УДК 519.681

LP-СТРУКТУРЫ ДЛЯ ОБОСНОВАНИЯ РЕФАКТОРИНГА В ОБЪЕКТНО-ОРИЕНТИРОВАННОМ ПРОГРАММИРОВАНИИ

М. Д. Шурлин

Воронежский государственный университет

Поступила в редакцию 31.08.2012 г.

Аннотация. В индустрии программного обеспечения важную роль играет разработка формальных моделей программируемых объектов. Такие модели создают основу для автоматизированной верификации и оптимизации программного кода. В настоящей работе рассматривается класс основанных на решетках алгебраических структур, описывающих семантику иерархии типов в объектно-ориентированной программной системе. Исследуются свойства таких структур, включая замкнутость и эквивалентность преобразований. Методология предназначена для верификации и модернизации иерархий типов, важным направлением которой является автоматизированное устранение избыточности кода. Впервые приводятся полные доказательства полученных результатов.

Ключевые слова: иерархия типов, рефакторинг, алгебраическая модель, совмещение атрибутов.

Annotation. In the software industry an important direction is connected with the development of formal models for programming objects. Such models provide a basis for automated verification and optimization of the program code. In this paper a class of lattice-based algebraic structures describing semantics of a type hierarchy in an object-oriented program system is considered. The properties of such structures, including closedness and equivalent transformations are studied. The methodology is designed to verify and upgrade type hierarchies and is focused on automatic elimination of code redundancy. Complete proofs of the results obtained are given for the first time.

Keywords: type hierarchy, refactoring, algebraic model, common attributes.

ВВЕДЕНИЕ

Алгебраические системы предоставляют основу формального построения и исследования компьютерных программ, основанных на различных парадигмах [1–2]. Это относится и к логическому программированию, включая широко распространенные на практике системы продукционного типа [3]. В ряде работ (см. [4] и библиографию в ней) предложена методология алгебраизации продукционных систем на основе решеток и отношений. Получены результаты для обоснования эквивалентных преобразований, верификации и оптимизации таких систем. Решетка с заданным на ней дополнительным продукционным отношением названа LP-структурой (lattice production structure).

Исследования показали применимость данной методологии в различных областях информатики. В работе [5] впервые установлено, что отношения обобщения и агрегации типов обладают свойствами продукционно-логического вывода. В результате построен класс LP-структур для моделирования иерархий типов в целях рефакторинга – модернизации кода. В этой модели в LP-структуре из решеточных операций использовалось лишь объединение. Данное обстоятельство ограничило возможности теории формализацией единственного метода рефакторинга – поднятия общих атрибутов (обзор известных методов содержится в [6–7]). Статья [8] развивает теорию LP структур для иерархий типов. В результате «инвертирования» модели [5] предложен новый метод рефакторинга. Его суть состоит в замене нескольких атрибутов класса их общим потомком (совмещение атрибутов).

Настоящая работа развивает алгебраическую модель, предложенную в [8]. Для нового вида LP-структур рассматриваются следующие вопросы: логическое замыкание, его архитектура, эквивалентные преобразования. Впервые представлены полные доказательства полученных результатов.

Решение родственных задач методами анализа формальных понятий (FCA) представлено в [9], где элементам множества классов предлагается в некотором смысле оптимально назначить наборы атрибутов – элементов другого независимого множества. В соответствии с выбранными назначениями формируется иерархия классов. В постановке, которая рассматривается в настоящей работе, в отличие от [9], атрибуты сами относятся к исследуемой иерархии классов (типов), что усложняет задачу и не оставляет возможности непосредственного применения методов FCA.

В разделе 1 приводятся необходимые сведения из теории отношений и решеток. Раздел 2 содержит обсуждение рассматриваемых задач на конкретных примерах, а также их приложений. Данное обсуждение служит иллюстрацией и отправной точкой для построения LP-структуры в разделе 3, где также формулируются и доказываются основные теоретические результаты.

1. ОСНОВНЫЕ ПОНЯТИЯ И ОБОЗНАЧЕНИЯ

Бинарное отношение R на множестве F называется:

• рефлексивным, если для всех $a \in F$ справедливо $(a,a) \in R$;

• транзитивным, если для любых $a,b,c \in F$ из $(a,b),(b,c) \in R$ следует $(a,c) \in R$.

Существует замыкание R^* произвольного отношения R относительно свойств рефлексивности и транзитивности – рефлексивно-транзитивное замыкание (РТЗ) [10]. Пара элементов $a,c \in F$ называется транзитивной в R, если $(a,c) \in R_1^*$, где R_1^* – РТЗ отношения $R_1 = R \setminus \{(a,c)\}$.

Необходимые для чтения статьи сведения о решетках содержатся в [11]. Решеткой называется частично упорядоченное множество $\mathbb{F}$, в котором наряду с отношением $\le$ («не больше», «содержится») определены также две двуместные операции $\wedge$ («пересечение») и $\vee$ («объединение»), вычисляющие соответственно точную нижнюю и верхнюю грани любой пары $a,b \in \mathbb{F}$.

Решетка называется ограниченной, если она содержит общие верхнюю и нижнюю грани – такие два элемента O,I, что $O \le a \le I$ для любого $a \in \mathbb{F}$.

2. ОБСУЖДЕНИЕ ЗАДАЧ И ПРИЛОЖЕНИЙ

Рассмотрим иерархию типов $\mathbb{F}$ в объектно-ориентированной программной системе. Между парами типов могут существовать как минимум два вида связей – наследование (тип наследует атрибуты типа-предка) и агрегация (тип содержит в качестве атрибута представителя другого типа) [6].

Отношение наследования порождает на $\mathbb{F}$ частичный порядок: если тип b является потомком a, то $b \le a$. Требуется, чтобы для любых $a,b \in \mathbb{F}$ были определены две решеточные операции: пересечение $a \wedge b$ – наибольший общий потомок; объединение $a \vee b$ – наименьший общий предок a,b (первая операция актуальна в прикладных системах с множественным наследованием). Для ограниченности решетки добавим к $\mathbb{F}$ два специальных элемента: I – универсальный тип (общий предок, имеется в ряде современных систем программирования) и O – фиктивный потомок всех типов.

