

УДК 004.272.43

А. Э. Саак, д-р техн. наук, доц., e-mail: saak@tgn.sfedu.ru,
Южный федеральный университет, г. Ростов-на-Дону — Таганрог

Диспетчеризация массивов заявок кругового и гиперболического типа в Grid-системах

В среде ресурсных прямоугольников предлагаются варианты уровневых алгоритмов, позволяющие диспетчирование массивами ресурсных прямоугольников, не обладающих свойством монотонности горизонтальных или вертикальных измерений. Для указанных массивов, состоящих из заявок кругового и гиперболического типа, предлагаются варианты уровневых алгоритмов по высоте и протяженности (с недостатком, превышением, минимальным отклонением). Сравнительный анализ позволяет рекомендовать ряд таких алгоритмов к использованию в Grid-системах с централизованной структурой.

Ключевые слова: Grid-система, централизованная структура системы диспетчирования, мультисайтный режим обслуживания заявки, обслуживание заявки без прерывания, заявка кругового типа, заявка гиперболического типа, не-эвклидова эвристическая мера, полиномиальная трудоемкость алгоритма, уровневый алгоритм по высоте, уровневый алгоритм по протяженности

Введение и постановка задачи

Возрастающая потребность пользователей в вычислительной мощности, наряду с развитием технологии, ростом числа процессоров, увеличением быстродействия телекоммуникационных сетей, стимулировала в конце 90-х годов переход от многопроцессорных систем, метакомпьютинга к Grid-компьютингу, который развивается и в настоящее время [1, 2], наряду с такими парадигмами распределенных вычислений [3], как облачные вычисления [4] и вычислительные джунгли [5].

В Grid-диспетчировании выделяют четыре основных этапа: выявление ресурсов (*resource discovery*), выбор ресурсов (*resource selection*), генерация расписания (*schedule generation*), выполнение задачи (*job execution*) [6]. Наше основное внимание сосредоточено на этапе генерации расписания алгоритмом диспетчирования (*scheduling algorithm*). При этом алгоритм диспетчирования определяет способ, которым задачи назначаются на ресурсы, а под расписанием понимается назначение задач на ресурсы в определенные периоды времени [7].

В настоящей работе рассматриваются вычислительные заявки [8] с заранее известным временем решения [9], в которых число требуемых процессоров определяет пользователь при подаче в систему [10]. Предполагается, что политика диспетчирования поддерживает обслуживание заявки без прерывания (*non-preemptive, run-to-completion*), т. е. выделенные заявке процессоры удерживаются в течение всего времени ее выполнения от начала до

окончания [11, 12]. В этих условиях при выделении для обработки заявки процессоров с последовательными номерами задача генерации расписания алгоритмом диспетчирования эквивалентна задаче распределения ресурсных прямоугольников. При представлении заявки пользователя для обслуживания диспетчером Grid-системы ресурсным прямоугольником горизонтальное и вертикальное измерения принимаются равными соответственно числу единиц ресурса времени и процессоров, требуемому для обработки заявки [13]. Символом $[a(j), b(j)]$ обозначается j -я заявка, требующая $a(j)$ единиц времени и $b(j)$ единиц процессоров.

Grid-системы с централизованной структурой системы диспетчирования [8, 14], состоящие из сайтов, содержащих параллельные системы, функционирующие в режиме мультисайтного диспетчирования [14, 15], моделируются ресурсным квадратом [16, 17].

Для управления распределением вычислительных-временных ресурсов и заявками пользователей Grid-систем в качестве теоретической основы алгоритмического обеспечения диспетчирования с полиномиальной трудоемкостью в [16—20] определена среда ресурсных прямоугольников. В [16] введены операции динамического интегрирования ресурсных прямоугольников по вертикали и по горизонтали, состоящие в наилучшем приближении заданного значения с недостатком. Состав операций расширен динамическим интегрированием ресурсных прямоугольников по вертикали [21] и по

горизонтали [22] с превышением и с минимальным отклонением. На основе этих операций разработаны полиномиальные алгоритмы [21, 22], адаптированные под соответствующий квадратичный тип массива заявок. В [16] круговой, гиперболический и параболический типы определены для массивов из не менее двух заявок. Квадратичный тип одной заявки введен в [23], где исследована адаптируемость полиномиальных алгоритмов под массивы, состоящие из заявок кругового типа.

В настоящей статье ставится вопрос об адаптированности полиномиальных алгоритмов для массивов, состоящих из заявок кругового и гиперболического квадратичного типа.

Расширение класса полиномиальных уровневых алгоритмов обслуживания в Grid-системах

В работе [20] предложен уровневый алгоритм по высоте (с недостатком), в [16] — уровневый алгоритм по протяженности (с недостатком). Уровневый алгоритм по высоте с превышением и уровневый волновой алгоритм (с минимальным отклонением) введены в работе [21]. Рассмотрим варианты этих алгоритмов для массивов ресурсных прямоугольников, не обладающих свойством монотонности горизонтальных или вертикальных измерений, присутствующих массивам кругового и гиперболического типов: *V*-уровневых алгоритмов по высоте (с недостатком, с превышением, с минимальным отклонением) и *H*-уровневых алгоритмов по протяженности (с недостатком, с превышением, с минимальным отклонением).

Функционирование *V*-уровневых алгоритмов по высоте аналогично приведенным в [21], с тем отличием, что на каждом шаге правая сторона ресурсной оболочки определяется суммой значения правой стороны достигнутой ресурсной оболочки и максимального горизонтального измерения элементов в слое, а не горизонтальным измерением начального элемента слоя. Так, на *s*-м шаге (рис. 1) значение правой стороны ресурсной оболочки Y_{s+1} равно $Y_{s+1} = Y_s + \max_{j_{s-1} \leq j < j_s} a(j)$, где j_{s-1}, j_s — начальный и конечный номера ресурсных прямоугольников в вертикальном слое.

В рассматриваемых *V*-уровневых алгоритмах по высоте значение уровня определяется величиной

$$\sqrt[k]{\sum_{j=0}^{k-1} a(j)b(j)}, \text{ где } k \text{ — число}$$

заявок в массиве.

Функционирование *H*-уровневых алгоритмов по протяженности аналогично приведенному в работе [16]. Для *H*-уровневого алгоритма по протяженности с превышением отличие состоит в том, что на каждом шаге используется введенная в [22] операция динамического интегрирования ресурсных прямоугольников по горизонтали с превыше-

нием, состоящая в наилучшем приближении заданной протяженности с избытком. Для *H*-уровневого алгоритма по протяженности с минимальным отклонением отличие состоит в том, что на каждом шаге используется введенная в работе [22] операция динамического интегрирования ресурсных прямоугольников по горизонтали с минимальным отклонением. При этом верхняя сторона ресурсной оболочки определяется суммой значения верхней стороны достигнутой ресурсной оболочки и максимального вертикального измерения элементов в слое, а не вертикальным измерением начального элемента слоя. Так, на *s*-м шаге (рис. 2) значение

