

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ  
«САРАТОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра математической  
кибернетики и компьютерных наук

ПОСТРОЕНИЕ СИСТЕМЫ АНАЛИЗА ЛОГОВ С ИСПОЛЬЗОВАНИЕМ  
ТЕХНОЛОГИИ MAPREDUCE

КУРСОВАЯ РАБОТА

Студента 4 курса 411 группы  
направления 010501.65 — Прикладная математика и информатика  
факультета КНиИТ  
Коновалова Никиты Алексеевича

Научный руководитель  
доцент, к. ф.-м. н.

\_\_\_\_\_

А. С. Иванова

Заведующий кафедрой  
к.ф.-м.н.

\_\_\_\_\_

А. С. Иванов

Саратов 2013

## Содержание

|  |    |
|--|----|
| ВВЕДЕНИЕ .....   | 3  |
| 1 Пример оформления текста .....                           | 4  |
| 1.1 Пример основных элементов математического текста ..... | 4  |
| 1.2 Еще элементы математического текста .....              | 4  |
| 1.3 Снова математический текст .....                       | 6  |
| 2 Раздел с подразделами .....                              | 8  |
| 2.1 Текст с формулами и леммой .....                       | 8  |
| 2.2 Название другого подраздела .....                      | 9  |
| 2.2.1 Более мелкий подраздел .....                         | 9  |
| 2.2.2 Текст с таблицей .....                               | 9  |
| 2.2.3 Текст с кодом программы .....                        | 10 |
| ЗАКЛЮЧЕНИЕ .....   | 13 |
| СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....                     | 14 |
| ПРИЛОЖЕНИЕ А Нумеруемые объекты в приложении .....         | 15 |
| ПРИЛОЖЕНИЕ Б Листинг программы .....                       | 16 |
| ПРИЛОЖЕНИЕ В Многостраничная таблица .....                 | 18 |

## ВВЕДЕНИЕ

В настоящее время возникает множество задач обработки больших массивов данных. Как правило, стандартные решения для хранения и обработки таких данных неприменимы в силу того, что они рассчитаны на работу внутри одной вычислительной системы. Современные сервера способны обеспечить высокую производительность вычислений, но какова бы она ни была, одной вычислительной системы будет не достаточно для работы с большими данными.

Существуют различные технологии получения, хранения и обработки данных, рассчитанные на работу одновременно нескольких вычислительных систем(серверов). Такие технологии называют распределенными.

Одной из распределенных технологий обработки больших данных является технология MapReduce.

Целью данной работы является построение системы обработки текстовых лог-файлов на основе технологии MapReduce. Задачами можно считать следующие требования к системе:

- система должна иметь возможность обработать сколь угодно большой объем лог-файлов;
- система должна быть масштабируема

## 1 Пример оформления текста

### 1.1 Пример основных элементов математического текста

Внутритекстовая формула  $\frac{1}{\varepsilon^*} = \frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0}$ . Пример одиночной ссылки на литературу [1]. Пример множественной ссылки на литературу [2–8]. Ещё множественная ссылка [9, 10].

$$F(x) = \int_a^b f(x) dx. \quad (1)$$

Ссылка на рисунок 1.

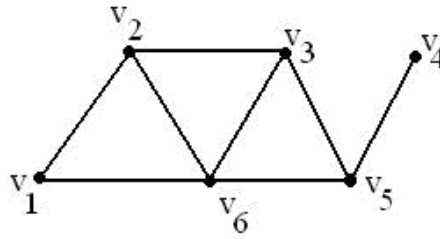


Рисунок 1 – Подпись к рисунку

Если разность энергий электронно-дырочных уровней  $E_2 - E_1$  близка к энергии предельного оптического фонона  $\hbar\Omega_{LO}$ , то в разложении волновых функций полного гамильтониана можно ограничиться нулевым приближением для всех состояний, за исключением близких по значению к  $E_2$ . Волновые функции последних представляют собой следующие комбинации почти вырожденных состояний [8].

