УДК: 681.3.062

МУЛЬТИМНОЖЕСТВА: ОБЗОР БИБЛИОГРАФИИ, ПОСТРОЕНИЕ РЕШЕТКИ МУЛЬТИМНОЖЕСТВ

Ю.А. Богатырева

Киевский национальный университет имени Тараса Шевченко,
03680, Киев, проспект Академика Глушкова 2, корп. 6, (044)521 3345, j_bogatyreva@ukr.net

Дается обзор современной библиографии по теории мультимножеств и ее применениям; строится решетка мультимножеств.

The review of the modern bibliography of the multisets' theory and its applications is given; the multisets' lattice is constructed.

Обзор современной библиографии по мультимножествам

Актуальность данной работы обусловлена необходимостью решения прикладных задач, особенностью которых являются множественность и повторяемость данных. Для решения таких задач в качестве математического объекта используют мультимножества (multisets, bags).

Обзор современной литературы по мультимножествам показал, что исследования, посвященные этой тематике, можно условно разделить на два вида: работы по теории мультимножеств и работы, связанные с применением мультимножеств в различных прикладных областях.

Теорию мультимножеств рассматривали в своих работах Дж. Альберт (J. Albert) [1], В. Близард (W. Blizard) [2], А. Сиропоулос (A. Syropoulos) [3], А.Б. Петровський [4], а также украинские исследователи В.Н. Редько, Ю.И. Брона, Д.Б. Буй, С.А. Поляков [5].

В. Близард в своей работе [2] представил развернутый обзор развития теории мультимножеств, различные определения мультимножества, а также некоторые их специфические применения. В работе излагаются основные идеи и достижения, полученные исследователями в этом направлении. Отмечено при этом, что у различных авторов понятие мультимножества возникает под разными названиями.

Одним из первых, кто обратил внимание на то, что существует необходимость рассмотрения мультимножества как отдельного математического объекта, был Д. Кнут. В своей книге [6] он дает содержательное определение мультимножеству и определяет операции объединения, пересечения и сложения мультимножеств.

Дж. Альберт в [1] не только дает определения мультимножествам и операций над ними, но и представляет некоторые результаты, относящиеся к алгебраическим свойствам мультимножеств.

В работе [3] А. Сиропоулос систематизирует все то, что имеет отношение к мультимножествам: дает определение мультимножеству и операций над ними, описывает так называемые гибридные множества, мультимножества в теории категорий, нечеткие и частично-упорядоченные мультимножества.

А.Б. Петровский в своей монографии [4] вводит основные определения, относящиеся к теории мультимножеств: определение мультимножества, его характеристической функции, операций над ними. Кроме того он рассматривает некоторые свойства основных операций над мультимножествами, методы графического представления и краткий обзор применений их в различных областях. Однако следует заметить, что изложенная теория требует внесения уточнений и дополнений. Автор продолжает свое исследования в книге [7], в которой рассматриваются метрические пространства множеств (мультимножеств), устанавливаются их основные свойства мер, описываются новые типы пространств измеримых, а также новые виды метрик.

Определение мультимножества в терминах табличных баз данных приведено в книге Г. Гарсия-Молина и др. [8]. Мультимножество рассматривается как совокупность кортежей с возможными повторениями. Над мультимножествами вводится как основные (объединение, пересечение, разность, произведение, соединение), так и дополнительные (агрегирование, группирование, сортировка) операции.

В работе [9] рассматриваются различные представления мультимножеств: в мультипликативной и линейной формах, в виде последовательности, семейства множеств, числовой функции. Определяются операции над мультимножествами, а также приводится краткий обзор применений мультимножеств в математике, компьютерных науках и других областях.

Применение мультимножеств в базах данных (БД), является естественным применением их возможностей. Этот вопрос освещен в работе Ж. Ламперти (G. Lamperti) и др. [10]. Авторы отмечают, что современные коммерческие реляционные системы БД позволяют проводить мультимножественно-ориентированные манипуляции над таблицами, даже если они основаны на формальной множественно-ориентированной модели.

Мультимножества также достаточно широко используются в SQL-подобных языках [11]. Начиная со стандарта SQL:2003, в язык был введен конструктор типа MULTISET. Значения мультимножеств задаются путем использования специальной конструкции значений-мультимножеств (multiset value constructor). Кроме того, для мультимножеств поддерживаются операции объединения (multiset union), пересечения (multiset intersect) и разности (multiset except). Также введены новые агрегатные функции (collect, fusion, intersect). Введение конструктора типа мультимножества открывает новые возможности для применения языка SQL.