На решетке $\mathbb{F}$ рассмотрим второе, соответствующее агрегации, отношение R: если экземпляр типа a в качестве атрибута содержится в определении типа b, то $(b,a) \in R$. Оба отношения ($\le$ и R) имеют общую семантику: в каждом случае, $b \le a$ или $(b,a) \in R$, тип b получает возможности типа a в виде доступа к его атрибутам. Семантически понятно, что это общее отношение «обладания набором возможностей» (обозначим его $\xleftarrow{R}$) обязано быть рефлексивным и транзитивным. Обсудим другие свойства введенных отношений.

Пусть для элементов $a,b_1,b_2 \in \mathbb{F}$ справедливо $b_1 \le a$, $b_2 \le a$. Тогда по определению решетки имеем $b_1 \vee b_2 \le a$. Это естественное для отношения $\le$ свойство (согласно [4]) называется $\vee$-дистрибутивностью. Посмотрим, что будет означать обладание этим же свойством для отношения $\xleftarrow{R}$. Пусть $b_1 \xleftarrow{R} a$ и $b_2 \xleftarrow{R} a$, то есть каждый тип b_1 и b_2 обладает возможностями типа a. Тогда в силу предполагаемой $\vee$-дистрибутивности имеем $b_1 \vee b_2 \xleftarrow{R} a$. Последнее означает, что тип $b_1 \vee b_2$ также обладает возможностями типа a. С точки зрения проектирования типов это не обязательно. Однако, если более одного типа-наследника (в данном случае – b_1 и b_2) содержат одинаковые атрибуты, то, согласно принципу рефакторинга [6], целесообразно «поднять» общие атрибуты, то есть поместить один такой атрибут в общий тип-предок $b_1 \vee b_2$, после чего каждый b_1 и b_2 получит возможности a в порядке наследования. В рассмотренной ситуации $\vee$-дистрибутивность отношения $\xleftarrow{R}$ содержит решение важной задачи – устранение дублирования кода.

Последнее из отмеченных свойств формализовано и исследовано в работе [5]. В частности, показано, что отношение $\xleftarrow{R}$, обладая свойством транзитивности вне зависимости от контекста, не может во всех ситуациях удовлетворять свойству $\vee$-дистрибутивности, так как это приведет к некорректным результатам. Формально описаны ситуации подобных коллизий и построена соответствующая LP-структура с ограниченным свойством $\vee$-дистрибутивности. Принятая стратегия предполагает отказ от «поднятия» атрибутов при невыполнении $\vee$-дистрибутивности. Теоретически возможны и другие подходы, более тонко учитывающие особенности конкретных систем.

Как известно из алгебраической логики [12], «решеточные» операции объединения и пересечения порождают в алгебраических системах двойственные свойства (например, законы де Моргана). Подобная закономерность имеет место и в LP-структурах [4], где наряду с $\vee$-дистрибутивностью отношений присутствует симметричное свойство – $\wedge$-дистрибутивность. Применительно к модели иерархии типов такое свойство впервые рассмотрено в [8]. В настоящей работе уточняются и доказываются утверждения, анонсированные в [8].

Выясним, что означает свойство $\wedge$-дистрибутивности применительно к указанному выше отношению $\xleftarrow{R}$. Предположим, что для элементов $a_1, a_2, b \in \mathbb{F}$ выполнено $b \le a_1$, $b \le a_2$. Тогда по определению решетки $b \le a_1 \wedge a_2$. Это свойство частичного порядка $\le$ на решетке называется $\wedge$-дистрибутивностью [4]. Распространим его на отношение $\xleftarrow{R}$. Пусть $b \xleftarrow{R} a_1$ и $b \xleftarrow{R} a_2$, то есть тип b обладает возможностями типов a_1 и a_2. В силу предполагаемой $\wedge$-дистрибутивности получим $b \xleftarrow{R} a_1 \wedge a_2$. Таким образом, тип b также обладает возможностями типа $a_1 \wedge a_2$. Как и выше, с точки зрения проектирования типов последнее соотношение обязательным не является. Его семантика такова: если тип имеет два или более различных атрибута, то эти атрибуты можно заменить единственным атрибутом, относящимся к ближайшему общему потомку типов исходных атрибутов. Очевидно, что такая реорганизация сделает определение типа-контейнера более компактным с сохранением его функциональности. Этот новый метод рефакторинга (по крайней мере, он отсутствует в известном перечне [6]) назван «совмещением атрибутов». Данный метод актуален лишь для иерархий типов с множественным наследованием. Тем не менее, его существование оправдано популярностью языка C++.

Выясним далее вопрос о том, насколько в данной алгебраической модели типов свойство $\wedge$-дистрибутивности отношения $\xleftarrow{R}$ универсально, и не окажутся ли его ограничения двойственными по отношению к описанным в [5] ограничениям $\vee$-дистрибутивности.

Пример 1. Пусть $b \xleftarrow{R} a$ и $b \xleftarrow{R} b$. При $\wedge$-дистрибутивности отношения $\xleftarrow{R}$ должно быть выполнено $b \xleftarrow{R} b \wedge a$. Согласно принципам объектно-ориентированного программирования, тип b не имеет права что-либо знать о своем типе-наследнике $b \wedge a$. По этой причине тип b, являясь предком типа $b \wedge a$, может обладать его возможностями лишь в случае $b \le a$, иначе соотношение $b \xleftarrow{R} b \wedge a$ окажется некорректным.

Пример 2. Пусть при $b \xleftarrow{R} a_1, b \xleftarrow{R} a_2$ элементы b и $a_1 \wedge a_2$ имеют объединение $b \vee (a_1 \wedge a_2) = d \ne I$, причем $b < d$ и $a_1 \wedge a_2 < d$. Если в данном случае допустить $(b, a_1 \wedge a_2) \in R$, то окажется, что тип b обладает возможностями другого типа d одновременно по двум линиям, а именно – и как его потомок, и как контейнер типа $a_1 \wedge a_2$, также имеющего возможности d в порядке наследования. Недостатки такого кода заключаются в его избыточности – разрыв связи $a_1 \wedge a_2 < d$ не приведет к потере функциональности системы типов.

Пример 3. Пусть $(b, a_1), (b, a_2), (b, a_3) \in R$, причем элементы a_1, a_2, a_3, $a_1 \wedge a_2$, $a_2 \wedge a_3$ попарно различны и $(a_1 \wedge a_2) \vee (a_2 \wedge a_3) = a_2$. Тогда пары $(b, a_1), (b, a_2)$ «конфликтуют» с парами $(b, a_2), (b, a_3)$. Этот факт означает, что если в обоих случаях совместить атрибуты (применив свойство $\wedge$-дистрибутивности), то тип b получит возможности типа a_2 одновременно посредством двух атрибутов – $a_1 \wedge a_2$ и $a_2 \wedge a_3$, что также ухудшит код.