### 1.2 Ещё элементы математического текста

Нейрон является составной частью нейронной сети. Он состоит из элементов трех типов: умножителей (синапсов), сумматора и нелинейного преобразователя. Синапсы осуществляют связь между нейронами, умножают входной сигнал на число, характеризующее силу связи (вес синапса). Сумматор выполняет сложение сигналов, поступающих по синаптическим связям от других нейронов, и внешних входных сигналов. Нелинейный преобразователь реализует нелинейную функцию одного аргумента — выхода сумматора. Эта функция называется функцией активации или передаточной функцией. На рисунке 2 приведено строение одного нейрона.

Нейрон в целом реализует скалярную функцию векторного аргумента. Математическая модель нейрона:

$$s = \sum_{i=1}^n w_i x_i + b,$$

$$y = f(s),$$

где  $w_i$  — вес синапса;  $i = 1, \dots, n$ ;  $b$  — значение смещения;  $s$  — результат суммирования;  $x_i$  —  $i$ -тый компонент входного вектора (входной сигнал),  $i = 1, \dots, n$ ;  $y$  — выходной сигнал нейрона;  $n$  — число входов нейрона;  $f(s)$  — нелинейное преобразование (функция активации).

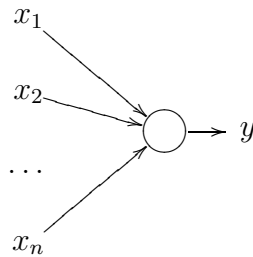


Рисунок 2 – Нейрон

В качестве функции активации нейронов берут обычно одну из следующих:

- пороговая функция активации;
- экспоненциальная сигмоида;
- рациональная сигмоида;
- гиперболический тангенс.

Данные функции активации обладают таким важным свойством как нелинейность. Нелинейность функции активации принципиальна для построения нейронных сетей. Если бы нейроны были линейными элементами, то любая последовательность нейронов также производила бы линейное преобразование и вся нейронная сеть была бы эквивалентна одному нейрону (или одному слою нейронов в случае нескольких выходов). Нелинейность разрушает суперпозицию и приводит к тому, что возможности нейросети существенно выше возможностей отдельных нейронов.

### 1.3 Снова математический текст

Опишем самую популярную архитектуру — многослойный персептрон с последовательными связями и сигмоидальной функцией активации (Feedforward Artificial Neural Network, FANN).

В многослойных нейронных сетях с последовательными связями нейроны делятся на группы с общим входным сигналом — слои. Стандартная сеть состоит из  $L$  слоев, пронумерованных слева направо. Каждый слой содержит совокупность нейронов с едиными входными сигналами. Внешние входные сигналы подаются на входы нейронов входного слоя (его часто нумеруют как нулевой), а выходами сети являются выходные сигналы последнего слоя. Кроме входного и выходного слоев в многослойной нейронной сети есть один или несколько скрытых слоев, соединенных последовательно в прямом направлении и не содержащих связей между элементами внутри слоя и обратных связей между слоями. Число нейронов в слое может быть любым и не зависит от количества нейронов в других слоях. Архитектура нейронной сети прямого распространения сигнала приведена на рисунке 3.

На каждый нейрон первого слоя подаются все элементы внешнего входного сигнала. Все выходы нейронов  $i$ -го слоя подаются на каждый нейрон слоя  $i + 1$ .

Нейроны выполняют взвешенное суммирование элементов входных сигналов. К сумме прибавляется смещение нейрона. Над результатом суммирования выполняется нелинейное преобразование — функция активации (передаточная функция). Значение функции активации есть выход нейрона. Приведем схему многослойного персептрона. Нейроны представлены кружками, связи между нейронами — линиями со стрелками.

