

УДК 681.3.06: 681.323 (519.6)

С.С. Белоконова

ПОИСК ОБЪЕКТОВ РАЗЛИЧНЫХ ТИПОВ ПО НЕСКОЛЬКИМ ПРИЗНАКАМ НА ОСНОВЕ СОРТИРОВКИ

Описана схема детерминированного поиска на основе сортировки данных различных типов одновременно по нескольким маскам. Метод обладает параллелизмом в силу максимальной параллельности сортировки и параллелизма обработки отдельно взятых фрагментов. Поиск конструируется как идентификация с помощью сортировки экстремальных элементов числовой последовательности, сопоставленной просматриваемым фрагментам, файлам или группам объектов.

Схема; поиск.

S.S. Belokonova

SEARCH OF OBJECTS OF VARIOUS TYPES IN SEVERAL SIGNS ON THE BASIS OF SORTING

The scheme of the determined search on the basis of sorting the given various types simultaneously on several masks is described. The method possesses parallelism by virtue of the maximal parallelism of sorting and parallelism of processing of separately taken fragments. Search is designed as identification by means of sorting extreme elements of the numerical sequence compared looked through fragments, files or groups of objects.

Scheme; Search.

Введение и постановка вопроса. Проблема поиска, сбора и обработки информации принадлежит к числу основных задач информатики. Ее актуальность возрастает с ростом объема информации в электронном виде, с ростом ресурсов, доступных в сети Internet. Особую актуальность приобрели вопросы, связанные с поиском и распознаванием оцифрованной информации различного формата и типа данных, включая текстовую, графическую, аудио- и видеoinформацию. Существующие методы и подходы не вполне обеспечивают точность, релевантность результатов поиска запросу, не совмещают в требуемой мере поиск с распознаванием. В частности, это относится к поиску данных различных типов.

Оператор локализации экстремальных элементов последовательности. В основе схемы лежит применение адресной устойчивой распараллеливаемой сортировки по ключу, сохраняющей на выходе входные индексы упорядоченных элементов, причем в порядке их расположения после сортировки. Одна из таких сортировок представляет собой разновидность сортировки подсчетом, для выполнения которой применяется матрица сравнений (МС). На пересечении i -й строки и j -го столбца МС находится результат сравнения элементов $c[i]$ и $c[j]$, отмечаемый знаком «+», если $c[i] < c[j]$, знаком «-», если $c[i] > c[j]$, и знаком «0», если $c[i] = c[j]$. Число нулей и плюсов в j -м столбце над диагональю, включая диагональный элемент, складывается с числом минусов j -й строки справа от диагонали. Значение суммы k становится индексом j -го элемента в отсортированном массиве: $c1[k] := c[j]$, с тем же индексом запоминается входной

номер переставленного элемента: $e[k] := j$. Дополнительной составляющей схемы поиска является оператор локализации экстремальных элементов. Локально минимальный элемент массива (в дальнейшем – минимум) определяется как элемент меньший предшествующего и не больший последующего по отношению « \leq ». Оператор локализации минимумов последовательности из n элементов имеет вид

```

j:=1; while j<= n do begin
  FOR L:=1 TO j-1 do if abs(e[j]-e[j-L])<=eps then goto
22; Writeln (' ',c[e[j]],' ',e[j]);
22:  j:=j+1; end;

```

Присоединение к процедуре сортировки оператора локализации минимумов влечет программную идентификацию всех локально минимальных элементов входного массива в окрестности радиуса eps . При этом eps задается априори и произвольно фиксировано. Аналогично могут быть идентифицированы локальные максимумы [1,2], с изменением знака неравенства на противоположный – глобальные экстремумы [1]. На основании параллелизма сортировки экстремумы идентифицируются по максимально параллельной схеме.