Анализируя примеры 1–3 в сравнении с соответствующими примерами [5], приходим к выводу о двойственном характере ограничений свойств $\wedge$-дистрибутивности и $\vee$-дистрибутивности отношения $\xleftarrow{R}$, которые необходимы для адекватного моделирования иерархий типов. Отмеченный факт служит дополнительным подтверждением естественности разрабатываемых алгебраических моделей, и, в частности, вводимых ограничений дистрибутивности.

На основе представленных выше соображений в следующем разделе определяется логическое бинарное отношение на решетке. Оно отражает свойство «обладания набором возможностей» в иерархии типов. Логическое замыкание произвольного отношения на решетке предоставляет все такие пары (b, a), что в типе b доступны возможности типа a. Решив задачу построения логического замыкания, можно автоматизировать верификацию системы типов. Существование логического замыкания открывает также возможности для автоматизированных эквивалентных преобразований иерархий типов.

3. LP-СТРУКТУРЫ И ИХ ОСНОВНЫЕ СВОЙСТВА

В предыдущем разделе изложены общие идеи формализованного представления иерархии типов на основе LP-структуры. Ниже дается формальное определение соответствующей алгебраической системы.

3.1. ОСНОВНЫЕ ПОНЯТИЯ

Вначале сформулируем некоторые понятия и условия, связанные с ограничением свойства $\wedge$-дистрибутивности продукционно-логических отношений. Как отмечалось в п.2, это свойство описывает совмещение атрибутов в определении типа. В данном случае трудности формализации обусловлены попыткой ввести статическое формальное условие для описания динамического процесса. Более точно – условие $\wedge$-дистрибутивности отношения R на любых подходящих парах (b, a_1), $(b, a_2) \in R$ должно быть выполнено как до совмещения атрибутов, относящихся к типам a_1, a_2, так и после него. В дальнейшем при ссылках на примеры для краткости будут использоваться фразы «совмещение атрибутов a_1, a_2», или просто «совмещение a_1, a_2», если это не вызовет противоречий. Такое тем более возможно потому, что формальные определения данного раздела связаны не с типами, а с элементами абстрактной решетки.

Пара (b, a) элементов ограниченной решетки $\mathbb{F}$ называется простой, если $a \vee b = I$ – верхняя грань $\mathbb{F}$.

Определение 1. Пусть R – бинарное отношение на ограниченной решетке $\mathbb{F}$. Две пары вида (b, a_1), $(b, a_2) \in R$ называются $\wedge$-совместимыми в R, если существуют такие $c_1, c_2 \in \mathbb{F}$, что $c_1 \wedge c_2 \le a_1 \wedge a_2$, причем пара $(b, c_1 \wedge c_2)$ простая, а пары (b, c_1), (b, c_2) нетранзитивны в $R \bigcup \le$. При этом набор элементов $T = (b, c_1, c_2)$ будем называть $\wedge$-дистрибутивной тройкой, а $C = (b, a_1, a_2, c_1, c_2)$ – $\wedge$-дистрибутивным кортежем (в R).

Поясним определение 1 на примере. Предположим, что тип b непосредственно содержит атрибуты a_1 и a_2, причем пара $(b, a_1 \wedge a_2)$ является простой. Тогда, в соответствии с обсуждением п.2, атрибуты a_1 и a_2 могут быть совмещены, то есть, заменены одним атрибутом $a_1 \wedge a_2$. До этого действия имеем $c_1 = a_1, c_2 = a_2$. После совмещения атрибутов условие определения 1 остается выполненным, но с другими промежуточными элементами: $c_1 = c_2 = a_1 \wedge a_2$.

Сформулированное в определении 1 условие нетранзитивности пар (b, c_1), (b, c_2) обеспечивает совмещение атрибутов, относящихся непосредственно к типу b, а не косвенных, связанных с b цепочкой наследований и агрегаций. Конечно, атрибуты неявно совмещаются вместе со своими «возможностями», но именно непосредственные атрибуты должны удовлетворять определению 1.

Следующее понятие формально описывает возможные варианты конфликтов между тройками элементов, претендующими на свойство $\wedge$-дистрибутивности (см. также о конфликте пар в примере 3).

Определение 2. Рассмотрим $\wedge$-дистрибутивную тройку $T = (b, c_1, c_2)$, а также тройку $T' = (b', c_1', c_2')$, тестируемую на аналогичное свойство. Тройка T называется нейтрализующей для T' ($T' \prec T$), если выполнено одно из условий:

1) $b' = b$, $c_1 \wedge c_2 \ne c_1' \wedge c_2'$ и справедливо хотя бы одно из неравенств $c_1 \wedge c_2 < c_i'$;

2) $b' < b$, $c_1 \wedge c_2 \ne c_1' \wedge c_2'$ и выполнено хотя бы одно из неравенств $c_1 \wedge c_2 \le c_i'$;

3) $b' < b$, $c_1 \wedge c_2 = c_1' \wedge c_2'$.

Говоря неформально, данное определение описывает ситуации, когда наличие нейтрализующей тройки T «угрожает» свойству $\wedge$-дистрибутивности тройки T'. Рассмотрим с этой точки зрения условия 1)–3) определения 2.

Пусть имеет место 1). Тогда, после возможного совмещения атрибутов c_1 и c_2 (то есть их замены атрибутом $c_1 \wedge c_2$), в силу соотношений $b' = b$, $(b, c_1 \wedge c_2) \in R$, $c_1 \wedge c_2 < c_i'$ пара (b', c_i') окажется транзитивной в $R \bigcup \le$, что сделает невозможной $\wedge$-дистрибутивность тройки T'.

Условие 1), в частности, содержит конфликт, который иллюстрируется в примере 3 предыдущего раздела работы. Для него справедливы оба соотношения $T' \prec T$ и $T \prec T'$.

Если для T, T' выполнен вариант 2), то после совмещения атрибутов c_1 и c_2 получим соотношения $b' < b$, $(b, c_1 \wedge c_2) \in R$, $c_1 \wedge c_2 < c_i'$, что вновь означает транзитивность в $R \bigcup \leq$ пары (b', c_i') и, соответственно, аннулирует свойство $\wedge$-дистрибутивности тройки T'.