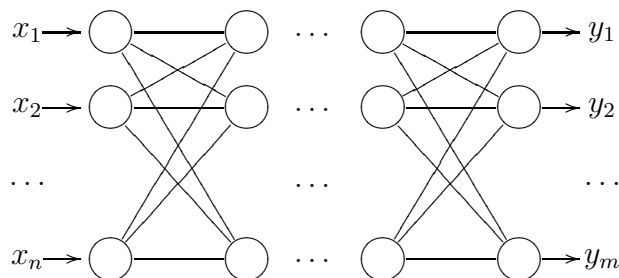


Рисунок 3 – Архитектура многослойной сети прямого распространения

Функционирование сети выполняется в соответствии с формулами:

$$s_j^{[k]} = \sum_{i=1}^{N_{k-1}} w_{ji}^{[k]} y_i^{[k-1]} + b_j^{[k]}, \quad j = 1, \dots, N_k, \quad k = 1, \dots, L;$$

$$y_j^{[k]} = f(s_j^{[k]}), \quad j = 1, \dots, N_k, \quad k = 1, \dots, L - 1,$$

$$y_j^{[L]} = s_j^{[L]},$$

где

- $y_i^{[k-1]}$  — выходной сигнал  $i$ -го нейрона  $(k - 1)$ -го слоя;
- $w_{ji}^{[k]}$  — вес связи между  $j$ -м нейроном слоя  $(k - 1)$  и  $i$ -м нейроном  $k$ -го слоя;
- $b_j^{[k]}$  — значение смещения  $j$ -го нейрона  $k$ -го слоя;
- $y = f(s)$  — функция активации;
- $y_j^{[k]}$  — выходной сигнал  $j$ -го нейрона  $k$ -го слоя;
- $N_k$  — число узлов слоя  $k$ ;
- $L$  — общее число основных слоев;
- $n = N_0$  — размерность входного вектора;
- $m = N_L$  — размерность выходного вектора сети.

На рисунке 4 представлена сеть прямого распространения сигнала с 5 входами, 3 нейронами в скрытом слое и 2 нейронами в выходном слое.

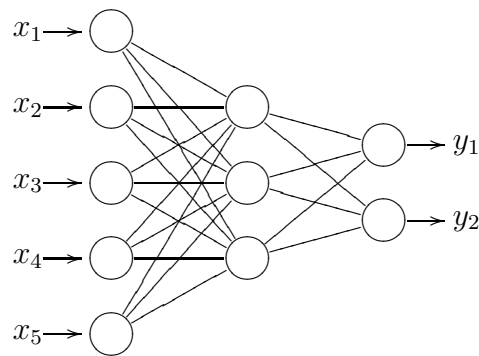


Рисунок 4 – Пример нейронной сети

## 2 Раздел с подразделами

### 2.1 Текст с формулами и леммой

Обозначим  $[y_0, y_1, \dots, y_p; f]$  разделенную разность порядка  $p$  функции  $f$  по узлам  $y_0 < y_1 < \dots < y_p$ .

Обозначим  $L_p f(x; y_0, y_1, \dots, y_p)$  интерполяционный полином Ньютона функции  $f$  по узлам  $y_0, y_1, \dots, y_p$ :

$$L_p f(x; y_0, y_1, \dots, y_p) = \sum_{j=0}^p [y_0, \dots, y_j; f] \cdot \prod_{i=0}^{j-1} (x - y_i), \quad x - y_{-1} \stackrel{\text{def}}{=} 1 \quad (2)$$

Лемма 1. Если  $0 \leq x_0 < x_1 < \dots < x_p \leq 1$  и  $f \in C[0, 1]$  удовлетворяет условиям

1.  $f(x) \geq 0, \quad x \in [0, 1];$
2.  $[y_0, \dots, y_{p+1}; f] \geq 0$  для всех  $y_i \in [0, 1], \quad i = 0, \dots, p+1,$

тогда

$$L_p f(x; x_0, \dots, x_p) \geq 0 \quad (3)$$

для всех  $x \in [x_{p-(2k+1)}, x_{p-2k}], \quad k = 0, \dots, [p/2], \quad x_{-1} \stackrel{\text{def}}{=} -\infty.$