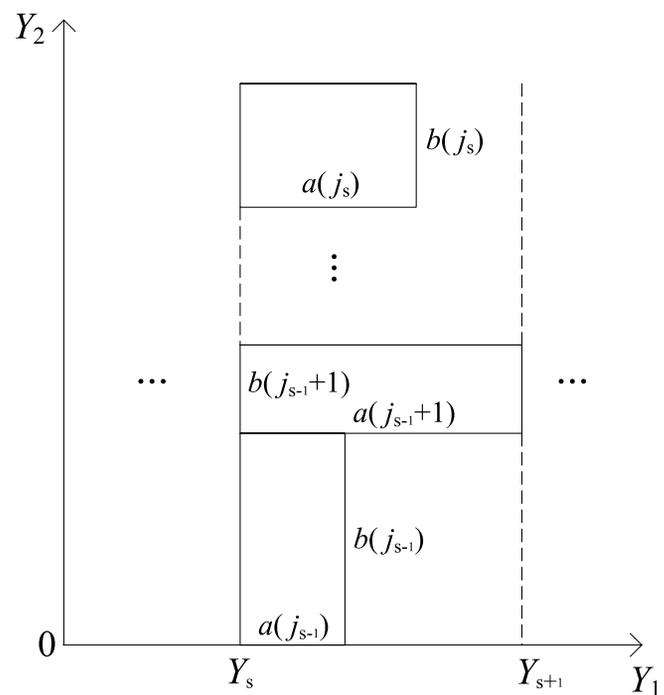


Рис. 1. Вертикальный *s*-й слой *V*-уровневых алгоритмов

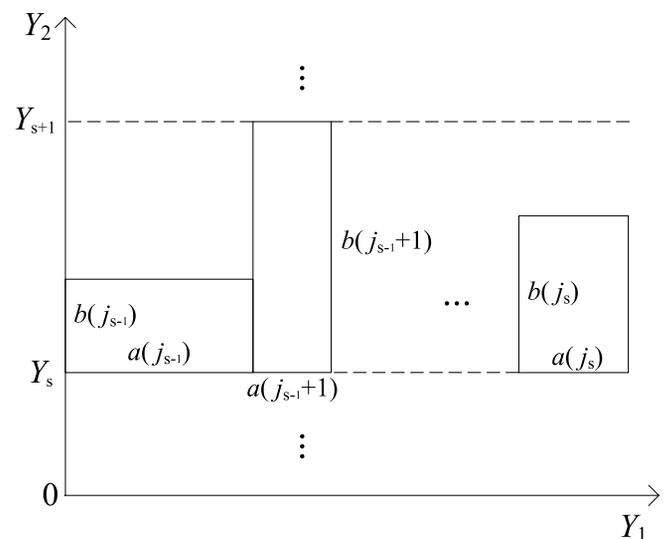


Рис. 2. Горизонтальный *s*-й слой *H*-уровневых алгоритмов

верхней стороны ресурсной оболочки Y_{s+1} равно $Y_{s+1} = Y_s + \max_{j_{s-1} \leq j < j_s} b(j)$, где j_{s-1}, j_s — начальный и конечный номера ресурсных прямоугольников в горизонтальном слое.

В рассматриваемых H -уровневых алгоритмах по протяженности значение требуемой протяженности определяется величиной

$$\sqrt{\sum_{j=0}^{k-1} a(j)b(j)}.$$

Далее исследуется адаптированность полиномиальных V -уровневых и H -уровневых алгоритмов для массивов, состоящих из заявок кругового и гиперболического квадратичного типа.

Диспетчеризация уровней алгоритмами массива, состоящего из заявок кругового и гиперболического типов

Для оценки качества диспетчеризации эвристических алгоритмов в работе [16] была предложена

неевклидова эвристическая мера $\frac{1}{2} \left(\frac{AB + (A - B)^2}{\sqrt{\sum_{j=0}^{k-1} a(j)b(j)}} \right)^2$,

учитывающая как площадь AB , так и форму $(A - B)^2$ занятой ресурсной области, где A — протяженность, B — уровень по вертикали ресурсной прямоугольной оболочки. Возможный минимум эвристической меры, равный $1/2$, достигается при беспустотной укладке в квадрат.

Массивы прямоугольников, индуцированные элементами диссекции квадрата на различные прямоугольники с отношением сторон $1:2$ ($\frac{a(j)}{b(j)} = \frac{1}{2}$

или $\frac{a(j)}{b(j)} = 2$), используются нами в качестве тестового примера [24]. Такой массив, в соответствии

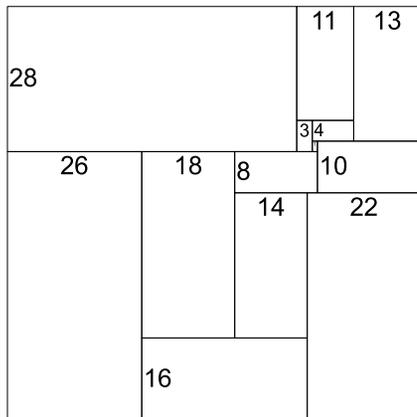


Рис. 3. Диссекция квадрата на 13 прямоугольников с отношением сторон $1:2$

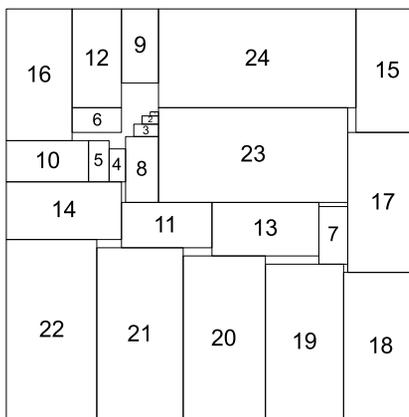


Рис. 4. Укладка 24 последовательных прямоугольников с отношением сторон $1:2$ в квадрат с минимальной стороной [25]

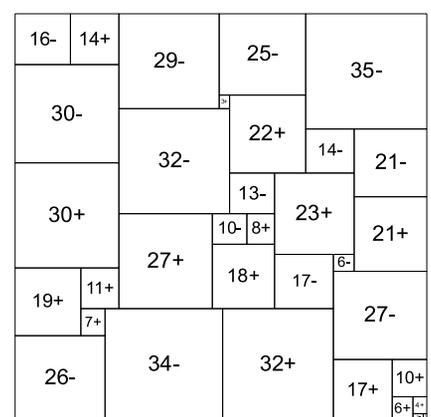


Рис. 5. Облицовка почти квадрата, содержащая 34 последовательных почти квадратов [28]

с определениями [23], содержит прямоугольники кругового типа при горизонтальной ориентации $a(j) \geq b(j)$ и гиперболического типа — при вертикальной $a(j) < b(j)$. В [24] приведены диссекции квадрата на 11, 12 и 13 (рис. 3) прямоугольников. На прямоугольнике указано значение меньшей стороны.

Указанные массивы ресурсных прямоугольников, упорядоченные по убыванию высот, обозначим следующим образом: для диссекции порядка 11 — массив I, для диссекции порядка 12 — массив II, для диссекции порядка 13 — массив III.

Следующие тестовые примеры порождены укладками последовательных прямоугольников $[j, 2j]$, $j = 1, 2, \dots, k$, в квадрат с минимальной стороной, полученные Э. Фридманом (Erich Friedman) в 2014 г. для $k = 21, 22, 23, 24$ (рис. 4) [25]. На прямоугольнике указано значение меньшей стороны.