Л. Либкин (L. Libkin) и Л. Вонг (L. Wong) рассматривали теоретические вопросы баз данных, основой которых выступают мультимножества. В работе [12] они строят язык запросов BQL (Bag Query Language) для мультимножеств и исследуют связь между полученным языком и, так называемой вложенной реляционной алгеброй (nested relation algebra). Работа [13] посвящена изучению выразительной силы языка запросов для мультимножеств, а также обсуждению проблемы использования конструкций типа "структурная рекурсия" и "ограниченный цикл" для мультимножеств, множеств и списков.

Авторы К. Росс (K. Ross) и Ю. Стоянович (J. Stoyanovich) в [14] представляют симметрическую связь между *k*-арными сущностями БД как мультимножество мощности *k*, где *k* – натуральное число. Мультимножества, ограниченные по мощности (cardinality-bounded multiset), естественным образом возникают при решении реальных задач. В работе приводятся аргументы о необходимости поддержки базами данных мультимножеств, ограниченных по мощности, и предлагаются методы реализации. Авторы также описывают синтаксис расширения SQL, что позволит формулировать запросы над такими симметрическими связями.

Мультимножества также используются в декларативных языках программирования. Дж. Ллойд (J.W. Lloyd) в статьях [15, 16] предлагает новый способ поддержки мультимножеств в декларативном языке программирования Escher. Он вводит стандартное определение мультимножества, а потом определяет его соответствующими средствами языка. Автор также реализует операции над мультимножествами (суммирование, объединение, пересечение, разность) и ряд вспомогательных функций.

В декларативном языке ограничений OCL (входящем в современный универсальный язык моделирования UML) такой структурный тип, как мультимножество (BAG), определен явно. Такой тип является одним из разновидностей коллекций и имеет набор соответствующих операций [17].

Одним из возможных применений теории мультимножеств является представление и кодирование информации в терминах данной теории. Этому вопросу посвящена работа [18]. Информационный ресурс в данном случае рассматривается как ресурс, порождающий мультимножественные сообщения (т. е. сообщения, состоящие из мультимножества символов). Исследуется норма энтропии такого мультимножественного информационного ресурса.

Д. Кнут в своей работе [19] использует мультимножества в контекстно-свободных мультиязыках. Он определяет мультиязык как мультимножество строк и строит контекстно-свободный мультиязык. Автор обращает внимание на то, что замена множества строк на мультимножество строк является более естественным с точки зрения программирования.

Мультимножества применяют при определении основных понятий сетей Петри [20, 21] и в задачах распознавания символов [22].

Кроме компьютерных наук мультимножества используются в математике (в λ-исчислении [23]), физике, философии, логике, лингвистике [2, 9], а также в новой области знаний – так называемых вычислениях на ДНК [24].

Анализ литературы свидетельствует о достаточно широком применении мультимножеств при решении практических задач, что, в свою очередь, вызывает необходимость в дальнейшем расширении и уточнении соответствующих разделов теории мультимножеств. В данной работе рассматривается построение решетки мультимножеств, что поясняет структуру семейства мультимножеств.

Решетка мультимножеств

Введем формальное определение мультимножества. Мультимножество α с основой U – это функция вида $\alpha : U \to N^+$, где U – некоторое множество (в классическом канторовском понимании), а $N^+ = \{1, 2, ...\}$ – множество натуральных чисел без нуля [4, 5].

Пусть задано мультимножество α с основой $U_\alpha = dom\,\alpha$. Здесь $dom\,\alpha$ – множество первых компонент пар, которые составляют функцию, т. е. область определения мультимножества как функции.

Характеристической функцией мультимножества α называется функция вида $\chi_\alpha : D \to N$, значение которой задается следующей кусочной схемой:

$$\chi_\alpha(d) = \begin{cases} \alpha(d), \text{если } d \in dom\alpha, \\ 0, \text{иначе;} \end{cases}$$

для всех $d \in D$, где D – универсум элементов основ мультимножеств [4, 5]. Очевидно, что по характеристической функции соответствующее мультимножество восстанавливается однозначно.

Введем бинарное отношение включения на мультимножествах. Мультимножество β включается в мультимножество α ($\beta \preceq \alpha$), если для их характеристических функций выполняется утверждение: $\chi_\beta(d) \le \chi_\alpha(d),\ \forall d \in D$. Непосредственно проверяется, что отношение включения является частичным порядком.

Дадим определение операциям объединения и пересечения мультимножеств. Операция $\bigcup_{All}$ мультимножествам α и β сопоставляет мультимножество $\alpha \bigcup_{All} \beta$, значение характеристической функции которого на произвольном аргументе *d* задается выражением: $\max(\chi_\alpha(d), \chi_\beta(d))$.

Операция $\bigcap_{All}$ мультимножествам α и β сопоставляет мультимножество $\alpha \bigcap_{All} \beta$, значение характеристической функции которого на произвольном аргументе *d* задается выражением: $\min(\chi_\alpha(d), \chi_\beta(d))$.