Доказательство. Возьмем  $x \in [x_{p-(2k+1)}, x_{p-2k}], \quad k = 0, \dots, [p/2].$

Из условия 1 леммы следует, что

$$[x_0, \dots, x_{p-(2k+1)}, x, x_{p-2k}, \dots, x_p; f] \geq 0,$$

т. е.

$$\Delta_p f(x; x_0, \dots, x_p) \stackrel{\text{def}}{=} \begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^p & f(x_0) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{p-(2k+1)} & x_{p-(2k+1)}^2 & \dots & x_{p-(2k+1)}^p & f(x_{p-(2k+1)}) \\ \stackrel{\text{def}}{=} 1 & x & x^2 & \dots & x^p & f(x) \\ 1 & x_{p-2k} & x_{p-2k}^2 & \dots & x_{p-2k}^p & f(x_{p-2k}) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_p & x_p^2 & \dots & x_p^p & f(x_p) \end{vmatrix} \geq 0 \quad (4)$$



Из равенства

$$\Delta_p f(x; x_0, \dots, x_p) = (L_p f(x; x_0, \dots, x_p) - f(x)) \prod_{0 \leq i < j \leq p} (x_j - x_i).$$

и (4) следует, что

$$L_p f(x; x_0, \dots, x_p) \geq f(x).$$

С учетом условия 1 леммы мы получаем утверждение (3).  $\square$

## 2.2 Название другого подраздела

### 2.2.1 Более мелкий подраздел

Если разность энергий электронно-дырочных уровней  $E_2 - E_1$  близка к энергии предельного оптического фонона  $\hbar\Omega_{LO}$ , то в разложении волновых функций полного гамильтониана можно ограничиться нулевым приближением для всех состояний, за исключением близких по значению к  $E_2$ .

### 2.2.2 Текст с таблицей

В таблице 1 представлены результаты сокращения словарей неисправностей для схем из каталога ISCAS'89.

Таблица 1 – Результат сокращения словарей неисправностей при помощи масок

| 1     | 2    | 3     | 4        | 5   | 6      | 7      | 8      |
|-------|------|-------|----------|-----|--------|--------|--------|
| S298  | 177  | 1932  | 341964   | 61  | 10797  | 3,16%  | 0,61   |
| S344  | 240  | 1397  | 335280   | 59  | 14160  | 4,22%  | 0,53   |
| S349  | 243  | 1474  | 358182   | 62  | 15066  | 4,21%  | 0,60   |
| S382  | 190  | 12444 | 2364360  | 55  | 10450  | 0,44%  | 3,78   |
| S386  | 274  | 2002  | 548548   | 91  | 24934  | 4,55%  | 1,40   |
| S400  | 194  | 13284 | 2577096  | 58  | 11252  | 0,44%  | 4,28   |
| S444  | 191  | 13440 | 2567040  | 60  | 11460  | 0,45%  | 4,26   |
| S510  | 446  | 700   | 312200   | 70  | 31220  | 10,00% | 0,63   |
| S526  | 138  | 13548 | 1869624  | 38  | 5244   | 0,28%  | 2,41   |
| S641  | 345  | 5016  | 1730520  | 132 | 45540  | 2,63%  | 7,06   |
| S713  | 343  | 3979  | 1364797  | 131 | 44933  | 3,29%  | 5,61   |
| S820  | 712  | 21185 | 15083720 | 244 | 173728 | 1,15%  | 126,99 |
| S832  | 719  | 21603 | 15532557 | 253 | 181907 | 1,17%  | 135,18 |
| S953  | 326  | 322   | 104972   | 91  | 29666  | 28,26% | 0,27   |
| S1423 | 293  | 750   | 219750   | 93  | 27249  | 12,40% | 0,57   |
| S1488 | 1359 | 22230 | 30210570 | 384 | 521856 | 1,73%  | 541,69 |

### 2.2.3 Текст с кодом программы

Термин «разреженная матрица» впервые был предложен Гарри Марковицем. В 1989 он был награжден премией имени Джона фон Неймана в том числе и за вклад в теорию методов для разреженных матриц.