Поиск объектов различных типов одновременно по нескольким признакам. Исходная (мультипликативная [1,2]) схема строится следующим образом. Исследуемой строке сопоставляется одномерный числовой массив, количеством элементов которого совпадает с количеством слов заданной строки. Элементы сопоставленного числового массива, соответствующие различным маскам, перемножаются, причем сомножители априори задаются так, чтобы произведения взаимно однозначно соответствовали различным маскам, а локальные минимумы в их последовательности единственным образом соотносились с полным набором масок. Сопоставление массиву слов Sl числового массива $C = (c_1, c_2, \dots, c_m)$ выполняется по соотношениям:

$$c_i = \prod_{j=1}^n a_j, a_j = \begin{cases} j+1, & \text{если } i - \text{е слово совпадает} \\ & \text{с } j - \text{й маской поиска} \\ j+1+n, & \text{если } i - \text{е слово не совпадает} \\ & \text{с } j - \text{й маской поиска} \end{cases} \quad (1)$$

При этом индексы элементов $c[i]$ совпадают с индексами элементов $Sl[i]$. К отсортированному массиву C применяется оператор локализации минимумов. Если локализация минимума выполняется в окрестности наперед заданного радиуса, не меньшего числа сгруппированных масок, то идентифицированным окажется искомое сочетание одновременно нескольких масок. Тем самым можно вести поиск не по разрозненному сочетанию нескольких масок, а по их взаимосвязанному положению в исследуемом массиве слов.

Мультипликативная схема [1,2] позволяет выполнить поиск объекта в том случае, если объект обладает не более чем одним свойством и при этом принадлежит одному типу данных. В частности, слово исследуемого текста может совпадать не более, чем с одной маской. С целью видоизменения поиска для объектов различных типов мультипликативная схема поиска модифицируется следующим образом.

Пусть дан массив Sl из m слов, в котором требуется выполнить поиск по заданной последовательности из n масок $M = (m_1, m_2, \dots, m_n)$, расположен-

ных в определенном и фиксированном порядке. На первом шаге массиву слов Sl сопоставляется двумерный числовой массив $mass[n, m]$, где n – число масок поиска, m – число слов в исследуемом массиве. Элементы массива $mass$ первоначально полагаются равными нулю. Рассматривается первая маска поиска m_1 . Исследуемый массив слов просматривается, начиная с номера $i=1$, и i -е слово строки сравнивается с первой маской поиска m_1 . Если i -е слово массива Sl совпадает с маской m_1 , то элементу $mass[1, i]$ присваивается число 1, в противном случае значение равно 0. Затем номер i увеличивается на 1 и выполняется переход к следующему слову. Процесс продолжается, пока не будут просмотрены все слова исследуемого массива. Второй проход повторяет действия, описанные в первом проходе, для второй маски поиска m_2 . Такой процесс выполняется для всех n масок поиска и массиву Sl сопоставляется массив $mass[n, m]$:

$$mass[j, i] = \begin{cases} 1, & \text{если } i\text{-е слово совпадает с } j\text{-й маской поиска,} \\ 0, & \text{если } i\text{-е слово не совпадает с } j\text{-й маской поиска.} \end{cases} \quad (2)$$

В (2) и ниже $i=1, 2, \dots, m$, $j=1, 2, \dots, n$.

Для получения одномерного числового массива на основе (2) формируется промежуточный двумерный массив $mass1$:

$$mass1[j, i] = \begin{cases} j+1, & \text{если } mass[j, i] = 1, \\ j+1+n, & \text{если } mass[j, i] = 0. \end{cases} \quad (3)$$

Затем путем перемножения элементов столбцов (3) формируется одномерный массив c :

$$c_i = \prod_{j=1}^n mass1[j, i]. \quad (4)$$

Изложенную схему можно следующим образом перевести в схему поиска по совпадению с множеством признаков или свойств. С помощью соотношений (3), (4) массив (2) переводится в одномерный числовой массив при условии, что при каждом i исследуемому объекту с номером i соответствует i -й столбец массива $mass$, в котором не более одного единичного значения. Формирование массива $mass$ происходит с помощью проверки соответствия или несоответствия исследуемого объекта заданному признаку при фиксированном порядке обхода объектов [1].