В случае 3) после совмещения атрибутов c_1 и c_2 придем к ситуации, когда обе пары (b', c_i'), $i = 1,2$ окажутся транзитивными в $R \bigcup \leq : b' < b$, $(b,\ c_1 \wedge c_2) \in R$, $c_1 \wedge c_2 = c_1' \wedge c_2' < c_i'$.

Тройку $T' = (b', c_1', c_2')$ будем называть неконфликтной, если для нее не существует ни одной нейтрализующей $\wedge$-дистрибутивной тройки.

Понятия, введенные в определении 2 (и далее) для троек вида $T = (b, c_1, c_2)$ и $T' = (b', c_1', c_2')$, автоматически распространяются на соответствующие им кортежи $C = (b, a_1, a_2, c_1, c_2)$ и $C' = (b', a_1', a_2', c_1', c_2')$. Две $\wedge$-совместимые пары (b, a_1), (b, a_2) называются неконфликтно $\wedge$-совместимыми, если для них существует неконфликтный $\wedge$-дистрибутивный кортеж (b, a_1, a_2, c_1, c_2).

Отношение R на решетке $\mathbb{F}$ называется ограниченно $\wedge$-дистрибутивным, если для любых неконфликтно $\wedge$-совместимых в R пар (b, a_1), (b, a_2) справедливо $(b, a_1 \wedge a_2) \in R$.

Определение 3. Отношение называется логическим с ограничением пересечений (в данной работе – просто логическим), если оно содержит $\leq$, транзитивно и ограниченно $\wedge$-дистрибутивно. Логическим замыканием отношения R называется наименьшее логическое отношение, содержащее R и его множество неконфликтно $\wedge$-совместимых пар.

Два отношения R_1 и R_2, определенные на общей решетке, называются эквивалентными ($R_1 \sim R_2$), если их логические замыкания совпадают.

Для выяснения вопроса о существовании логического замыкания введем следующее понятие логической связи.

Определение 4. Пусть задано произвольное бинарное отношение R на решетке $\mathbb{F}$. Будем констатировать, что упорядоченная пара $b, a \in \mathbb{F}$ логически связана отношением R ($b \xleftarrow{R} a$), если выполнено одно из следующих условий:

1) $(b, a) \in R$;

2) $b \leq a$;

3) существуют такие $a_1, a_2 \in \mathbb{F}$, что $a = a_1 \wedge a_2$, причем $b \xleftarrow{R} a_1, b \xleftarrow{R} a_2$ и пары $(b, a_1), (b, a_2)$ неконфликтно $\wedge$-совместимы;

4) существует элемент $c \in \mathbb{F}$ такой, что $b \xleftarrow{R} c$ и $c \xleftarrow{R} a$.

Условия 1)–4) определения 4 будем также называть правилами (вывода).

При получении некоторой логической связи на основе определения 4 шагом вывода будем называть применение ровно одного правила вывода из этого определения, возможно, одновременно к некоторому конечному множеству элементов решетки. Например, если $b_t \xleftarrow{R} c_t, c_t \xleftarrow{R} a_t$, $t \in T$, то $b_t \xleftarrow{R} a_t$, $t \in T$.

Уровнем рекурсии в соотношении $b \xleftarrow{R} a$ будем называть количество шагов вывода, необходимое для получения этой связи. При этом учитываются лишь применение правил 3)–4). Для связи, основанной только на правилах 1)–2), уровень рекурсии считается равным нулю. Поскольку в общем случае связь $b \xleftarrow{R} a$ может быть получена не единственным набором правил, будем лишь оценивать ее уровень рекурсии, не указывая его точного значения.

Применительно к шагам вывода соотношения $b \xleftarrow{R} a$ можно употреблять слова «начальный», «последний», «предыдущий», «следующий» и так далее. При этом имеется в виду продвижение в направлении прямого логического вывода, то есть от пар исходного отношения ($R \bigcup \leq$) к рассматриваемой паре отношения $\xleftarrow{R}$. Заметим, что рекурсивное определение 4 сформулировано в соответствии с принципом обратного вывода.

3.2. СВОЙСТВА ДИСТРИБУТИВНЫХ ТРОЕК И СОВМЕСТИМЫХ ПАР

Для дальнейшего исследования необходимо установить ряд свойств $\wedge$-дистрибутивных троек и $\wedge$-совместимых пар.

Лемма 1. Пусть $T = (b, c_1, c_2)$ – неконфликтная $\wedge$-дистрибутивная тройка и элемент $c_3 \in \mathbb{F}$ таков, что пары $(b, c_1 \wedge c_3), (b, c_2 \wedge c_3)$ простые, а пара $(b, c_3) \in R$ нетранзитивна в $R \bigcup \leq$. Тогда $c_1 \wedge c_2 \wedge c_3 = c_1 \wedge c_2 = c_2 \wedge c_3 = c_1 \wedge c_3$.

Доказательство. Сразу заметим, что интерес представляет лишь случай, когда $c_3 \neq c_1$, $c_3 \neq c_2$, иначе утверждение леммы тривиально. Рассмотрим тройку $T' = (b, c_2, c_3)$. В силу условия леммы она $\wedge$-дистрибутивна. Кроме того, поскольку $c_2 \wedge c_3 \le c_2$ и пары (b, c_2), (b, c_3) нетранзитивны в $R \bigcup \le$, то $c_2 \wedge c_3 < c_2$. Последнее неравенство означает, что тройка T' может оказаться нейтрализующей для $T = (b, c_1, c_2)$. Чтобы этого не случилось (ведь по условию T неконфликтна!) в силу условия 1) определения 3.2 необходимо выполнение равенства $c_1 \wedge c_2 = c_2 \wedge c_3$, из которого сразу следуют соотношения $c_1 \wedge c_2 \wedge c_3 = c_1 \wedge c_2 = c_2 \wedge c_3$. Недостающее равенство $c_1 \wedge c_2 \wedge c_3 = c_1 \wedge c_3$ можно получить, если вместо T' аналогично рассмотреть тройку (b, c_1, c_3). □

Лемма 2. Пусть R – произвольное бинарное отношение на $\mathbb{F}$ и $T = (b, c_1, c_2)$ – неконфликтная $\wedge$-дистрибутивная тройка. Тогда множество неконфликтно $\wedge$-совместимых пар отношения $R \bigcup \{(b, c_1 \wedge c_2)\}$ совпадает с множеством неконфликтно $\wedge$-совместимых пар исходного отношения R.