Массивы прямоугольников, индуцированные такими укладками, содержат прямоугольники с отношением сторон $1:2$ горизонтальной и вертикальной ориентации. Соответствующие массивы ресурсных прямоугольников, упорядоченные по убыванию высот, обозначим следующим образом: для $k = 21$ — массив IV, для $k = 22$ — массив V, для $k = 23$ — массив VI, $k = 24$ — массив VII.

Заключительные тестовые примеры индуцированы укладкой массива $[j, (j + 1)]$, $j = 1, 2, \dots, k$, последовательных прямоугольников (измерения сторон отличаются на единицу, называемых почти квадратами [26—28], в соответствующий почти квадрат без пустот. В [28] приведены такие укладки, называемые облицовкой (tiling), для $k = 20, 34$ (рис. 5). Следуя [28], на прямоугольнике указано значение горизонтального измерения, а знаки "+" или "-" означают больше или меньше на 1 значение вертикального измерения.

Так как условие отнесения прямоугольника к круговому или гиперболическому типу $(b(j) - a(j))^2 \leq a(j)b(j)$, сформулированное в работе [23], для

почти квадратов приводит к очевидному неравенству $1 \leq j(j + 1)$, то почти квадраты относятся к круговому или гиперболическому типу в зависимости от горизонтальной или вертикальной ориентации [23]. Массивы ресурсных прямоугольников, индуцированные облицовкой, упорядоченные по убыванию высот, обозначим следующим образом: для $k = 20$ — массив VIII, $k = 34$ — массив IX.

Вычислим эвристические меры ресурсных оболочек, получаемых при диспетчировании этих массивов V -уровневыми алгоритмами по высоте (с недостатком, с превышением, с минимальным отклонением) и H -уровневыми алгоритмами по протяженности (с недостатком, с превышением, с минимальным отклонением).

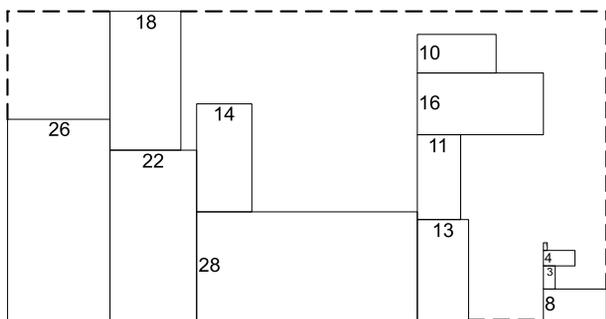


Рис. 6. Укладка массива III V -уровневым алгоритмом по высоте с недостатком

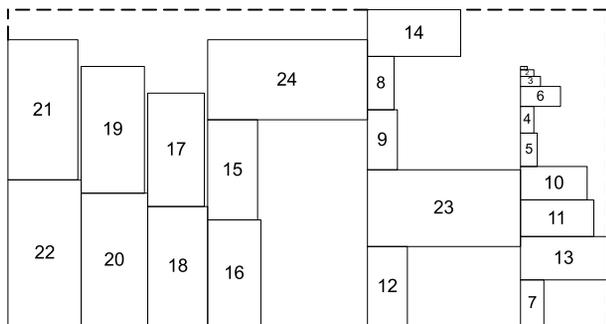


Рис. 7. Укладка массива VII V -уровневым алгоритмом по высоте с недостатком

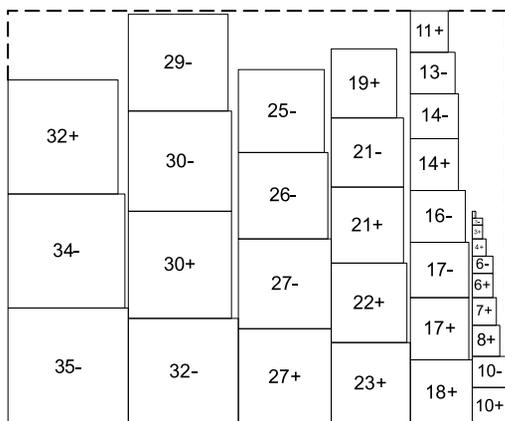


Рис. 8. Укладка массива IX V -уровневым алгоритмом по высоте с недостатком

Построения V -уровневым алгоритмом по высоте с недостатком для массивов III, VII, IX приведены на рис. 6—8.

Эвристические меры ресурсных оболочек V -уровневого алгоритма по высоте с недостатком для массива из заявок кругового и гиперболического квадратичного типов приведены в табл. 1.

Видим, что эвристические меры ресурсных оболочек V -уровневого алгоритма по высоте с недостатком не превосходят значения $0,5 + 0,86$.

Построения V -уровневым алгоритмом по высоте с превышением для массивов III, VII, IX приведены на рис. 9—11.

Эвристические меры ресурсных оболочек V -уровневого алгоритма по высоте с превышением для

Таблица 1
Эвристические меры ресурсных оболочек V -уровневого алгоритма по высоте с недостатком

Номер массива	Эвристическая мера	Номер массива	Эвристическая мера	Номер массива	Эвристическая мера
I	1,03	IV	0,81	VII	1,24
II	1,3	V	0,76	VIII	0,71
III	1,36	VI	0,96	IX	0,63

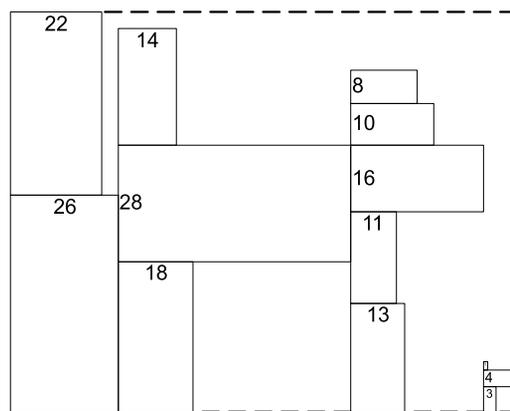


Рис. 9. Укладка массива III V -уровневым алгоритмом по высоте с превышением

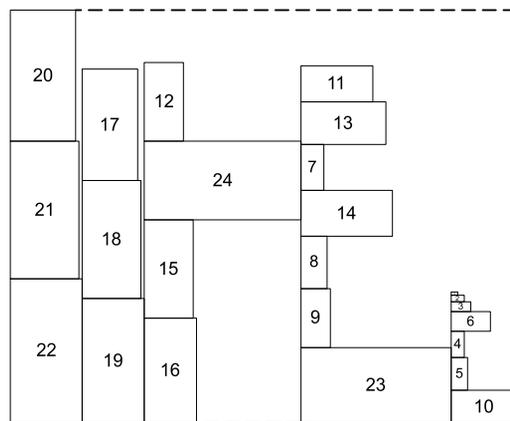


Рис. 10. Укладка массива VII V -уровневым алгоритмом по высоте с превышением

массива из заявок кругового и гиперболического квадратичного типов приведены в табл. 2.

Видим, что эвристические меры ресурсных оболочек *V*-уровневого алгоритма по высоте с превышением не превосходят значения $0,5 + 0,54$.

Построения *V*-уровневым алгоритмом по высоте с минимальным отклонением для массивов III, VII, IX приведены на рис. 12—14.

Эвристические меры ресурсных оболочек *V*-уровневого алгоритма по высоте с минимальным отклонением для массива из заявок кругового и гиперболического квадратичного типа приведены в табл. 3.

Видим, что эвристические меры ресурсных оболочек *V*-уровневого алгоритма по высоте с минимальным отклонением не превосходят значения $0,5 + 0,47$.