В большинстве источников, разреженной матрицей называется матрица, в которой мало ненулевых элементов. Это нельзя назвать определением из-за слова «мало». В [11] понятие разреженной матрицы определяется так: «Мы можем называть матрицу разреженной, если применение к ней методов, описываемых в книге, экономит память и/или время». Таким образом, следует дать определение алгоритму для разреженных матриц. Алгоритмом для разреженных матриц будем называть алгоритм, у которого время работы и необходимый объем памяти зависят от количества ненулевых элементов в матрице.

Размерность квадратной матрицы  $A$  будем обозначать  $n$ , а количество ненулевых элементов в ней  $|A|$ .

Плотные матрицы обычно хранятся в качестве двумерного массива  $n \times n$ . Будем обозначать такой массив  $a$ . Разреженные матрицы не стоит хранить таким способом из-за слишком большого потребления памяти, которая будет занята в основном нулевыми элементами.

Один из вариантов представления разреженных матриц в памяти компьютера — в виде трех массивов: `column`, `value` и `rowIndex`. Размеры массивов `column` и `value` равны  $|A|$ . Размер `rowIndex` равен  $n + 1$ . Ненулевые элементы матрицы  $A$  хранятся последовательно по строкам в этих массивах. Элемент `column[i]` содержит номер столбца, в котором содержится  $i$ -й ненулевой элемент, а `value[i]` — его величину. Массив `rowIndex[i]` содержит в себе индекс первого ненулевого элемента  $i$ -й строки. Все ненулевые элементы  $i$ -й строки содержатся в массивах `column` и `value` в элементах с индексами от `rowIndex[i]` по `rowIndex[i + 1] - 1`. Для удобства полагают `rowIndex[n] = |A|`.

Для примера рассмотрим следующую матрицу:

$$\begin{pmatrix} 1 & 0 & 5 & 0 & 0 \\ 0 & 2 & 7 & 4 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 9 & 6 & 0 & 3 & 0 \\ 0 & 0 & 3 & 0 & 5 \end{pmatrix}$$

Массивы `column`, `value` и `rowIndex` для этой матрицы представлены в таблице 2.

Таблица 2 – Массивы `column`, `value` и `rowIndex`

|                       | 0 | 1 | 2 | 3 | 4 | 5  | 6 | 7 | 8 | 9 | 10 | 11 |
|-----------------------|---|---|---|---|---|----|---|---|---|---|----|----|
| <code>column</code>   | 0 | 2 | 1 | 2 | 3 | 2  | 0 | 1 | 3 | 2 | 4  |    |
| <code>value</code>    | 1 | 5 | 2 | 7 | 4 | 1  | 9 | 6 | 3 | 3 | 5  |    |
| <code>rowIndex</code> | 0 | 2 | 5 | 6 | 9 | 11 |   |   |   |   |    |    |

Неизвестный вектор и вектор правой части хранятся в виде массивов размера  $n$ . Массив неизвестного вектора обозначают `x`, а массив правой части — `rhs`.

Рассмотрим пример алгоритма для разреженных матриц. Алгоритм решения СЛАУ, представленной нижнетреугольной матрицей `a`, можно реализовать двумя вложенными циклами по `n`:

```

1 for(int i = 0; i < n; ++i){
2     x[i] = rhs[i];
3     for(int j = 0; j < i; ++j)
4         x[i] -= a[i][j] * x[j];
5     x[i] /= a[i][i];
6 }
```

Но, если матрица `a` хранится в разреженном виде, то в данном алгоритме можно проходить только по ненулевым элементам `a`:

```

1 for(int i = 0; i < n; ++i){
2     x[i] = rhs[i];
3     for(int j = rowIndex[i]; j < rowIndex[i + 1] - 1; ++j)
4         x[i] -= value[j] * x[column[j]];
5     x[i] /= value[rowIndex[i + 1] - 1];
6 }
```

В первом случае оценка времени работы будет  $O(n^2)$ , а во втором  $O(|A|)$ .