Доказательство. Вначале докажем, что в результате этой операции, а именно – добавления $(b, c_1 \wedge c_2)$ к R, исходное множество не сужается. Из определения 1 следует, что множество $\wedge$-совместимых пар полностью определяется множеством нетранзитивных пар отношения $R \bigcup \le$. Таким образом, достаточно рассмотреть те пары нового отношения $R \bigcup \{(b, c_1 \wedge c_2)\}$, нетранзитивность которых может нарушиться в результате присоединения к R пары $(b, c_1 \wedge c_2)$ (например, (b', c') при $c_1 \wedge c_2 \le c'$ и $b' \le b$).

Покажем, что любые неконфликтно $\wedge$-совместимые пары $(b', d_1), (b', d_2)$ останутся таковыми после добавления к R указанной в лемме пары $(b, c_1 \wedge c_2)$. Соответствующий парам $(b', d_1), (b', d_2)$ неконфликтный $\wedge$-дистрибутивный кортеж обозначим $C' = (b', d_1, d_2, c_1', c_2')$. Согласно определению 1, справедливо $c_1' \wedge c_2' \le d_1 \wedge d_2$, пара $(b', c_1' \wedge c_2')$ простая и пары $(b', c_1'), (b', c_2') \in R$ нетранзитивны в $R \bigcup \le$. Их нетранзитивность может нарушиться в результате добавления пары $(b, c_1 \wedge c_2)$ к отношению R. Покажем, как в этом случае заменить соответствующие элементы c_1', c_2' на другие $\tilde{c}_1, \tilde{c}_2$, чтобы для пар $(b', d_1), (b', d_2)$ вновь выполнялись все требования определения 1.

Рассмотрим вначале случай $c_1 \wedge c_2 = c_1' \wedge c_2'$. Поскольку обе тройки $T = (b, c_1, c_2)$ и $T' = (b', c_1', c_2')$ неконфликтны, то согласно условию 3) определения 2 имеем $b = b'$. Пусть при этом $c_1 \wedge c_2 < c_i'$ ($i = 1$ или $i = 2$). Тогда после добавления $(b, c_1 \wedge c_2)$ пара (b', c_i') становится транзитивной в $R \bigcup \le$. Обозначив $\tilde{c}_i = c_1 \wedge c_2$, заметим, что пара $(b', \tilde{c}_i)$ нетранзитивна в $R \bigcup \le$. Если же неравенство $c_1 \wedge c_2 < c_i'$ не выполнено, то пара (b', c_i') остается нетранзитивной, и можно выбрать $\tilde{c}_i = c_i'$. Проверим, что выбранные элементы $\tilde{c}_1, \tilde{c}_2$ наряду с нетранзитивностью пар $(b', \tilde{c}_1), (b', \tilde{c}_2) \in R$ обеспечивают и остальные требования $\wedge$-совместимости пар $(b', d_1), (b', d_2)$.

Во-первых, по построению $\tilde{c}_i$ справедливо $\tilde{c}_i \le c_i'$, отсюда имеем $\tilde{c}_1 \wedge \tilde{c}_2 \le c_1' \wedge c_2'$. Из этого неравенства и соотношения $c_1' \wedge c_2' \le d_1 \wedge d_2$ следует $\tilde{c}_1 \wedge \tilde{c}_2 \le d_1 \wedge d_2$, что и требуется в определении 1. Покажем далее, что $\tilde{c}_1 \wedge \tilde{c}_2 = c_1' \wedge c_2'$. Поскольку рассматривается вариант $c_1 \wedge c_2 = c_1' \wedge c_2'$, то из построения $\tilde{c}_i$ ($\tilde{c}_i = c_i'$ либо $\tilde{c}_i = c_1 \wedge c_2$) следует, что $c_1' \wedge c_2' \le \tilde{c}_i$, $i = 1, 2$. Отсюда вытекает соотношение $c_1' \wedge c_2' \le \tilde{c}_1 \wedge \tilde{c}_2$. С другой стороны, как было замечено в начале абзаца, $\tilde{c}_1 \wedge \tilde{c}_2 \le c_1' \wedge c_2'$. В итоге имеем $\tilde{c}_1 \wedge \tilde{c}_2 = c_1' \wedge c_2'$. Это равенство означает, что пара $(b', \tilde{c}_1 \wedge \tilde{c}_2)$ является простой, так как пара $(b', c_1 \wedge c_2)$ простая по определению. Таким образом, для элементов $\tilde{c}_1, \tilde{c}_2$ выполнены все требования определения 1.

Рассмотрим теперь вариант $c_1 \wedge c_2 \ne c_1' \wedge c_2'$ и в нем потенциально опасный случай $b' \le b$. При этом из-за неконфликтности тройки T' в силу условий 1)–2) определения 2 невозможно выполнение неравенств вида $c_1 \wedge c_2 \le c_i'$. Рассматривая условия 1)–2) отдельно, замечаем, что в каждом из вариантов сохраняется нетранзитивность пар (b', c_i'), $i = 1, 2$, а вместе с ней и $\wedge$-совместимость пар $(b', d_1), (b', d_2)$. Таким образом, при исходном неравенстве $c_1 \wedge c_2 \ne c_1' \wedge c_2'$ кортеж C' не меняет своих свойств, то есть можно выбрать $\tilde{c}_i = c_i'$, $i = 1, 2$.