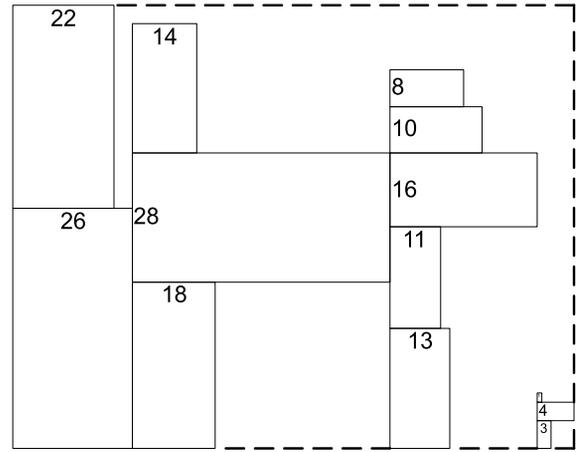


Рис. 12. Укладка массива III *V*-уровневым алгоритмом по высоте с минимальным отклонением

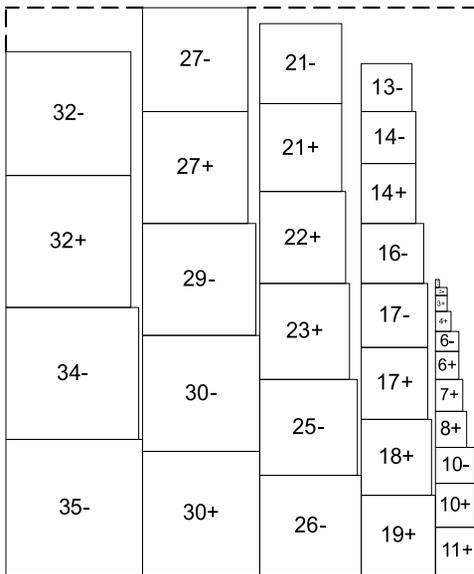


Рис. 11. Укладка массива IX *V*-уровневым алгоритмом по высоте с превышением

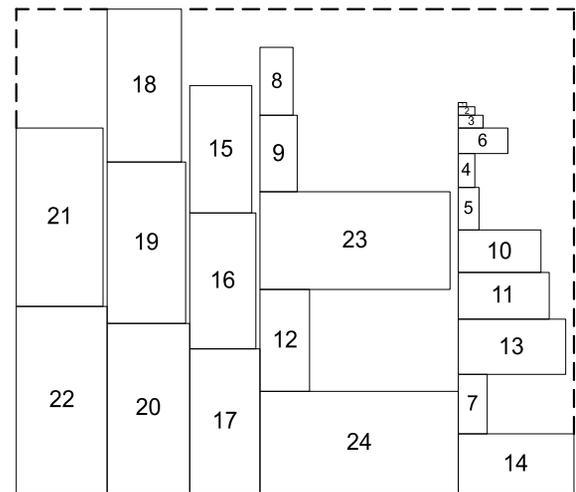


Рис. 13. Укладка массива VII *V*-уровневым алгоритмом по высоте с минимальным отклонением

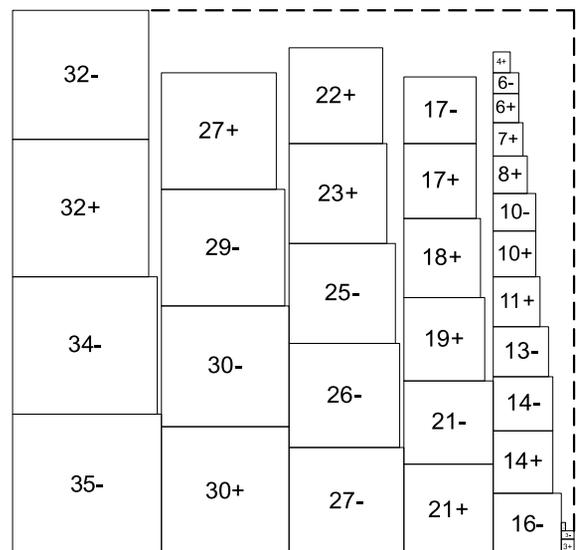


Рис. 14. Укладка массива IX *V*-уровневым алгоритмом по высоте с минимальным отклонением

Таблица 2

Эвристические меры ресурсных оболочек *V*-уровневого алгоритма по высоте с превышением

Номер массива	Эвристическая мера	Номер массива	Эвристическая мера	Номер массива	Эвристическая мера
I	0,86	IV	0,74	VII	1,04
II	0,81	V	0,88	VIII	0,64
III	0,97	VI	0,92	IX	0,62

Таблица 3

Эвристические меры ресурсных оболочек *V*-уровневого алгоритма по высоте с минимальным отклонением

Номер массива	Эвристическая мера	Номер массива	Эвристическая мера	Номер массива	Эвристическая мера
I	0,86	IV	0,72	VII	0,81
II	0,78	V	0,93	VIII	0,62
III	0,97	VI	0,77	IX	0,61

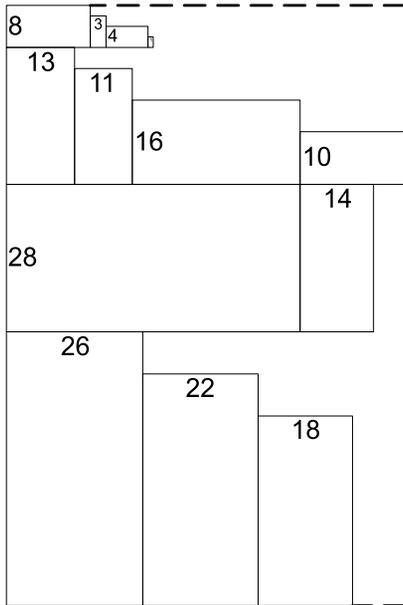


Рис. 15. Укладка массива III *H*-уровневым алгоритмом по протяженности с недостатком

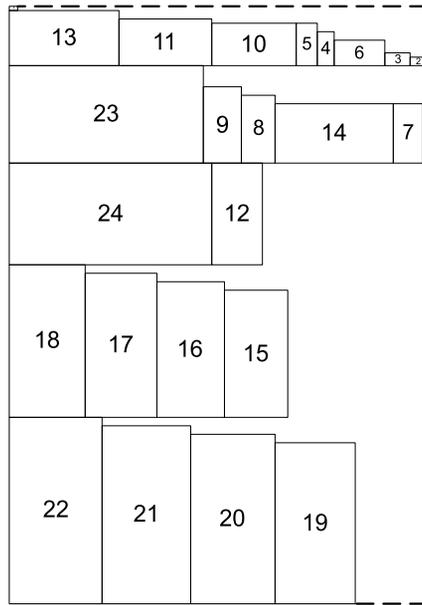


Рис. 16. Укладка массива VII *H*-уровневым алгоритмом по протяженности с недостатком

