

5. Ромм Я.Е., Дзюба А.С. Идентификация рукописных символов с применением подстановки индексов при сортировке полярных координат // Известия ЮФУ. Технические науки. 2013. – № 7 (144). – С. 154-160.
6. Демин А.А. Обзор интеллектуальных систем для оценки каллиграфии // Инженерный вестник (МГТУ им. Н.Э. Баумана). Электронный журнал. – 2012. – № 9.
7. Гайдуков Н.П., Савкова Е.О. Обзор методов распознавания рукописного текста // Международная научно-техническая конференция студентов, аспирантов и молодых учёных "Информационно-управляющие системы и компьютерный мониторинг – 2012".
8. Садыхов Р.Х., Ваткин М.Е. Алгоритм обучения нейронной сети «неокогнитрон» для распознавания рукописных символов // Весті НАН Беларусі. Сер. фіз.-тэхн. навук. 2002. – № 3. – С. 1.
9. Садыхов Р.Х., Дудкин А.А. Обработка изображений и идентификация объектов в системах технического зрения // Штучний інтелект. 2005 – № 3. – С. 670-679.

Статью рекомендовал к опубликованию д.т.н., профессор В.П. Карелин.

Ромм Яков Евсеевич – Федеральное государственное бюджетное образовательное учреждение «Таганрогский государственный педагогический институт имени А.П. Чехова»; e-mail: romm@list.ru; 347926, г. Таганрог, ул. Инициативная, 48; тел.: 89094081126; кафедра информатики; д.т.н.; профессор.

Дзюба Андрей Сергеевич – e-mail: dzuba_ni@list.ru; тел.: 89094080776; кафедра информатики; аспирант.

Romm Yakov Evseevich – Federal State Budget Educational Establishment of Higher Professional Education «Taganrog State Pedagogical Institute named after A.P. Chehov»; e-mail: romm@List.ru; 48, Initsiativnaya, Taganrog, 347926, Russia; phone: +79094081126; the department of information science; dr. of eng. sc.; professor.

Dzuba Andrew Sergeevich – e-mail: dzuba_ni@list.ru; phone: +79094080776; the department of information science; postgraduate student.

УДК 681.3.06: 681.323 (519.6)

Я.Е. Ромм, С.С. Белоконова

ДЕТЕРМИНИРОВАННЫЙ ПОИСК ДАННЫХ РАЗЛИЧНОГО ТИПА ПО ЭКСТРЕМАЛЬНЫМ ЗНАЧЕНИЯМ ОТСОРТИРОВАННЫХ ЭЛЕМЕНТОВ

Излагается распараллеливаемый метод детерминированного поиска на основе идентификации локально экстремальных элементов числовой последовательности, которая взаимно однозначно сопоставляется исследуемому массиву данных фиксированного типа. Способ сопоставления использует разложение элементов последовательности на простые множители. Метод использует алгоритм максимально распараллеливаемой сортировки с взаимно-однозначным соответствием входных и выходных индексов, обобщается на поиск одновременно по произвольно заданному конечному количеству масок различного типа в множестве файлов. Особенностью метода является его функциональная возможность выполнять поиск как данных нечислового, так и числового типа. Многообразие масок используется для поиска в зависимости от их взаимного сочетания, от индексного расстояния между ними. Параллелизм основан на максимальной параллельности сортировки и параллелизме обработки отдельно взятых файлов и их фрагментов, достигается оценка временной сложности максимально параллельного поиска на данной основе $O(1)$. Количество процессоров зависит как от размеров обрабатываемых файлов или числа объектов, так и от количества масок или искомым свойств.

Информационный поиск; данные различных типов; параллельные алгоритмы сортировки и поиска; идентификация экстремумов на основе сортировки.

Ya.E. Romm, S.S. Belokonova

DETERMINISTIC SEARCH DIFFERENT TYPES OF DATA THE EXTREME ON VALUES OF THE SORTED ELEMENTS

The parallelizable method of the determined search on the basis of identification of locally extreme elements of numerical sequence which is one-to-one compared to a studied data file of the fixed type is stated. The comparison method uses elements decomposition of sequence on simple multipliers. The method uses algorithm of the parallelizable sorting with biunique compliance of input and output indexes. Also this method is generalized on search at the same time by randomly set final quantity of masks of various type in a set of files. Feature of a method is its functionality to carry out search as data non-numerical, and numerical type. The variety of masks is used for search depending on their mutual combination, on index distance between them. Parallelism is based on the maximum sorting parallelism and on processing parallelism overlapping of separately taken files and their fragments. In the case of most parallel search this method provides temporary complexity estimation, which equal $O(1)$. The number of processors depends as on the sizes of processed files or number of objects, and on quantity of masks or required properties.