Конструкция используется для идентификации объектов одновременно различных типов или, в более общем случае, объектов, характеризующихся свойствами различной природы. Схема имеет то ограничение, что каждый объект может удовлетворять не более чем одному признаку (ниже «признак» и «свойство» – синонимы).

Схема видоизменяется для объектов, обладающих набором свойств.

Каждому объекту с заданным набором признаков следующим единственным образом сопоставляется числовое значение.

Пусть дано проиндексированное множество V из m однотипных объектов, в котором требуется выполнить поиск по заданной последовательности из n признаков $M = (m_1, m_2, \dots, m_n)$, расположенных в определенном

фиксированном порядке. Как и прежде, формируются массивы $mass$, $massI$ и c , но при этом соотношения (2), (3) заменяются соотношениями (5), (6):

$$mass[j,i] = \begin{cases} 1, & \text{если } i\text{-й объект соответствует} \\ & j\text{-му признаку поиска,} \\ 0, & \text{если } i\text{-й объект не соответствует} \\ & j\text{-му признаку поиска,} \end{cases} \quad (5)$$

$$massI[j,i] = \begin{cases} p[j], & \text{если } mass[j,i] = 1, \\ p[j+n], & \text{если } mass[j,i] = 0, \end{cases} \quad (6)$$

где p - конечная последовательность из $2n$ упорядоченных простых чисел (2, 3, 5,...). Имеет место взаимно однозначное соответствие между набором признаков, присущих объекту $V[i]$, и сопоставленным числом $c[i]$: значение $c[i]$ по теореме о простых числах единственным образом разложимо на простые множители, по этому разложению данное число обратимо указывает на те признаки (5), которыми обладает объект соответственно простым сомножителям $massI[j,i]$ из (6).

Каждому набору признаков, присущих объекту, соответствует единственное значение произведения элементов столбца, образующее элемент сопоставленной числовой последовательности (4).

Числовые идентификаторы наборов признаков могут быть априори рассчитаны, на этой основе выполняется поиск по полному и частичному набору заданных свойств. После сортировки массива c из (4) применяется оператор локализации минимумов, который по наличию локальных минимумов идентифицирует индексы искоемых объектов по аналогии с тем, как идентифицируются индексы слов, совпадающих с масками поиска [1,2].

На основе совпадения индексов числового массива и массива исходных объектов выполняется переход от числовых экстремумов к искомым объектам в исходном виде.

Числовые идентификаторы, по которым локализуются экстремальные значения, могут быть рассчитаны наперед. Поэтому анализ результатов поиска может вестись по совпадению с такими числовыми значениями в локализованной окрестности.

Поиск на этой основе может выполняться не только по полному набору заданных свойств, но и по их частичному набору. При этом путем анализа окрестности локализованных минимумов можно идентифицировать не только один объект, но и группу таких объектов.

Пример 1. Среди изображений фрагментов кардиограмм на рис. 1 (представленных на входе метода оцифрованными элементами массива $V = (V_1, V_2, \dots, V_7)$) указать те, которые удовлетворяют признакам:

1. Преобладание в QRS-комплексе R-зубца над S-зубцом.
2. Отсутствие аритмии (максимальное отклонение от среднего расстояния между R-зубцами более 10%).
3. Средняя частота сердечных сокращений в пределах нормы (60-90 ударов в минуту).