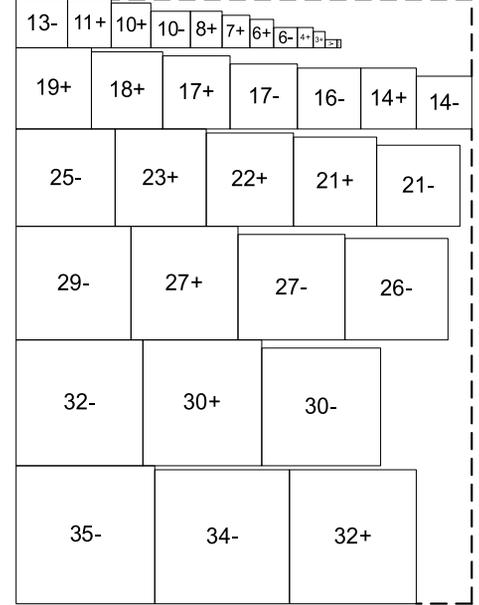


Рис. 17. Укладка массива IX *H*-уровневым алгоритмом по протяженности с недостатком

Таблица 4

Эвристические меры ресурсных оболочек *H*-уровневого алгоритма по протяженности с недостатком

Номер массива	Эвристическая мера	Номер массива	Эвристическая мера	Номер массива	Эвристическая мера
I	0,71	IV	0,69	VII	0,80
II	0,73	V	0,70	VIII	0,67
III	0,79	VI	0,72	IX	0,64

Построения *H*-уровневым алгоритмом по протяженности с недостатком для массивов III, VII, IX приведены на рис. 15–17.

Эвристические меры ресурсных оболочек *H*-уровневого алгоритма по протяженности с недостатком для массива из заявок кругового и гиперболического квадратичного типов приведены в табл. 4.

Видим, что эвристические меры ресурсных оболочек *H*-уровневого алгоритма по протяженности с недостатком не превосходят значения $0,5 + 0,30$.

Построения *H*-уровневым алгоритмом по протяженности с превышением для массивов III, VII, IX приведены на рис. 18–20.

Эвристические меры ресурсных оболочек *H*-уровневого алгоритма по протяженности с превышением для массива из заявок кругового и гиперболического квадратичного типов приведены в табл. 5.

Видим, что эвристические меры ресурсных оболочек *H*-уровневого алгоритма по протяженности с превышением не превосходят значения $0,5 + 0,43$.

Построения *H*-уровневым алгоритмом по протяженности с минимальным отклонением для массивов III, VII, IX приведены на рис. 21–23.

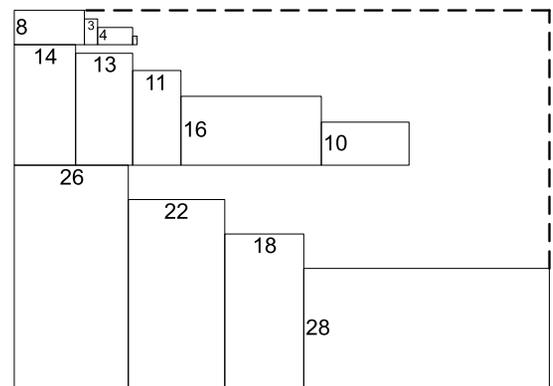


Рис. 18. Укладка массива III *H*-уровневым алгоритмом по протяженности с превышением

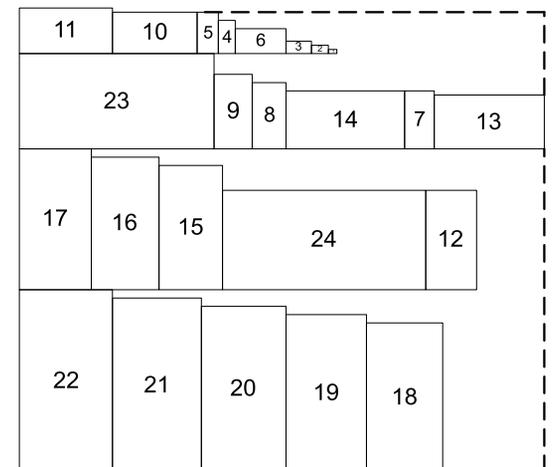


Рис. 19. Укладка массива VII *H*-уровневым алгоритмом по протяженности с превышением

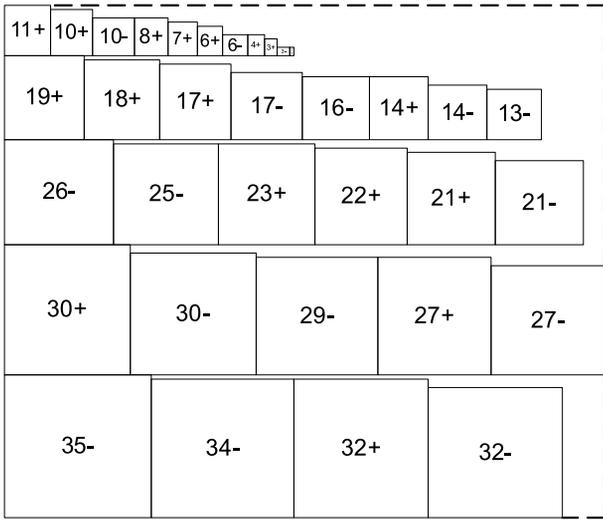


Рис. 20. Укладка массива IX H -уровневым алгоритмом по протяженности с превышением

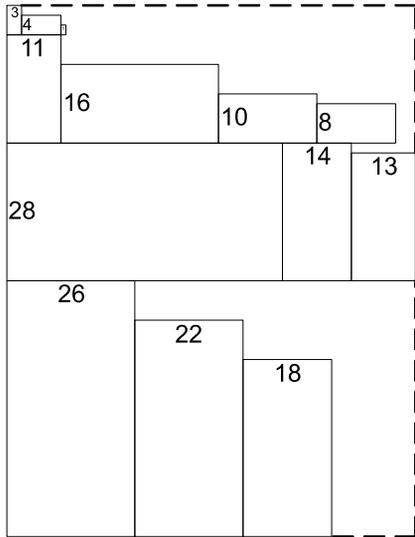


Рис. 21. Укладка массива III H -уровневым алгоритмом по протяженности с минимальным отклонением

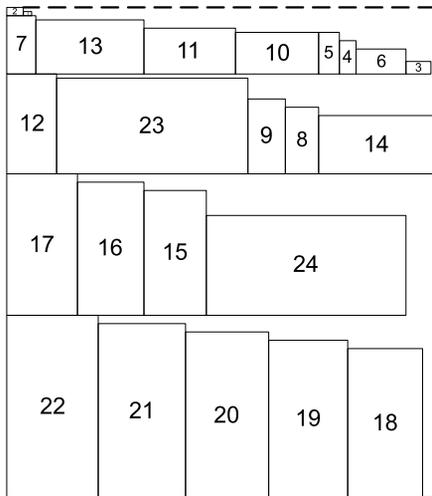


Рис. 22. Укладка массива VII H -уровневым алгоритмом по протяженности с минимальным отклонением

Таблица 5

Эвристические меры ресурсных оболочек H -уровневого алгоритма по протяженности с превышением

Номер массива	Эвристическая мера	Номер массива	Эвристическая мера	Номер массива	Эвристическая мера
I	0,67	IV	0,62	VII	0,72
II	0,81	V	0,68	VIII	0,62
III	0,93	VI	0,64	IX	0,63

Таблица 6

Эвристические меры ресурсных оболочек H -уровневого алгоритма по протяженности с минимальным отклонением

Номер массива	Эвристическая мера	Номер массива	Эвристическая мера	Номер массива	Эвристическая мера
I	0,71	IV	0,70	VII	0,63
II	0,67	V	0,66	VIII	0,64
III	0,75	VI	0,71	IX	0,61

Эвристические меры ресурсных оболочек H -уровневого алгоритма по протяженности с минимальным отклонением для массива из заявок кругового и гиперболического квадратичного типов приведены в табл. 6.