Операции объединения и пересечения мультимножеств имеют стандартные свойства.

Лемма (о идемпотентности, коммутативности и ассоциативности операций объединения и пересечения). Операции $\bigcup_{All}$ и $\bigcap_{All}$ идемпотентны (т.е. $\alpha \bigcup_{All} \alpha = \alpha$, $\alpha \bigcap_{All} \alpha = \alpha$), коммутативны и ассоциативны.

Доказательство вытекает из того, что теоретико-числовые операции $\max$, $\min$ имеют те же самые свойства.

Таким образом, можно рассматривать две коммутативные идемпотентные полугруппы: $\langle \boldsymbol{A}, \bigcup_{All} \rangle$ и $\langle \boldsymbol{A}, \bigcap_{All} \rangle$, где $\boldsymbol{A}$ – семейство мультимножеств соответствующего универсума D.

Используя результат теории решеток [25], (§ 8, с. 151, теорема 1), можно полугруппу по объединению превратить в верхнюю полурешетку, а полугруппу по пересечению – в нижнюю. Частичные порядки верхней полурешетки и нижней полурешетки задаются соответственно:

$$\alpha \mathrel{\bar{\le}} \beta \Leftrightarrow \alpha \bigcup_{All} \beta = \beta,\ \alpha \mathrel{\bar{\trianglelefteq}} \beta \Leftrightarrow \alpha \bigcap_{All} \beta = \alpha,$$

причем $\sup_{\bar{\le}}\{\alpha, \beta\} = \alpha \bigcup_{All} \beta$, $\inf_{\bar{\trianglelefteq}}\{\alpha, \beta\} = \alpha \bigcap_{All} \beta$.

Непосредственно проверятся, что эти порядки совпадают с порядком включения мультимножеств $\preceq$. Таким образом, семейство мультимножеств $\boldsymbol{A}$ с частичным порядком $\preceq$ является одновременно и верхней, и нижней полурешеткой, т. е. решеткой.

Такой способ построения решетки мультимножеств явно не использовал законы поглощения. Покажем в общем случае их роль при построении решетки по двум коммутативным идемпотентным полугруппам, сигнатурные операции которых связаны законами поглощения.

Рассмотрим две коммутативные идемпотентные полугруппы $\langle A, + \rangle$ и $\langle A, \cdot \rangle$, где A – некоторое абстрактное множество. Используя хорошо известный результат теории решеток о связи коммутативных идемпотентных полугрупп и полурешеток (полуструктур) [25], (§ 8, с. 151, теорема 1), полугруппу $\langle A, + \rangle$ можно превратить в верхнюю полурешетку, а полугруппу $\langle A, \cdot \rangle$ – в нижнюю. Соответствующие частичные порядки верхней и нижней полурешеток задаются выражениями:

$$a \le b \Leftrightarrow a + b = b,\ a \triangleleft b \Leftrightarrow ab = a,$$

причем $\sup_{\le}\{a, b\} = a + b$, $\inf_{\triangleleft}\{a, b\} = ab$. Заметим, что, согласно стандартным соглашениям [25, 26], знак операции умножения “ $\cdot$ ” в выражениях опускается.

Теорема (критерий совпадения порядков верхней и нижней полурешеток). Частичные порядки верхней и нижней полурешеток совпадают тогда и только тогда, когда выполняются законы поглощения.

Доказательство. Сначала докажем необходимость. Если порядки совпадают, то заданное множество является одновременно и верхней и нижней полурешеткой, а, значит, решеткой. Для решеток законы поглощения выполняются [25, 26].

Докажем достаточность. Для этого нужно показать, что для $\forall\, a, b$ выполняется эквивалентность: $a + b = b \Leftrightarrow ab = a$. Допустим, что равенство $a + b = b$ выполняется. Тогда $ab = a\,(a + b)$. По закону поглощения $a\,(a + b) = a$, поэтому $ab = a$. Аналогично сделаем допущение, что $ab = a$. В этом случае $a + b = ab + b$. По закону поглощения $ab + b = b$, значит, $a + b = b$.

Естественно, этот общий результат применим к построению решетки мультимножеств. Для этого надо убедится в выполнении законов поглощения: $\alpha \bigcap_{All} (\alpha \bigcup_{All} \beta) = \alpha$ и $\alpha \bigcup_{All} (\alpha \bigcap_{All} \beta) = \alpha$, что делается непосредственно (и, в свою очередь, вытекает из выполнения законов поглощения для теоретико-числовых функций $\max$, $\min$)[1].

[1] Автор выражает благодарность Д.Б. Бую за неоднократные обсуждения.