Показано, что после добавления $(b, c_1 \wedge c_2)$ к R, пары $(b', d_1), (b', d_2)$ сохраняют свойство $\wedge$-совместимости. Для доказательства необходимого вложения множеств осталось проверить, что сохраняется и их свойство неконфликтности. Достаточно показать, что для тройки $\tilde{T} = (b', \tilde{c}_1, \tilde{c}_2)$ нет ни одной нейтрализующей $\wedge$-дистрибутивной тройки $N = (p, n_1, n_2)$. С этой целью предположим противное, а именно – что она существует ($\tilde{T} \prec N$), и при этом проанализируем условия 1)–3) определения 2. Учитывая полученные выше соотношения $\tilde{c}_i \le c_i'$ и $\tilde{c}_1 \wedge \tilde{c}_2 = c_1' \wedge c_2'$, нетрудно заметить, что для $T' = (b', c_1', c_2')$ также выполнено $T' \prec N$. Если при этом тройка N (вместе со своими свойствами) существовала до добавления пары $(b, c_1 \wedge c_2)$

к R, то имеем противоречие, так как тройка T' изначально была неконфликтной. Рассмотрим ситуацию, когда сама тройка (p, n_1, n_2) или ее свойства (пара $(p, n_1 \wedge n_2)$ простая, а пары (p, n_1), $(p, n_2) \in R$ нетранзитивны в $R \cup \le$) появились в результате включения $(b, c_1 \wedge c_2)$ в отношение R. Очевидно, что свойство нетранзитивности любой пары не может *возникнуть* в результате добавления к отношению другой пары, равно как и свойство простоты фиксированной пары элементов решетки. Таким образом, единственно возможным вариантом остается лишь появление самой тройки (p, n_1, n_2), когда $p = b$ и $n_j = c_1 \wedge c_2$ ($j = 1$ либо $j = 2$). Положим для определенности $j = 1$, то есть $N = (b, c_1 \wedge c_2, n_2)$ (альтернативный случай симметричен). Далее заметим, что поскольку пара $(b, c_1 \wedge c_2 \wedge n_2)$ простая, то для тройки $T = (b, c_1, c_2)$ и пары (b, n_2) выполнены условия леммы 1. В силу этой леммы $c_1 \wedge c_2 \wedge n_2 = c_1 \wedge c_2$. С учетом последнего равенства рассмотрим применительно к $\tilde{T} \prec N$ каждое из условий 1)–3) определения 2 в отдельности.

Пусть $b = b'$, $c_1 \wedge c_2 \wedge n_2 \ne \tilde{c}_1 \wedge \tilde{c}_2$ и справедливо одно из неравенств $c_1 \wedge c_2 \wedge n_2 < \tilde{c}_i$, то есть имеет место условие 1). Тогда выполнены и соотношения $c_1 \wedge c_2 \wedge n_2 = c_1 \wedge c_2 < c_i'$. Кроме того, в рассматриваемом варианте $c_1 \wedge c_2 \ne \tilde{c}_1 \wedge \tilde{c}_2 = c_1' \wedge c_2'$. Получено, что изначально тройка $T = (b, c_1, c_2)$ была нейтрализующей для тройки $T' = (b', c_1', c_2')$, что противоречит условию неконфликтности T'.

Рассмотрим следующий вариант, когда тройка $N = (b, c_1 \wedge c_2, n_2)$ является нейтрализующей для $\tilde{T}$ в силу условия 2). Согласно этому условию, $b' < b$, $c_1 \wedge c_2 \wedge n_2 \ne \tilde{c}_1 \wedge \tilde{c}_2$ и выполнено хотя бы одно из неравенств $c_1 \wedge c_2 \wedge n_2 \le \tilde{c}_i$. Тогда, как и в предыдущем случае, справедливо $c_1 \wedge c_2 \le c_i'$ и $c_1 \wedge c_2 \ne c_1' \wedge c_2'$, то есть $T' \prec T$ и имеем аналогичное противоречие.

Наконец, если $b' < b$, $c_1 \wedge c_2 \wedge n_2 = \tilde{c}_1 \wedge \tilde{c}_2$, то и $c_1 \wedge c_2 = c_1' \wedge c_2'$. И в данном случае оказывается, что $T' \prec T$ было выполнено до добавления $(b, c_1 \wedge c_2)$ к R, чего не могло быть по условию леммы.

Полученные противоречия показывают, что в результате добавления $(b, c_1 \wedge c_2)$ к R произвольные $\wedge$-совместимые пары $(b', d_1), (b', d_2)$ сохраняют свое свойство неконфликтности. Таким образом, доказано, что в результате данной операции исходное множество неконфликтно $\wedge$-совместимых пар не сужается. Для полного доказательства леммы 2 осталось установить, что оно и не расширяется. С этой целью вначале покажем, что расширение было бы возможным лишь за счет единственной новой $\wedge$-дистрибутивной тройки $T_\wedge = (b, c_1 \wedge c_2, c_1 \wedge c_2)$. Предположим противное, то есть что после добавления $(b, c_1 \wedge c_2)$ в R появились новые неконфликтно $\wedge$-совместимые пары $(b', d_1), (b', d_2)$ с соответствующей им тройкой $T' = (b', c_1', c_2')$, не совпадающей с $T_\wedge$. Она не могла существовать до добавления $(b, c_1 \wedge c_2)$ к R. При этом, как и выше, единственно возможным вариантом является лишь появление самой тройки (b', c_1', c_2'), когда $b' = b$ и $c_1 \wedge c_2 = c_j'$ ($j = 1$ или $j = 2$). Снова не ограничивая общности можно считать, что $j = 1$, то есть $T' = (b, c_1 \wedge c_2, c_2')$. Кроме того, в силу сделанного предположения имеем $c_2' \ne c_1 \wedge c_2$. Так как пара $(b, c_1 \wedge c_2 \wedge c_2')$ простая, то для тройки $T = (b, c_1, c_2)$ и пары (b, c_2') выполнены условия леммы 1, по которой имеем $c_1 \wedge c_2 \wedge c_2' = c_1 \wedge c_2$, то есть $c_1 \wedge c_2 < c_2'$. Последнее неравенство приводит к противоречию, так как из него следует транзитивность пары (b, c_2') в $R \cup \le$, чего не может быть по определению дистрибутивной тройки T'.

Таким образом, в результате описанной в лемме операции появилась единственная новая $\wedge$-дистрибутивная тройка $T_\wedge = (b, c_1 \wedge c_2, c_1 \wedge c_2)$. Однако, ее появление не порождает новых дистрибутивно совместимых пар $(b, a_1), (b, a_2)$, поскольку неравенство $c_1 \wedge c_2 = (c_1 \wedge c_2) \wedge (c_1 \wedge c_2) \le b_1 \wedge b_2$ было бы выполнено и ранее для тройки $T = (b, c_1, c_2)$.

Итак, лемма 2 полностью доказана. □

Замечание 1. Пользуясь определениями 1–3, легко показать, что множество $\wedge$-совместимых пар (с их конфликтами) определяется множеством нетранзитивных пар отношения $R \cup \le$. Отсюда следует инвариантность множества неконфликтно $\wedge$-совместимых пар отношения R относительно еще одной операции – расширения R за счет транзитивных пар отношения $R \cup \le$.