Видим, что эвристические меры ресурсных оболочек H -уровневого алгоритма по протяженности с минимальным отклонением не превосходят значения $0,5 + 0,25$.

Графики эвристической меры ресурсных оболочек V -уровневыми алгоритмами по высоте (с недостатком, с превышением, с минимальным отклонением) и H -уровневыми алгоритмами по протя-

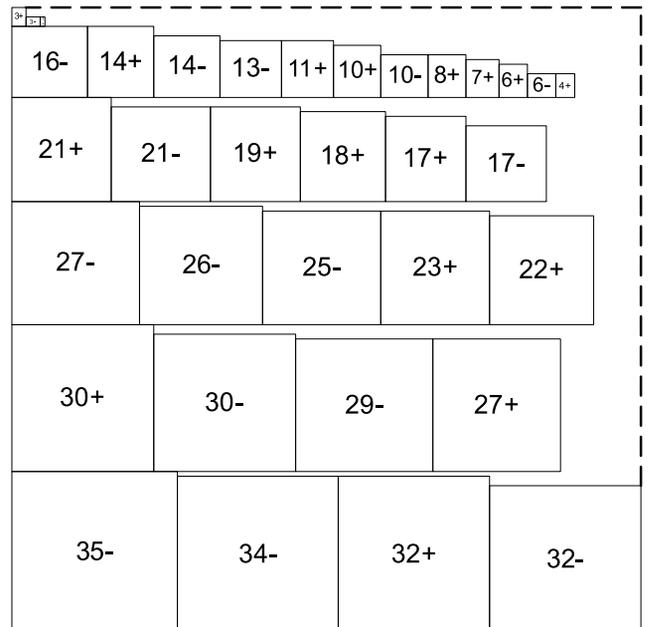


Рис. 23. Укладка массива IX H -уровневым алгоритмом по протяженности с минимальным отклонением

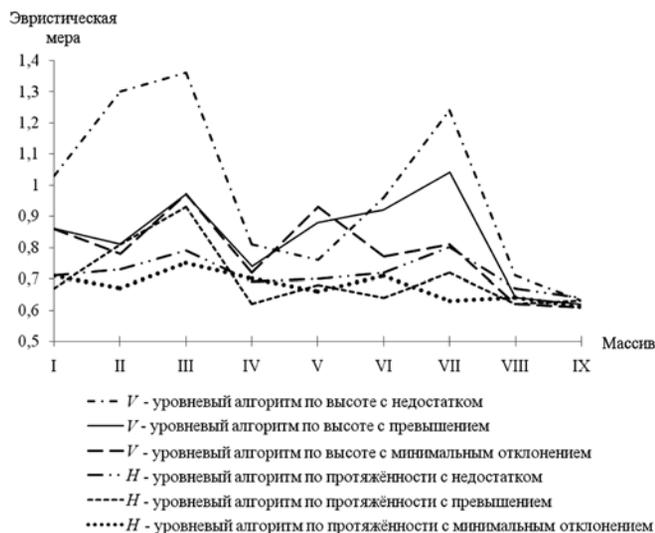


Рис. 24. Эвристические меры ресурсных оболочек V-уровневых и H-уровневых алгоритмов

женности (с недостатком, с превышением, с минимальным отклонением) при диспетчеризации массивами I—IX показаны на рис. 24.

Видим, что H-уровневые алгоритмы по протяженности имеют меньшую эвристическую меру ресурсных оболочек. При этом H-уровневый алгоритм по протяженности с минимальным отклонением имеет наименьшее значение максимума эвристической меры $0,5 + 0,25$ на рассматриваемых тестовых массивах ресурсов прямоугольников. Проведенный анализ позволяет рекомендовать предложенные полиномиальные алгоритмы к использованию в Grid-системах с централизованной структурой и мультисайтным режимом диспетчеризации при обслуживании массивов, состоящих из заявок кругового и гиперболического квадратичного типов.

Заключение

В среде ресурсных прямоугольников предлагаются варианты уровневых алгоритмов по высоте и по протяженности, позволяющие диспетчеризацию массивами ресурсных прямоугольников, не обладающих свойством монотонности горизонтальных или вертикальных измерений. Для девяти тестовых массивов заявок кругового и гиперболического квадратичного типов вычисляются эвристические меры ресурсных оболочек шести вариантов уровневых алгоритмов. Исследование показало адаптированность ряда предложенных полиномиальных алгоритмов к рассматриваемому классу массивов заявок пользователей Grid-систем.

Список литературы

1. Schwegelshohn U., Badia R., Bubak M., Danelutto M., Dustdar S., Gagliardi F., Geiger A., Hluchy L., Kranzlmüller D., Laure E., Priol T., Reinefeld A., Resch M., Router A., Rienhoff O., Rütter T., Sloot P., Talia D., Ullmann K., Yahyapour R., Voigt G.

Perspectives on grid computing // Future Generation Computer Systems. 2010. Vol. 26, Iss. 8. P. 1104—1115.

2. Bencivenni M., Michelotto D., Alfieri R., Brunetti R., Ciccanti A., Cesini D., Costantini A., Fattibene E., Gaido L., Misurelli G., Ronchieri E., Salomoni D., Veronesi P., Venturi V., Vistoli M. Accessing Grid and Cloud services through a scientific web portal // J. Grid Computing. 2015. Vol. 13, Iss. 2. P. 159—175.

3. Kahanwal B., Singh T. The distributed computing paradigms: p2p, grid, cluster, cloud, and jungle // International Journal of Latest Research in Science and Technology. 2012. Vol. 1, Iss. 2. P. 183—187.

4. Buyya R., Broberg J., Goscinski A. (eds.). Cloud computing: principles and paradigms. Hoboken: John Wiley & Sons Inc., 2011, 644 p.

5. Seinstra F., Maassen J., Nieuwpoort R., Drost N., Kessel T., Werkhoven B., Urbani J., Jacobs C., Kielmann T., Bal H. Jungle computing: Distributed supercomputing beyond clusters, grids, and clouds / M. Cafaro, G. Aloisio (eds.) // Grids, Clouds and Virtualization. London: Springer-Verlag, 2011, 235 p., pp. 167—197.

6. Li M., Baker M. The grid: core technologies. Chichester: John Wiley & Sons Ltd, 2005, 452 p.

7. Pugliese A., Talia D., Yahyapour R. Modeling and Supporting Grid Scheduling // J. Grid Computing. 2008. Vol. 6, Iss. 2. P. 195—213.

8. Rahman M., Ranjan R., Buyya R., Benattallah B. A taxonomy and survey on autonomic management of applications in grid computing environments // Concurrency Computat.: Pract. Exper. 2011. Vol. 23, Iss. 16. P. 1990—2019.

9. Ye D., Zhang G. On-Line Scheduling of Parallel Jobs // LNCS. 2004. Vol. 3104. P. 279—290.

10. Keitelson D., Rudolph L. Toward convergence in job schedulers for parallel supercomputers // LNCS. 1996. Vol. 1162. P. 1—26.