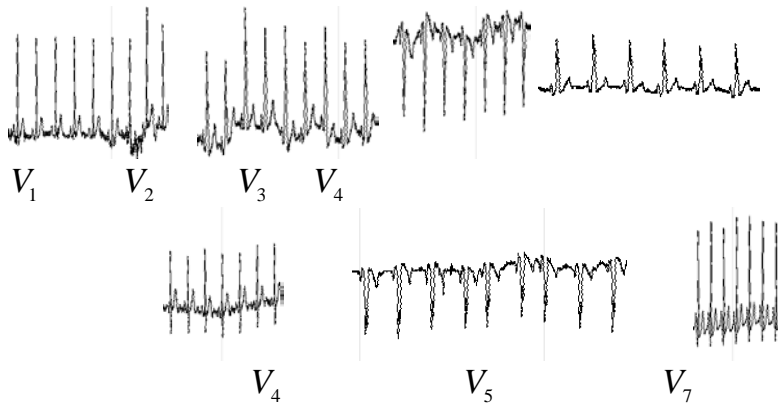


Рис. 1. Фрагменты кардиограмм

По соотношению (5) на первом шаге массиву V сопоставляется двумерный массив $mass$. Первоначально каждый элемент этого массива равен нулю. Затем в каждом объекте V_i ($i=1,2,\dots,7$) проверяется наличие каждого признака. Если i -я кардиограмма отвечает j -му признаку, то элементу $mass[j,i]$ присваивается 1, иначе 0. Массив $mass$ примет вид

Таблица 1

Вид модифицированного массива $mass$

	V_1	V_2	V_3	V_4	V_5	V_6	V_7
m_1	1	1	0	1	1	0	1
m_2	0	1	1	1	1	0	1
m_3	1	1	1	0	1	0	0

Образование массива $mass1$ и одномерного числового массива C в соответствии с (4), (6) иллюстрирует табл. 2.

Таблица 2

Сопоставление модифицированных массивов $mass1$ и с массиву $mass$

	V_1	V_2	V_3	V_4	V_5	V_6	V_7
m_1	2	2	7	2	2	7	2
m_2	11	3	3	3	3	11	3
m_3	5	5	5	13	5	13	13
массив C	110	30	105	78	30	1001	78

По построению схемы (4) – (6) глобально минимальный элемент соответствует объекту, обладающему всеми искомыми свойствами, глобально максимальный – объекту, не обладающему ни одним свойством. Иные числовые идентификаторы указывают объекты с частью признаков.

В массиве C идентифицируются экстремумы. Поскольку искомый объект удовлетворяет одновременно всем условиям поиска, то требуется найти гло-

бально минимальный элемент. Достаточно найти локально минимальный элемент, равный 30 (значение априори рассчитано и совпадает с глобальным минимумом). Локально минимальные элементы, равные 30, идентифицируются с индексами 2 и 5. Следовательно, искомые объекты на рис. 1 идентифицируются по номерам 2 и 5. Все остальные промежуточные значения соответствуют объектам, обладающим частью признаков.

Схему (4) – (6) можно адаптировать к поиску группы объектов, типы которых различны между собой. В качестве примера можно взять поиск каталога, содержащего файлы различных типов. Каталог рассматривается как объект, искомый в нем файл – как признак поиска. Каталог с такими признаками идентифицируется согласно (4) – (6), при этом схема поиска без изменения применима к файлам различного типа, если типы совместимы или могут быть приведены к одному типу в операционной системе, например, *.txt, *.dpr, *.pas, *.xls, *.doc. Для реализации поиска объектов, которые не приводятся к одному и тому же типу, например, *.dat, *.bmp, *.jpg, предлагается описываемая в дальнейшем специальная схема. Исследуемому файлу сопоставляется группа файлов, например, текстовый файл *.txt, непосредственно содержащий текст исследуемого файла, графические файлы *.bmp, содержащие внедренные в исследуемый файл рисунки, типизированный файл *.dat, который содержит все числовые значения, присутствующие в данном файле и т.д. Поиск в исходном файле сводится к поиску в файлах сопоставленной группы. Каждая группа рассматривается как объект со многими признаками поиска, а наличие или отсутствие в объекте файла, соответствующего маске поиска, как признак объекта для его идентификации. Таким образом, используется рекуррентное вложение схем поиска. Объединенный результат поиска можно получить по схеме с использованием (4) – (6), что позволяет свести поиск к идентификации локальных экстремумов числовой последовательности и в этом сравнительно общем случае.