Лемма 3. Множество неконфликтно $\wedge$-совместимых пар отношения $\xleftarrow{R}$ совпадает с таковым множеством базового отношения R.

Доказательство. Достаточно установить, что при построении $\xleftarrow{R}$ из отношения R за счет правил 2)–4) определения 4 указанное в лемме множество сохраняется. Для правил 2) и 4) подлежащее доказательству утверждение сразу следует из замечания 1. Покажем, что аналогичное верно и для правила 3).

Пусть пары $(b,a_1),(b,a_2)$ неконфликтно $\wedge$-совместимы. Рассмотрим соответствующий им неконфликтный кортеж $C=(b,a_1,a_2,c_1,c_2)$. Как отмечалось выше, пары $(b,c_1),(b,c_2)$ также неконфликтно $\wedge$-совместимы, поэтому отношение $\xleftarrow{R}$ должно содержать и пару $(b,c_1\wedge c_2)$. При этом, по лемме 2, операция добавления к R пары $(b,c_1\wedge c_2)$ сохраняет его множество неконфликтно совместимых пар.

Рассматривая далее пары $(b,c_1\wedge c_2),(c_1\wedge \wedge c_2,a_1\wedge a_2)$, замечаем, что пара $(b,a_1\wedge a_2)$ транзитивна в $R\bigcup\le$ при новом (расширенном) отношении R. Эта пара также должна быть добавлена к R по правилам 2) и 4). При этом, как указано в начале доказательства, исходное множество неконфликтно совместимых пар расширяемого отношения R не изменится. □

3.3. ЛОГИЧЕСКОЕ ЗАМЫКАНИЕ И ПРЕОБРАЗОВАНИЯ

В настоящем подразделе исследуются вопросы о существовании и архитектуре логического замыкания рассматриваемой LP-структуры, а также ее эквивалентных преобразованиях.

Лемма 4. При произвольном R отношение $\xleftarrow{R}$ является логическим.

Доказательство. Необходимо проверить, что данное отношение содержит $\le$, транзитивно и ограниченно $\wedge$-дистрибутивно. Первые два указанных свойства непосредственно следуют из п.2) и п.4) определения 4. Свойство ограниченной $\wedge$-дистрибутивности требует более подробного рассмотрения.

Пусть $b,a_1,a_2\in\mathbb{F}$, $b\xleftarrow{R}a_1, b\xleftarrow{R}a_2$ и пары $(b,a_1),(b,a_2)$ неконфликтно $\wedge$-совместимы в $\xleftarrow{R}$. Необходимо доказать, что тогда $b\xleftarrow{R}a_1\wedge a_2$. С этой целью покажем, что при сделанных предположениях пары $(b,a_1),(b,a_2)$ неконфликтно $\wedge$-совместимы в R. Тогда подлежащее доказательству утверждение сразу будет следовать из п.3) определения 4.

Итак, по определению 1 имеем: существуют такие $c_1,c_2\in\mathbb{F}$, что $c_1\wedge c_2\le a_1\wedge a_2$, пара $(b,c_1\wedge c_2)$ простая, $b\xleftarrow{R}c_1, b\xleftarrow{R}c_2$ и пары $(b,c_1),(b,c_2)$ нетранзитивны в $\xleftarrow{R}$ (поскольку $(R\bigcup\le)\subseteq\xleftarrow{R}$, то и в $R\bigcup\le$). Последнее означает, что соотношения $b\xleftarrow{R}c_1, b\xleftarrow{R}c_2$ были получены без применения п.4) определения 4. Простота пары $(b,c_1\wedge c_2)$ означает также, что не использовался и п.2). Таким образом, из п.1) и п.3) определения 4 следует, что для пар $(b,c_1),(b,c_2)$ остается лишь такой вариант, когда $c_1=c_1^1\wedge\ldots\wedge c_1^{m_1}$, $c_2=c_2^1\wedge\ldots\wedge c_2^{m_2}$, где $(b,c_j^{k_j})\in R$, $k_j=1,\ldots,m_j$, $j=1,2$, причем все пары $(b,c_j^{k_j})$ нетранзитивны в $R\bigcup\le$.

Рассмотрим вначале случай $m_1=m_2=1$. Тогда $(b,c_1),(b,c_2)\in R$, и лемма уже доказана (неконфликтность тройки $T=(b,c_1,c_2)$ в R следует из ее неконфликтности в большем множестве $\xleftarrow{R}$).

Предположим теперь, что $m_1>1$. В этом случае тройка $T'=(b,c_1^1,c_1^2)$ неконфликтна. Поскольку пара $(b,c_1\wedge c_2)$ простая, то для T и любой оставшейся пары вида $(b,c_j^{k_j})$ выполнены условия леммы 1. По ее утверждению справедливо равенство $c_1^1\wedge c_1^2\wedge c_j^{k_j}=c_1^1\wedge c_1^2$. Складывая (в смысле операции пересечения $\wedge$) такие равенства по всем $k_j=1,\ldots,m_j$, $j=1,2$, получим $c_1\wedge c_2=c_1^1\wedge c_1^2$. В итоге приходим к следующему факту: существуют такие $c_1^1,c_1^2\in\mathbb{F}$, что $c_1^1\wedge c_1^2\le a_1\wedge a_2$, пара $(b,c_1^1\wedge c_1^2)$ простая, пары $(b,c_1^1),(b,c_1^2)\in R$ нетранзитивны в $R\bigcup\le$, тройка $T'=(b,c_1^1,c_1^2)$ неконфликтна в R. Он означает, что пары $(b,a_1),(b,a_2)$ неконфликтно $\wedge$-совместимы в R. Отсюда в силу п.3) определения 4 следует $b\xleftarrow{R}a_1\wedge a_2$, то есть отношение $\xleftarrow{R}$ ограниченно $\wedge$-дистрибутивно. □

Лемма 5. Пусть R – некоторое логическое отношение на решетке $\mathbb{F}$ и $a,b\in\mathbb{F}$. Тогда, если справедливо $b\xleftarrow{R}a$, то $(b,a)\in R$.

Доказательство. Проведем его с помощью индукции по m – оценке уровня рекурсии в соотношении $b\xleftarrow{R}a$. При $m=0$ имеет место одно из условий 1)–2) определения 4. Случай 1) непосредственно означает справедливость доказываемого утверждения. Если же справедливо 2), то и тогда $(b,a)\in R$, поскольку логическое отношение R содержит и $\le$.