Information search; data of different types; parallel algorithm sorting and searching; identification of extrema on the basis of sorting.

Введение. Актуальность проблемы поиска возрастает с ростом объема информации в электронном виде, с ростом ресурсов, доступных в сети Internet. В частности это относится к поиску и распознаванию оцифрованной информации различного формата и типа данных, включая текстовую, графическую, аудио- и видеоинформацию. Существующие методы не вполне обеспечивают точность, релевантность результатов поиска запросу, в особенности в случае данных различных несовместимых типов, при обработке плохо структурированной информации. Остается актуальной разработка новых методов, позволяющих повысить эффективность, расширить область применения схем поиска, включая полнотекстовый поиск при организации электронных библиотек и каталогов, где проблема определяется растущим количеством пользователей, не обладающих профессиональными навыками при поиске информации на языке запросов.

Постановка задачи. Излагается построение метода детерминированного поиска одновременно по произвольно заданному конечному количеству масок различного типа. Многообразие масок предполагается использовать для поиска в зависимости от их сочетания друг с другом, от расстояния между ними, от взаимного расположения всех этих и частичных их комбинаций. Помимо того, на данной основе конструируется мера сходства и отличия от эталонных значений просматриваемых фрагментов файла, обсуждается параллелизм конструируемой схемы.

Метод конструируется на основе сортировки и идентификации экстремальных элементов числовой последовательности, сопоставленной исследуемому множеству объектов, в окрестности произвольного радиуса. Базовые конструкции метода излагались в [1, 2], а также в [3, 4].

Описание метода. Пусть дано проиндексированное множество (массив) V из m однотипных объектов, в котором требуется выполнить поиск по заданной последовательности признаков $M = (m_1, m_2, \dots, m_n)$, расположенных в произвольно фиксированном порядке. Первоначально массиву объектов V сопоставляется двумерный числовой массив $mass[n, m]$, каждый элемент которого равен нулю.

Просматриваются все объекты, начиная с номера $i = 1$, при каждом значении i объекту с номером i сопоставляются n числовых значений в зависимости от наличия в объекте соответствующего признака. Если первый объект соответствует первому

признаку, то элементу $mass[1,1]$ присваивается числовое значение, равное единице, в противном случае значение остается нулевым. Затем выполняется проверка наличия второго признака без изменения номера объекта. Если объект удовлетворяет второму признаку, то элементу $mass[2,1]$ присваивается числовое значение, равное единице, в противном случае значение остается нулевым. В продолжение процесса при проверке n -го признака элементу $mass[n,1]$, аналогично, либо присваивается единица, либо ноль. На этом проверка первого объекта заканчивается. Далее, выполняется проверка второго объекта на наличие тех же n признаков, идентифицируемых в точности таким же образом, как у первого объекта. Вторые n числовых значений записываются в элементы $mass[1,2], mass[2,2], \dots, mass[n,2]$. Процесс продолжается, пока не будут рассмотрены все m объектов. При каждом i объекту с номером i соответствуют элементы массива $mass : (mass[1,i], mass[2,i], \dots, mass[n,i])$. При этом

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

где i -му объекту соответствует i -й столбец, в котором количество единичных значений $\leq n$ соответствует количеству признаков, которым отвечает объект.

Массиву $mass$ сопоставляется массив $mass1$ по соотношению:

$$mass1[j,i] = \begin{cases} p[j], & \text{если } mass[j,i] = 1, \\ p[j+n], & \text{если } mass[j,i] = 0, \end{cases} \quad (2)$$

и формируется одномерный числовой массив

$$c_i = \prod_{j=1}^n mass1[j,i], \quad (3)$$

при этом в (2) p – массив последовательных простых чисел $(2, 3, 5, \dots)$. Количество элементов массива p равно $2n$.

Единица, записанная в элемент $mass[j,i]$, означает совпадение i -го объекта с j -м признаком, ноль – несовпадение, и этому объекту будет соответствовать единственное произведение простых чисел, равное c_i . На самом деле имеет место взаимно-однозначное соответствие. Числовое значение c_i в силу теоремы о простых числах единственным образом разложимо на простые множители, а, следовательно, число обратимо и указывает на те признаки, которые присутствуют у данного объекта соответственно простым сомножителям в (3).