11. Feitelson D., Rudolph L., Schwegelshohn U., Sevcik K., Wong P. Theory and practice in parallel job scheduling // LNCS. 1997. Vol. 1291. P. 1—34.

12. Majumdar S., Parsons E. Parallel job scheduling: a performance perspective // LNCS. 2000. Vol. 1769. P. 233—252.

13. Caramia M., Giordani S., Iovanella A. Grid scheduling by on-line rectangle packing // Networks. 2004. Vol. 44, N. 2. P. 106—119.

14. Hamscher V., Schwegelshohn U., Streit A., Yahyapour R. Evaluation of job-scheduling strategies for grid computing // LNCS. 2000. Vol. 1971. P. 191—202.

15. Sonmez O., Mohamed H., Epema D. On the benefit of processor coallocation in multicluster grid systems // IEEE Trans. on Paral. and Distrib. Syst. 2010. Vol. 21, N. 6. P. 778—789.

16. Саак А. Э. Полиномиальные алгоритмы распределения ресурсов в Grid-системах на основе квадратичной типизации массивов заявок // Информационные технологии. 2013. № 7. Приложение. 32 с.

17. Саак А. Э. Управление ресурсами и заявками пользователей в Grid-системах с централизованной архитектурой // Труды XII Всероссийского совещания по проблемам управления ВСПУ-2014. Москва, 16—19 июня 2014 г. М.: Институт проблем управления им. В. А. Трапезникова РАН, 2014. С. 7489—7498.

18. Саак А. Э. Локально-оптимальные ресурсные распределения // Информационные технологии. 2011. № 2. С. 28—34.

19. Саак А. Э. Алгоритмы диспетчеризации в Grid-системах на основе квадратичной типизации массивов заявок // Информационные технологии. 2011. № 11. С. 9—13.

20. Саак А. Э. Сравнительный анализ полиномиальных алгоритмов диспетчеризации в Grid-системах // Информационные технологии. 2012. № 9. С. 28—32.

21. Саак А. Э. Уровневые алгоритмы диспетчеризации массивами заявок кругового типа в Grid-системах // Известия ЮФУ. Технические науки. 2015. № 6 (167). С. 223—231.

22. Саак А. Э. Кольцевые алгоритмы диспетчеризации массивами заявок в Grid-системах // Информационные технологии. 2016. № 3. С. 163—169.

23. Саак А. Э. Диспетчеризация заявок кругового типа в Grid-системах // Информационные технологии. 2016. № 1. С. 37—41.

24. Jepsen C. Dissections into 1:2 rectangles // Discrete Mathematics. 1996. Vol. 148. P. 107—117.

25. Friedman E. Dominoes in squares. 2014. URL: <http://www2.stetson.edu/~efriedma/domino/>

26. Korf R., Moffitt M., Pollack M. Optimal rectangle packing // Annals of Operations Research. 2010. Vol. 179. Iss. 1. P. 261—295.

27. Huang E., Korf R. Optimal rectangle packing: an absolute placement approach // Journal of Artificial Intelligence Research. 2013. Vol. 46. P. 47—87.

28. Friedman E. Tiling Almost Squares. 2013. URL: <http://www2.stetson.edu/~efriedma/almost/>

Scheduling of Sets of Hyperbolic-Type and Circular-Type Tasks in Grid-Systems

The annotation: Users' growing demand for computational capability, along with the progress in technologies, stimulate the transition from metacomputing to Grid-computing at the end of past century, which is still stepping up today jointly with such distributed computing paradigms as cloud computing and jungle computing. This paper considers computer tasks with a priori defined time of execution, in which the number of required processors is specified by a user at the moment of their delivery to the system. In these conditions when assigning sequentially numbered processors for task handling, the task of schedule generation by a scheduling algorithm is equivalent to the task of resource rectangles distribution. User's task, which needs service from Grid system's scheduler, is presented as a resource rectangle with horizontal and vertical dimensions, correspondingly, are set equal to the number of processors and time units required to complete a task. Grid-systems with centralized structure of their scheduling system which consist of sites which contain parallel systems which function in the mode of multisite scheduling are simulated by the resource quadrant. The author defined the resource rectangles environment, as a theoretical base of algorithmic support of polynomial complexity scheduling for the purpose of distribution management of computer and time resources and users' tasks in Grid-systems. In the environment of resource rectangles versions of level algorithms by height and by length are suggested. They allow scheduling by the sets of resource rectangles which don't have the feature of monotony of horizontal and vertical dimensions. This paper poses a question of suggested polynomial algorithms adaptiveness for the sets which consists of tasks of hyperbolic and quadratic, and circular and quadratic types. The comparative analysis allows recommending some of the algorithms to use in Grid systems with centralized structure.

Keywords: Grid system, centralized structure of scheduling system, multisite mode of task service, non-preemptive task service, circular-type task, hyperbolic-type task, Non-Euclidean heuristic measure, polynomial complexity of an algorithm, level algorithm by height, level algorithm by length