Предположим далее, что лемма верна для некоторого $m\ge 0$ и докажем ее утверждение при уровне рекурсии $m+1$. В этом случае новые для рассмотрения варианты могут дать условия 3)–4) определения 4.

Рассмотрим вариант, когда соотношение $b\xleftarrow{R}a$ происходит из условия 3) определения 4. При этом по предположению индукции соотношения $b\xleftarrow{R}a_1, b\xleftarrow{R}a_2$ имеют уровень рекурсии $\le m$, то есть $(b,a_1),(b,a_2)\in R$. Тогда, в силу ограниченной $\wedge$-дистрибутивности R, получим $(b,a)\in R$.

Наконец, пусть справедливо 4). И в этом случае по предположению индукции базовые соотношения $b \xleftarrow{R} c$ и $c \xleftarrow{R} a$ имеют уровень рекурсии $\leq m$. Следовательно, $(b,c),(c,a) \in R$. Отсюда, в силу транзитивности логического отношения R, имеем $(b,a) \in R$. □

Теорема 1. Для произвольного отношения R на решетке $\mathbb{F}$ логическое замыкание существует и совпадает с множеством $\xleftarrow{R}$ всех упорядоченных пар, логически связанных отношением R.

Доказательство. В силу п.1) определения 4 отношение $\xleftarrow{R}$ содержит R. По лемме 3 оно также содержит и множество неконфликтно $\wedge$-совместимых в R пар. Из леммы 4 следует, что $\xleftarrow{R}$ является логическим. Осталось показать, что это – наименьшее из таковых отношений.

Пусть R' – другое логическое отношение, содержащее R и его множество неконфликтно совместимых пар. Тогда очевидно, что если $b \xleftarrow{R} a$, то $b \xleftarrow{R'} a$. Отсюда по лемме 5 имеем $(b,a) \in R'$. Следовательно, отношение $\xleftarrow{R}$ содержится в произвольно выбранном R', и, таким образом, является наименьшим логическим отношением, содержащим R и его множество неконфликтно совместимых пар. □

Далее рассмотрим вопросы, связанные с эквивалентными преобразованиями логических структур на решетках типов. Пусть дано отношение R. Его эквивалентным преобразованием называется замена множества упорядоченных пар R, в результате которой полученное новое отношение P эквивалентно R.

В работе [4] для стандартных LP-структур показано, что локально-эквивалентное преобразование части исходного множества R приводит к логически эквивалентному общему отношению P. Для рассматриваемых в настоящей работе LP-структур это в общем неверно. Например, если при $T' \prec T$ преобразовать (эквивалентно, на основе $\wedge$-дистрибутивности) две пары из T', общий результат окажется неэквивалентным. Однако имеет место эквивалентность ряда конкретных видов преобразований.

Теорема 2. Пусть R – отношение на решетке $\mathbb{F}$. Тогда каждая из следующих операций над R приводит к эквивалентному отношению:

• добавление или исключение пары (b,a), если $b \leq a$;

• добавление пары $(b, c_1 \wedge c_2)$, если тройка $T = (b, c_1, c_2)$ $\wedge$-дистрибутивна и неконфликтна в R;

• добавление или исключение пары (c,a) при наличии пар $(c,b),(b,a) \in (R \cup \leq)$, $c \neq b$, $b \neq a$.

Доказательство непосредственно следует из определения 4 и доказанных выше лемм. □

4. ЗАКЛЮЧЕНИЕ

Представленный формализм позволяет проводить автоматизированные исследования иерархий типов, включая эквивалентные преобразования, верификацию и оптимизацию. Он может служить основой для практической реализации (или модернизации) типов.

СПИСОК ЛИТЕРАТУРЫ

1. *Подловченко Р. И.* Иерархия моделей программ / Р. И. Подловченко // Программирование. – 1981. – № 2. – С. 3–14.

2. *Замулин А. В.* Алгебраическая семантика императивного языка программирования / А. В. Замулин // Программирование. – 2003. – № 6. – С. 51–64.

3. *Davis R.* An overview of production systems / R. Davis, J. King // Mahine Intelligence. – Chichester : Ellis Horwood Limited. – 1977. – Vol 8. – P. 300–332.

4. *Махортов С. Д.* Алгебраический подход к исследованию и оптимизации баз знаний продукционного типа / С. Д. Махортов, С. Л. Подвальный // Информационные технологии. – 2008. – № 8. – С. 55–60.

5. *Махортов С. Д.* LP-структуры на решетках типов и некоторые задачи рефакторинга / С. Д. Махортов // Программирование. – 2009. – Т. 35, № 4. – С. 5–14.

6. *Фаулер М.* Рефакторинг: улучшение существующего кода : пер. с англ. / М. Фаулер. – СПб. : Символ-Плюс, 2004. – 432 с.

7. *Mens T.* A Survey of Software Refactoring / T. Mens, T. Tourw'e // IEEE Trans. on Software Engineering. – Feb. 2004. – Vol. 30(2). – P. 126–139.

8. *Махортов С. Д.* LP-структуры для обоснования и автоматизации рефакторинга / С. Д. Махортов // Программная инженерия. – 2010, № 2. – С. 13–25.

9. *Godin R.* Formal Concept Analysis-Based Class Hierarchy Design in Object-Oriented Software Development / R. Godin, P. Valtchev // Formal Concept Analysis / eds. B. Ganter, G. Stumme, R. Wille // Lecture Notes In Computer Science. – Springer Berlin/Heidelberg. – 2005. – V. 3626. – P. 304–323.

10. *Aho A. V.* The transitive reduction of a directed graph. SIAM J. Computing 1 : 2 / A. V. Aho, M. R. Garey, J. D. Ulman, 1972. – P. 131–137.

11. *Биркгоф Г.* Теория решеток : пер. с англ. / Г. Биркгоф. – М. : Наука, 1984. – 568 с.

12. *Расёва Е.* Математика метаматематики : пер. с англ. / Е. Расёва, Р. Сикорский. – М. : Наука, 1972. – 591 с.

Шурлин Максим Дмитриевич – преподаватель кафедры математического обеспечения ЭВМ Воронежского государственного университета. E-mail: mshurlin@gmail.com

Shurlin Maxim D. – Holding a part-time position of lecturer at Applied and System Software Department at Voronezh State University. E-mail: mshurlin@gmail.com