И будем считать, что для любого дерева понятий высотой утверждение истинно. Заменяем все обычные неоднородные записи, расположенные на уровне $h - 1$ списка, соответствующими итоговыми записями. Полученному списку L' будет соответствовать дерево понятий высотой $h - 1$, а значит для него условия леммы выполняются. Т.к. предельная вложенность записей h , то все подсписки на уровне $h - 1$ – однородные. Следовательно, пересечение путей образования понятий всех записей будет равно пути образования родового понятия списка L , даже после замены путей образования итоговых записей пересечением путей образования обычных записей подсписков уровня $h - 1$.

В процессе доказательства леммы выяснилось, что виды и подвиды родового понятия списка образуют дерево понятий, т.е. неоднородный атомарный список L обладает структурой, соответствующей структуре подвидов его родового понятия. Структура списка представляется в виде дерева. Поэтому *неоднородные атомарные списки* далее именуются *структурированными атомарными списками*.

Совместимость структурированных атомарных списков

В процессе доказательства леммы в предыдущем разделе использовался прием понижения высоты дерева списка путем замены обыкновенных записей итоговыми.

Далее списки L' , полученные в результате понижения высоты исходного списка L , будем называть **приведенными списками**, а саму операцию понижения высоты – **приведением списка**.

Если над списком L выполнить $h-1$ раз операцию приведения, то полученный список L' станет однородным. Такой приведенный однородный список будем обозначать – L^{hg} .

Признак полусовместимости. Пусть даны структурированные атомарные списки L и M с высотами h_L и h_M соответственно. Будем говорить, что список M полусовместим со списком L , если

– $h_M \leq h_L$;

– $\exists C(\bar{L}) \subseteq C(L) | C(\bar{L}) \equiv C(M)$;

– каждый приведенный однородный подсписок списка M совместим с соответствующим ему приведенным подсписком \bar{L} , следовательно, $\overline{L^{hg}}$ совместим с M^{hg} .

Другими словами, список M полусовместим со списком L , если в последнем можно выделить подсписок, совпадающий с M .

Формулировка признака требует пояснений. Во-первых, эквивалентность деревьев $C(M)$ и $C'(L)$ означает, что они совпадают по высоте. Во-вторых, совпадают структуры деревьев, т.е. совпадают понятия в соответствующих узлах этих деревьев. В-третьих, пути образования обоих деревьев совпадают.

Признак совместимости. Пусть даны структурированные атомарные списки L и M такие, что M полусовместим со списком L , тогда эти списки совместимы, если

- $h_M = h_L$;
- $C(L) \equiv C(M)$.

Интересен также особый случай совместимости структурированных списков, когда совместимость устанавливается между одним из приведенных списков M' и подсписком L , такая совместимость структурированных списков будет называться частичной совместимостью или полусовместимостью.

Заключение

Рассмотренные в статье формальные определения однородных и структурированных списков, содержащих атомарные и лингвистические атрибуты, а также признаки их совместимости могут лечь в основу разработки стандартов автоматизированного обмена данными между информационными системами. В частности, определения однородных и структурированных списков могут использоваться в процессе разработки обменных форматов, а также сервисов обмена данными между различными информационными системами.

В качестве направлений для будущего исследования управленческих списков можно назвать:

- построение следующих категорий управленческих списков, в частности, списки с нечеткими значениями атрибутов [9];
- разработка вопросов соединимости управленческих списков с классификаторами и справочниками;
- разработка формальных грамматик для различных лингвистических атрибутов;
- разработка методики преобразования управленческих списков к отношениям Кодда.