После сортировки массива C из (3) применяется оператор локализации минимумов [3, 5], который по наличию локальных числовых минимумов идентифицирует индексы искоемых объектов. По идентифицированным индексам, на основе совпадения индексов элементов числового массива и массива исходных объектов, выполняется переход от числовых экстремумов к искомым объектам в исходном виде.

С помощью предложенной схемы можно выполнить поиск объектов не только на полное совпадение со всеми искомыми признаками, но и объектов, удовлетворяющих части признаков. Числовые идентификаторы наборов признаков (полное и частичное совпадение) могут быть априори рассчитанными.

Этот подход допускает видоизменение для поиска объекта с одновременно несколькими его признаками [1, 3].

В качестве примера рассматривается поиск каталога, содержащего файлы различных типов. Каталог рассматривается как объект исследования, а наличие требуемого файла в исследуемом каталоге – как признак поиска. Описанная выше схема без изменения применима к файлам различного типа в том случае, если типы объектов совместимы или могут быть приведены к одному типу в операционной системе, например, *.txt, *.dpr, *.xls, *.doc. Для реализации поиска объектов, которые не приводятся к одному и тому же типу, например, *.dat, *.bmp предлагается описываемая в дальнейшем специальная схема.

Для поиска на рассматриваемой основе файлов, содержащих разнотипные данные внутри файла, включая данные текстового типа, рисунки в формате *.bmp, числа с плавающей точкой, программные коды в различных языках программирования, можно поступить следующим образом. 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 в результате 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 также получить по схеме (1)–(3), что позволяет свести поиск к идентификации локальных экстремумов числовой последовательности и в этом сравнительно общем случае.

Поиск в файлах *.dat искомого числового значения с точностью до границы погрешности ϵ_{ps} можно выполнить на основе следующей схемы [1, 3]. Информация из типизированного числового файла считывается в числовой массив с сохранением порядка следования элементов. Чтобы в числовой последовательности $a = (a_1, a_2, \dots, a_n)$ найти заданное число b , она преобразуется к виду

$$\tilde{a} = (|a_1 - b|, |a_2 - b|, \dots, |a_n - b|). \quad (4)$$

Преобразованная последовательность сортируется. К элементам отсортированного массива применяется оператор локализации минимумов, который идентифицирует все нули. Учитывая свойство взаимно-однозначного соответствия входных и выходных индексов сортируемых элементов (для сортировки из [5]), по индексам идентифицированных нулей последовательности (4) можно обратиться к элементам исходного массива a . Схема обобщается на поиск самих типизированных числовых файлов по признаку наличия в нем заданного числового значения с допустимой границей точности.

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. Как показано в [1, 3], вектор распознавания может быть сформирован с использованием схемы (1)–(3). 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.

Иллюстрируя возможность идентификации объектов произвольной природы, с оговоркой, что их природа идентифицируема в цифровом представлении, отметим, что по предложенной схеме идентифицируются цветные растровые изобра-

жения, матричные представления графов, разновидности логических функций по таблицам истинности и объекты других предметных областей. В [1, 3] приводятся примеры такой программной идентификации. Достаточным условием идентификации является выполнимость кодирования по схеме (1)–(3) тех свойств, по которым требуется выполнить поиск или распознавание. Применимость схемы к идентификации логических функций позволяет перенести ее на случай тестирования логических и цифровых устройств. При этом объекты должны иметь фиксированный порядок взаимного расположения до конца процесса рассматриваемой обработки. В [1, 3] даны соответственные конкретные схемы и алгоритмы с формализованным описанием.

Особенностью предложенных схем поиска и идентификации объектов является их распараллеливаемость. Параллелизм основан на максимальной параллельности используемых сортировок [6, 7], на параллелизме обработки отдельно взятых файлов и их фрагментов. В самом деле, предложенные схемы поиска и распознавания не зависят от начальной точки работы алгоритма (от начальных условий): все искомые фрагменты будут идентифицированы, если они располагаются по ходу дальнейшего поиска. В [1, 3] приводится оценка временной сложности максимально параллельного поиска на основе (1)–(3), которая в пределе достигает значения $O(1)$. В этой оценке не учитывается время выполнения обмена, архитектура параллельной вычислительной системы. Оценка соответствует модели неветвящихся параллельных программ [8]. Количество процессоров зависит как от размеров обрабатываемых файлов или числа объектов, так и от количества масок или искомых свойств. Отметим, что взаимная независимость всех рассматриваемых операций формально позволяет синтезировать параллельные схемы для произвольно заданного и фиксированного числа процессоров.