References

1. Schwiegelshohn U., Badia R., Bubak M., Danelutto M., Dustdar S., Gagliardi F., Geiger A., Hluchy L., Kranzlmüller D., Laure E., Priol T., Reinefeld A., Resch M., Reuter A., Rienhoff O., Rüttger T., Shot P., Talia D., Ullmann K., Yahyapour R., Voigt G. Perspectives on grid computing, *Future Generation Computer Systems*, 2010, vol. 26, Iss. 8, pp. 1104–1115.
2. Bencivenni M., Michelotto D., Alfieri R., Brunetti R., Ciccanti A., Cesini D., Costantini A., Fattibene E., Gaido L., Misurelli G., Ronchieri E., Salomoni D., Veronesi P., Venturi V., Vistoli M. Accessing Grid and Cloud services through a scientific web portal, *J. Grid Computing*, 2015, vol. 13, Iss. 2, pp. 1159–1175.
3. Kahanwal B., Singh T. The distributed computing paradigms: p2p, grid, cluster, cloud, and jungle, *International Journal of Latest Research in Science and Technology*, 2012, vol. 1, Iss. 2, p. 183–187.
4. Buyya R., Broberg J., Goscinski A. (eds.). *Cloud computing: principles and paradigms*. Hoboken: John Wiley & Sons Inc., 2011, 644 p.
5. Seinstra F., Maassen J., Nieuwpoort R., Drost N., Kessel T., Werkhoven B., Urbani J., Jacobs C., Kielmann T., Bal H. Jungle computing: Distributed supercomputing beyond clusters, grids, and clouds. M. Cafaro, G., Aloisio (eds.). *Grids, Clouds and Virtualization*. London: Springer-Verlag, 2011, 235 p., pp. 167–197.
6. Li M., Baker M. *The grid: core technologies*. Chichester: John Wiley & Sons Ltd, 2005, 452 p.
7. Pugliese A., Talia D., Yahyapour R. Modeling and Supporting Grid Scheduling, *J. Grid Computing*, 2008, vol. 6, Iss. 2, pp. 195–213.
8. Rahman M., Ranjan R., Buyya R., Benatallah B. A taxonomy and survey on autonomic management of applications in grid computing environments, *Concurrency Computat: Pract. Exper*, 2011, vol. 23, Iss. 16, pp. 1990–2019.
9. Ye D., Zhang G. On-Line Scheduling of Parallel Jobs, *LNCS*, 2004, vol. 3104, pp. 279–290.
10. Feitelson D., Rudolph L. Toward convergence in job schedulers for parallel supercomputers, *LNCS*, 1996, vol. 1162, pp. 1–26.
11. Feitelson D., Rudolph L., Schwiegelshohn U., Sevcik K., Wong P. Theory and practice in parallel job scheduling, *LNCS*, 1997, vol. 1291, pp. 1–34.
12. Majumdar S., Parsons E. Parallel job scheduling: a performance perspective, *LNCS*, 2000, vol. 1769, pp. 233–252.
13. Caramia M., Giordani S., Iovanella A. Grid scheduling by on-line rectangle packing, *Networks*, 2004, vol. 44, no. 2, pp. 106–119.
14. Hamscher V., Schwiegelshohn U., Streit A., Yahyapour R. Evaluation of job-scheduling strategies for grid computing, *LNCS*, 2000, vol. 1971, pp. 191–202.
15. Sonmez O., Mohamed H., Epema D. On the benefit of processor co location in multicluster grid systems, *IEEE Trans. on Paral. and Distrib. Syst.*, 2010, vol. 21, no. 6, pp. 778–789.
16. Saak A. Eh. Polinomialnye algoritmy raspredeleniya resursov v Grid-sistemakh na osnove kvadrachnoy tipizatsii massivov zayavok [Polynomial algorithms for resource allocation in Grid-based systems for quadratic typing, arrays applications], *Informacionnye tekhnologii* [Information Technologies], 2013, no. 7. Prilozhenie, 32 p.
17. Saak A. Eh. Upravleniye resursami i zayavkami polzovatelye v Grid-sistemakh s centralizovannoy arkhitektyroy [Resource and multi-processor task management in Grid system of centralized architecture], *Trudy XII Vserossiyskogo soveshanya po problem upravleniya VSPU-2014*. Moskva, 16 ijunya — 19 ijunya 2014 g. [Proceedings of XII all-Russian conference "Control problems" RCCP'2014. Moscow, 16 June — 19 June 2014]. Moscow: Institut problem upravleniya im. V. A. Trapeznikova RAN, 2014, pp. 7489–7498.
18. Saak A. Eh. Lokalno-optimalnye resursnye raspredeleniya [Locally optimal resource allocation], *Informatsionniye tekhnologii* [Information Technologies], 2011, no. 2, pp. 28–34.
19. Saak A. Eh. Algoritmy dispatcherizatsii v Grid-sistemakh na osnove kvadrachnoy tipizatsii massivov zayavok [Algorithms scheduling in Grid-based systems for quadratic typing, arrays applications], *Infoniatsionnye tekhnologii* [Information Technologies], 2011, no. 11, pp. 9–13.
20. Saak A. Eh. Sravnitelniy analiz polinomialnykh algoritmov dispatcherizatsii v GRID-sistemakh [Comparative analysis of poly-

nomial algorithms for scheduling in GRID-systems], *Informatsionnye tekhnologii* (Information Technologies), 2012, no. 9, pp. 28–32.

21. **Saak A. Eh.** Urovnevye algoritmy dispatcherizatsii massivamy zayavok krygovogo tipa v Grid-sistemakh [Level algorithms of scheduling by circle type task sets in grid systems], *Izvestiya YuFU. Tekhnichskie nauki* [Izvestiya SFedU. Engineering Sciences], 2015, no. 6 (167), pp. 223–231.

22. **Saak A. Eh.** Kolcivie algoritmy dispatcherizatsii massivamy zayavok v Grid-sistemakh [Ring algorithms for scheduling in grid systems by sets of tasks], *Informatsionnye tekhnologii* [Information Technologies], 2016, no. 3, pp. 163–169.

23. **Saak A. Eh.** Dispatcherizatsiya zayavok krygovogo tipa v Grid-sistemakh [Circular-typed multiprocessor tasks scheduling in

grid systems], *Informatsionnye tekhnologii* (Information Technologies), 2016, no. 1, pp. 37–41.

24. **Jepsen C.** Dissections into 1:2 rectangles, *Discrete Mathematics*. 1996, vol. 148, pp. 107–117.

25. **Friedman E.** *Dominoes in squares*. 2014. URL: <http://www2.stetson.edu/~efriedma/domino/>

26. **Korf R., Moffitt M., Pollack M.** Optimal rectangle packing, *Annals of Operations Research*, 2010, vol. 179, Iss. 1, pp. 261–295.

27. **Huang E., Korf R.** Optimal rectangle packing: an absolute placement approach, *Journal of Artificial Intelligence Research*, 2013, vol. 46, pp. 47–87.

28. **Friedman E.** *Tiling Almost Squares*. 2013. URL: <http://www2.stetson.edu/~efriedma/almost/>

УДК 004.912; 519.254

А. С. Мохов¹, ассистент, e-mail: asmokhov@mail.ru,
В. О. Толчеев², д-р техн. наук, проф., e-mail: tolcheevvo@mail.ru
НИУ "МЭИ"

Способы учета структуры научных документов в задачах обработки и анализа текстовой информации

Рассмотрены процедуры обработки и анализа текстовой информации на основе учета структуры документа. Приведены основные модели представления текстов в задачах машинного обучения. Показана эффективность использования частично структурированных моделей для информационного поиска, автоматического аннотирования, выявления нечетких дубликатов и классификации. Наряду с известными подходами в работе излагаются предложенные авторами процедуры, учитывающие особенности двуязычных библиографических документов и позволяющие проводить высокоточную классификацию.

Ключевые слова: интеллектуальный анализ текстовых данных, модель текстового документа, информационный поиск, автоматическое аннотирование, выявление нечетких дубликатов, классификация двуязычных библиографических документов

Введение

В рамках работ по интеллектуальному анализу текстовых данных (Text Mining) решаются задачи поиска информации, кластеризации, фильтрации и классификации документов, автоматического аннотирования и извлечения ключевых понятий из текстов, выявления плагиата и нечетких дубликатов в документальных массивах, создания онтологий и построения моделей предметных областей. Общей особенностью вышеуказанных задач является их слабая формализованность, сложность математического описания текстов.

Под структурой документа далее понимается его логическое построение в виде взаимосвязанных фрагментов. Структура текстов во многом обуславливается функциональным стилем, которому соответствует публикация. Различают публицистический, официально-деловой, научно-технический, разговорный, художественный стили. В данной работе рассматривается научно-технический стиль, который используется в монографиях, статьях, учебни-

ках, диссертациях, отчетах по НИР и т. д. Для научных статей в качестве фрагментов целесообразно рассматривать название, аннотацию, ключевые слова, введение, основную часть, заключение. Для учебного пособия или монографии это могут быть главы, подглавы, параграфы и т. п.

Характерной особенностью научно-технических документов является наличие наряду с полным текстом короткого библиографического описания (БО), включающего имена и фамилии авторов, название, аннотацию, ключевые слова, место издания и другую вспомогательную информацию. Чаще всего БО представлены на двух языках: русском и английском. В библиографических описаниях, как и в других коротких текстах (сообщения электронной почты, новостная информация, описания товаров, инструкции по применению, анкеты, справочники), наиболее четко прослеживается общая структура. В простейшем случае библиографический научный документ может быть описан кортежем $\langle T, A, K \rangle$, в который включены название (title) — T , аннотация (abstract) — A , ключевые слова (key