Список литературы

1. Кодд, Е.Ф. Реляционная модель данных для больших совместно используемых банков данных // Системы Управления Базами Данных. – 1995. – № 1. – С. 145–150.
2. Дейт, К.Дж. Введение в системы баз данных / 8-е издание; перевод с английского. – Москва: Вильямс, 2005. – 1328 с.
3. Потанов, А.С. Распознавание образов и машинное восприятие: Общий подход на основе принципа минимальной длины описания. – Санкт-Петербург: Политехника, 2011. – 548 с.
4. Цаленко, М.Ш. Моделирование семантики баз данных. – Москва: Наука, 1989. – 288 с.
5. Войшвилло, Е.К. Понятие как форма мышления: логико-гносеологический анализ. – Москва: Изд-во МГУ, 1989. – 239 с.
6. Гладков, С.Л. О требованиях к интеллектуальной модели данных // Образовательные ресурсы и технологии. – 2015'2(10). – С. 63–70.
7. Джон М. Смит, Диана К. Смит. Абстракция баз данных: Агрегация и обобщение // СУБД. – 1996. – № 2. – С. 141–160.
8. Карнап, Р. Значение и необходимость. Исследование по семантике и модальной логике. – Москва: Изд. иностр. литературы, 1959. – 384 с.
9. Заде, Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений / перевод с английского Н.И. Ринго; под редакцией Н.Н. Моисеева и С.А. Орловского. – Москва: Мир, 1976. – 165 с.
10. Noam Chomsky. On certain formal properties of grammars // Information and Control. – Vol. 2. – Issue 2, 1959. – Pp. 137–167.
11. Фу Кинг-Сун. Структурные методы в распознавании образов / К. Фу; перевод с английского Н.В. Завалишина [и др.]; под редакцией М.А. Айзермана. – Москва: Мир, 1977. – 319 с.

Reference

1. *Kodd, E.F.* Relyacionnaya model' dannyh dlya bol'shih sovmestno ispol'zuemyh bankov dannyh // *Sistemy Upravleniya Bazami Dannyh*. – 1995. – № 1. – S. 145–150.
2. *Dejt, K.Dzh.* Vvedenie v sistemy baz dannyh / 8-e izdanie; perevod s anglijskogo. – Moskva: Vil'yams, 2005. – 1328 s.
3. *Potapov, A.S.* Raspoznavanie obrazov i mashinnoe vospriyatie: Obschij podhod na osnove principa minimal'noj dliny opisaniya. – Sankt-Peterburg: Politehnika, 2011. – 548 s.
4. *Calenko, M.Sh.* Modelirovanie semantiki baz dannyh. – Moskva: Nauka, 1989. – 288 s.
5. *Vojshvillo, E.K.* Ponyatie kak forma myshleniya: logiko-gnoseologicheskij analiz. – Moskva: Izd-vo MGU, 1989. – 239 s.
6. *Gladkov, S.L.* O trebovaniyah k intellektual'noj modeli dannyh // *Obrazovatel'nye resursy i tekhnologii*. – 2015'2(10). – S. 63–70.
7. *Dzhon M. Smit, Diana K. Smit.* Abstrakciya baz dannyh: Agregaciya i obobshchenie // *SUBD*. – 1996. – № 2. – S. 141–160.
8. *Karnap, R.* Znachenie i neobhodimost'. Issledovanie po semantike i modal'noj logike. – Moskva: Izd. inostr. literatury, 1959. – 384 s.
9. *Zade, L.A.* Ponyatie lingvisticheskoj peremennoj i ego primenenie k prinyatiyu priblizhennyh reshenij / perevod s anglijskogo N.I. Ringo; pod redakciej N.N. Moiseeva i S.A. Orlovskogo. – Moskva: Mir, 1976. – 165 s.
10. *Noam Chomsky.* On certain formal properties of grammars // *Information and Control*. – Vol. 2. – Issue 2, 1959. – Pr. 137–167.
11. *Fu King-Sun.* Strukturnye metody v raspoznavanii obrazov / K. Fu; perevod s anglijskogo N.V. Zavalishina [i dr.]; pod redakciej M.A. Ajzermana. – Moskva: Mir, 1977. – 319 s.