Методы для разреженных матриц основаны на следующих главных принципах [11]:

1. Хранятся только ненулевые элементы матрицы.
2. Выполняются только те преобразования, которые действительно что-то изменяют. В примере не имеет смысла вычитать из  $x[i]$  значение  $x[j] * a[i][j]$ , если  $a[i][j]$  равно нулю.
3. Число «новых элементов», возникающих, например, во время исключения Гаусса, стараются уменьшить путем перестановок строк и столбцов матрицы.

## ЗАКЛЮЧЕНИЕ

В настоящей работы приведен пример оформления студенческой работы средствами системы L<sup>A</sup>T<sub>E</sub>X.

Показано, как можно оформить документ в соответствии:

- с правилами оформления курсовых и выпускных квалификационных работ, принятых в Саратовском государственном университете в 2012 году;
- с правилами оформления титульного листа отчета о прохождении практики в соответствии со стандартом.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Yoffe, A. D. Low-dimensional systems: quantum size effects and electronic properties of semiconductor microcrystallites (zero-dimensional systems) and some quasi-two-dimensional systems / A. D. Yoffe // Adv. Phys. — 1993. — Vol. 42. — Pp. 173–266.
- 2 Эфрос, Ал. . Межзонное поглощение света в полупроводниковом шаре / Ал. . Эфрос, . . Эфрос // Физика и техника полупроводников. — 1982. — Т. 16, № 7. — С. 1209–1214.
- 3 Ансельм, . . Введение в теорию полупроводников / . . Ансельм. — Москва: Наука, 1978.
- 4 Segall, B. // Proceedings of IXth Conference on the Physics of Semiconductors, Moscow, 1968 / Ed. by S. M. Ryvkin. — Leningrad: Nauka, 1968. — P. 425.
- 5 Spectroscopy and Excitation Dynamics of Condensed Molecular Systems / Ed. by V. M. Agranovich, R. M. Hochstrasser. — Modern Problems in Condensed Matter Sciences. Amsterdam: North-Holland, 1983.
- 6 Inp basic parameters at 300 k // Electronic archive New Semiconductor Materials. Characteristics and Properties / Ioffe Physico-Technical Institute. — St. Petersburg, 2001. <http://www.ioffe.rssi.ru/SVA/NSM/Semicond/InP/basic.html>.
- 7 Мищенко, . . Неупругое рассеяние света в системе взаимодействующих электронов и фононов: Ph.D. thesis / ИТФ им. Л. Д. Ландау. — 1996.
- 8 Скворцов, . . Флуктуационные и интерференционные эффекты в мезоскопических системах. — 2008.
- 9 Perelman, G. Finite extinction time for the solutions to the ricci flow on certain three-manifolds / G. Perelman.
- 10 Nielsen, E. A configuration interaction analysis of exchange in double quantum dots / E. Nielsen, R. P. Muller.
- 11 Эстербю, . Прямые методы для разреженных матриц / . Эстербю, . Златев. — М.: Мир.