1. *Albert J.* Algebraic properties of bag data types // Seventeenth International Conference on Very Large Data Bases. – Barcelona, Spain, 1991. – P. 211–219.
2. *Blizard W.* The Development of Multiset Theory // Notre Dame J. of Formal Logic. – 1989. – Vol. 30, N 1. – P. 36–66.
3. *Syropoulos A.* Mathematic of Multisets // Multiset Processing: Mathematical, Computer Science, and Molecular Computing Points of View, number 2235 in Lecture Notes in Computing Since. – Berlin: Springer-Verlag, 2001. – P. 347 – 358.
4. *Петровський А.Б.* Основные понятия теории мультимножеств. – М.: “Едиториал УРСС”, 2002. – 80 с.
5. *Реляційні* бази даних: табличні алгебри та SQL-подібні мови / В.Н. Редько, Ю.Й. Брона, Д.Б. Буй, С.А. Поляков. – К.: Видавничий дім “Академперіодика”, 2001. – 198 с.
6. *Кнут Д.* Искусство программирования: 2 том, 3-е изд.: пер. с англ. – М.: “Вильямс”, 2000. – 832 с.
7. *Петровський А.Б.* Пространства множеств и мультимножеств. – М.: “Едиториал УРСС”, 2003. – 248 с.
8. *Гарсиа-Молина Г., Ульман Дж., Уидом Дж.* Системы баз данных: пер. с англ. – М.: “Вильямс”, 2004. – 1088 с.
9. *Singh D., Ibrahim A.M., Yohanna T. , Singh J.N.* An Overview of the Applications of Multisets // Novi Sad Journal of Mathematics. – 2007. – Vol. 37, N 2. – P. 73–92.
10. *Lamperti G., Melchiori M., Zanella M.* On Multisets in Database Systems // Multiset Processing: Mathematical, Computer Science, and Molecular Computing Points of View, number 2235 in Lecture Notes in Computing Since. – Berlin: Springer-Verlag, 2001. – P. 147–215.
11. *Наиболее* интересные новшества в стандарте SQL:2003 [Электронный ресурс]. Режим доступа: http://www.nestor. minsk.by/sr/2004/03/40331.html.
12. *Libkin L., Wong L.* Query Language for Bags and Aggregates Function // J. of Computer and System Sciences. – 1997. – Vol. 55, N 1. – P. 241–272.
13. *Libkin L., Wong L.* Some Properties of Query Language for Bags // Proceedings of 4th International Workshop on Database Programming Languages. – New York, 1993. – P. 97–114.
14. *Ross K., Stoyanovich J.* Symmetric relations and cardinality-bounded multisets in database systems // Very Large Database Endowment: international conference, August 31 – September 03, 2004, Totonto, Canada: proceedings. – 2004. – Vol. 30. – P. 912–923.
15. *Lloyd J.* Programming with Sets and Multisets // Department of Computer Science University of Bristol, 1998.
16. *Lloyd J.* Programming with Multisets // Department of Computer Science University of Bristol, 1998.
17. *Кузнецов С.Д.* Концептуальное проектирование реляционных баз данных с использованием языка UML [Электронный ресурс] – Режим доступа: ftp://ftp.dol.ru/pub/users/cgntv/ download/sbornic/sbornic9/Doc13.doc.
18. *Bonchis C., Izbasa C., Ciobanu G.* Information Theory over Multiset // Research Institute “re-Austria”, Institute of Computer Science, 2005.
19. *Knuth D.* Context-Free Multilanguages // Theoretical Studies in Computer Science. – Academic Press, 1992. – P. 1–13.
20. *Башкин В.А. , Ломазова И.А.* Подобие обобщенных ресурсов в сетях Петри [Электронный ресурс] – Режим доступа: http://lvk.cs.msu.su/files/mco2005/bashkin.pdf.
21. *Сети* Петри [Электронный ресурс] / Режим доступа: http://www.iacp.dvo.ru/lab_11/otchet/ot2000/pn3.html#top.
22. *Славин О.А.* Использование мультимножеств в распознавании символов [Электронный ресурс] – Режим доступа: ftp://ftp.dol.ru/pub/users/cgntv/download/sbornic/sbornic9/Doc13.doc
23. *Барендрегт Х.* Ламбда-исчисление. Его синтаксис и семантика: пер. с англ. / Х. Барендрегт. – М.: “Мир”, 1985. – 606 с.
24. *Малинецкий Г.Г., Науменко С.А.* Вычисления на ДНК. Эксперименты. Модели. Алгоритмы. Инструментальные средства [Электронный ресурс] – Режим доступа: http://www.keldysh.ru/papers/2005/ prep57/prep2005_57.html.
25. *Скорняков Л.А.* Элементы алгебры. – М.: “Наука”, 1986. – 240 с.
26. *Мальцев А.И.* Алгебраические системы. – М.: “Наука”, 1970. – 392 с.