Заключение. Разработана единая схема поиска на основе сортировки в качестве базового конструктивного алгоритма, которая позволяет выполнить текстовый поиск, а также поиск объектов различных типов, распознавание и идентификацию объектов одновременно по нескольким разнотипным признакам, в том числе, распознавание, идентификацию растровых изображений.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Белоконова С.С. Разработка и исследование схем применения сортировки для поиска нулей и особенностей функций с приложением к идентификации плоских изображений: Автореф. дисс. ... канд. техн. наук. – Таганрог: ТРТУ. – 2005. – 16 с.
2. Ромм Я.Е., Гуревич М.Ю., Белоконова С.С., Соловьёва И.А. Вычисление нулей и полюсов функций на основе устойчивой адресной сортировки с приложением к поиску и распознаванию // Проблемы программирования. – 2004. – № 2-3. – С. 462-472.
3. Ромм Я.Е., Белоконова С.С. Детерминированный поиск объектов различных типов на основе сортировки. – Таганрог: Изд-во ТГПИ, 2011. – 227 с.
4. Ромм Я.Е., Белоконова С.С. Поиск и идентификация объектов различных типов на основе признаков экстремального вида. – Таганрог: ТГПИ, 2008. – 49 с. Деп. в ВИНТИ от 06.06.2008, № 490-В2008.
5. Ромм Я.Е. Метод вычисления нулей и экстремумов функций на основе сортировки с приложением к поиску и распознаванию // Кибернетика и системный анализ. – 2001. – № 5. – С. 81-101.
6. Ромм Я.Е. Параллельная сортировка слиянием по матрицам сравнений // Кибернетика и системный анализ. – 1994. – № 5. – С. 3-23.

7. Ромм Я.Е. Параллельная сортировка слиянием по матрицам сравнений // Кибернетика и системный анализ. – 1995. – № 4. – С. 13-37.
8. Солодовников В.И. Верхние оценки сложности решения систем линейных уравнений / В кн.: Теория сложности вычислений. Записки научных семинаров ЛОМИ АН СССР. – Л., 1982. – Т. 118. – С. 159-187.

Статью рекомендовал к опубликованию д.т.н., профессор В.П. Карелин.

Ромм Яков Евсеевич – Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Таганрогский государственный педагогический институт имени А.П. Чехова»; e-mail: romm@list.ru; 347936, г. Таганрог, ул. Инициативная, 48; тел.: 89094081126; кафедра информатики; зав. кафедрой; д.т.н.; профессор.

Белоконова Светлана Сергеевна – e-mail: belokonova@mail.ru; тел.: 89185093262; кафедра информатики; доцент.

Romm Yakov Evseevich – Federal state educational institution of higher professional education «Taganrog state pedagogical Institute named after A.P. Chekhov»; e-mail: romm@list.ru; 48, Initiativnaya street, Taganrog, 347926, Russia; phone: +79094091126; the department of information science; head the department; dr of eng. sc.; professor.

Belokonova Svetlana Sergeevna – e-mail: belokonova@mail.ru; phone: +79185093262; the department of information science; associate professor.

УДК: 004.023, 681.518

В.И. Финаев, И.В. Пушнина

НЕЧЁТКИЕ ЛЕКСИКОГРАФИЧЕСКИЕ ОТНОШЕНИЯ В ЗАДАЧЕ РАНЖИРОВАНИЯ КРИТЕРИЕВ ПРОИЗВОДСТВА И ПОТРЕБЛЕНИЯ ЭЛЕКТРОЭНЕРГИИ*

При решении задач управления производственными процессами производства и распределения энергии существует многокритериальная задача баланса между производством и потреблением. Данная задача связана с управлением производственными мощностями энергетических предприятий. Решать эту задачу классическими методами решения распределительных задач можно, однако, данное решение может быть осуществлено только для стационарных состояний и при отсутствии последействия. Это является упущенной ситуацией и не даёт достоверного решения. В реальных системах всегда существует неопределённость, так как потребление мощности является нестационарным процессом и существует последействие. Это первая причина, требующая расширения классической распределительной задачи при неопределённых исходных данных. Другая причина состоит в том, что задача оптимального распределения производимой энергии между потребителями является многокритерильной и не решаемой классическими методами многокритериальной оптимизации. Поэтому в данной работе осуществлена постановка начальной задачи многокритериальной оптимизации, а именно, задачи ранжирования критериев в виде лексикографического отношения.

Нечёткость; множество критериев; оптимизация; ранжирование критериев; лексикографическое отношение; производство; потребление.

* Материалы статьи подготовлены в рамках выполнения работ по гранту Российского научного фонда № 14-19-01533