Поиск в файлах *.dat заданного числового значения с точностью до *eps1* можно выполнить на основе схемы поиска элемента в числовом массиве с заданной границей погрешности. Информация из типизированного числового файла считывается в числовой массив с сохранением порядка следования элементов. Далее без изменений применяется схема, описанная в [1]. Схема обобщается на поиск самих типизированных числовых файлов по признаку наличия в нем заданного числового значения с допустимой границей погрешности. С целью поиска файлов типа *.bmp каждому файлу (предполагается, что файл содержит только один рисунок) сопоставляется вектор распознавания. Этот вектор можно сформировать на основе какой-либо известной схемы. Если вектор сформирован, то для текущего рисунка находится норма разности между вектором исследуемого рисунка и эталонным вектором. Рисунок считается найденным, если норма разности соответственной пары векторов не превосходит заданной границы погрешности.

В [1] доказаны следующие утверждения.

Предложение 1. Массив *c* единственным образом идентифицирует изображение.

Следствие 1. Описанное преобразование изображения обратимо.

С некоторыми ограничениями схема распространяется на поиск числовых, формульных, текстовых фрагментов и данных других типов. По построению изложенный подход обобщается на поиск объектов одновременно нескольких произвольных типов, если эти типы реализованы в языке программирования [1].

Таким образом, изложена схема поиска по нескольким маскам объектов различных типов, при этом схема совмещает поиск с распознаванием.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Белоконова С.С. Разработка и исследование схем детерминированного поиска на основе сортировки с приложением к идентификации оцифрованных объектов различных типов / Автореферат диссертации на соискание ученой степени кандидата технических наук. – Таганрог: Изд-во ТТИ ЮФУ. – 2007. – 20 с.
2. Ромм Я.Е., Белоконова С.С. Схема поиска данных различных типов по нескольким маскам на основе сортировки // Известия вузов. Северо-Кавказский регион. Техн. науки. Специальный выпуск «Математическое моделирование и компьютерные технологии». 2006. – С. 3 – 8.

Белоконова Светлана Сергеевна

Таганрогский государственный педагогический институт

E-mail: belokonova@mail.ru

347936, г. Таганрог, ул. Инициативная, д. 48. Тел: 88634 60-18-99

Belokonova Svetlana Sergeevna

Taganrog State Pedagogical Institute

E-mail: belokonova@mail.ru

48, Initsiativnaia, Taganrog, 347936. Phone: 88634 60-18-99

УДК 539:620.179.16

Е. С. Огурцов

ИССЛЕДОВАНИЕ КОМБИНИРОВАННЫХ ЭЛЕКТРОДИНАМИЧЕСКИХ ВЕТРОГЕНЕРАТОРОВ С ВЕРТИКАЛЬНОЙ ИЛИ ГОРИЗОНТАЛЬНОЙ ОСЬЮ

В статье представлены результаты последних разработок по созданию оптимальных комбинированных электродинамических ветрогенераторов с вертикальным или горизонтальным валом. Даны результаты численного и экспериментального моделирования. Показана фотография устройства, вырабатывающего электрическую энергию от невозобновляемого источника.

Оптимальные комбинированные электродинамические ветрогенераторы, устройство-маховик; статор; ротор; генератор электрической энергии; постоянные магниты; невозобновляемые источники энергии.

E.S. Ogurtsov

RESEARCH COMBINED ELECTRODYNAMIC WINDGENERATORS WITH VERTICAL AND HORIZONTAL AXIS

In article results of last workings out on creation optimum combined electrodynamic windgenerators with vertical and horizontal shaft are presented. Results of numerical and experimental modelling are yielded. The photo of the device developing electric energy from not renewed source is shown.