# ПРИЛОЖЕНИЕ А

## Нумеруемые объекты в приложении

Таблица А.1 – Results of pass-fail dictionary reduction with the help of masks

| Circuit | Number of modelled faults | Number of test vectors in the test set | The volume of pass-fail dictionary, bit | The volume of found mask | The volume of masked dictionary, bit | % of pass-fail dictionary | CPU running time, min |
|---------|---------------------------|--|---|--------------------------|--------------------------------------|---------------------------|-----------------------|
| S298    | 177                       | 322                                    | 56994                                   | 30                       | 5310                                 | 9,32%                     | 0,07                  |
| S344    | 240                       | 127                                    | 30480                                   | 29                       | 6960                                 | 22,83%                    | 0,04                  |
| S349    | 243                       | 134                                    | 32562                                   | 35                       | 8505                                 | 26,12%                    | 0,05                  |
| S382    | 190                       | 2074                                   | 394060                                  | 28                       | 5320                                 | 1,35%                     | 0,43                  |
| S386    | 274                       | 286                                    | 78364                                   | 65                       | 17810                                | 22,73%                    | 0,26                  |
| S400    | 194                       | 2214                                   | 429516                                  | 32                       | 6208                                 | 1,45%                     | 0,99                  |
| S444    | 191                       | 2240                                   | 427840                                  | 30                       | 5730                                 | 1,34%                     | 0,98                  |
| S526    | 138                       | 2258                                   | 311604                                  | 28                       | 3864                                 | 1,24%                     | 0,61                  |
| S641    | 345                       | 209                                    | 72105                                   | 58                       | 20010                                | 27,75%                    | 0,24                  |
| S713    | 343                       | 173                                    | 59339                                   | 58                       | 19894                                | 33,53%                    | 0,19                  |
| S820    | 712                       | 1115                                   | 793880                                  | 147                      | 104664                               | 13,18%                    | 9,09                  |
| S832    | 719                       | 1137                                   | 817503                                  | 151                      | 108569                               | 13,28%                    | 9,20                  |
| S953    | 326                       | 14                                     | 4564                                    | 13                       | 4238                                 | 92,86%                    | 0,01                  |
| S1423   | 293                       | 150                                    | 43950                                   | 58                       | 16994                                | 38,67%                    | 0,15                  |
| S1488   | 1359                      | 1170                                   | 1590030                                 | 158                      | 214722                               | 13,50%                    | 26,69                 |

$$F(x) = \int_a^b f(x) dx. \quad (\text{A.1})$$

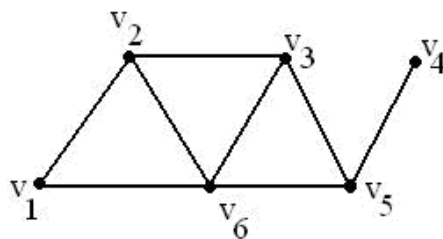


Рисунок А.1 – Подпись к рисунку

Таблица А.2

|   |   |
|---|---|
| 0 | 1 |
| 1 | 0 |

## ПРИЛОЖЕНИЕ Б

### Листинг программы

Код приложения task.pl.

```
1 use locale;
2 use encoding "cp866";
3 {
4     print "Имя папки: "; my $folder_name=<>;
5     chomp($folder_name);
6     my @files = 'chcp 866 & attrib $folder_name\\*.pl';
7     if (substr($files[1],0,15) eq 'Не найден путь:') {
8         print "Путь не найден. Попробуйте еще.\n";
9         redo;
10    }
11    elsif (substr($files[1],0,15) eq 'Не найден файл:') {
12        print "Папка не содержит файлов .txt .\n";
13        last;
14    }
15    else {
16        foreach my $file (@files[1 .. $#files]){
17            my $file_name = substr($file, 11);
18            chomp($file_name);
19            open(FH,"<$file_name") or die $!;
20            my %hash = ();
21            foreach $chunk (<FH>){
22                my @words = $chunk =~ /([\@%\$][a-zA-Z_0-9]+[\[\{\}])/g;
23                foreach my $word (@words) {
24                    $word = "\".$word, 1)
25                    if (substr($word, 0, 1) eq '@' &&
26                        substr($word, -1) eq '[');
27                    $word = "\".$word, 1). "["
28                    if (substr($word, 0, 1) eq '@');
29                    $word = "\".$word, 1). "{"
30                    if (substr($word, 0, 1) eq '%');
31                    $hash{$word}++;
32                };
33            };
34            my @xs = keys %hash;
35            print @xs;
36            close(FH);
```



```
37         my $ans = scalar(@xs);
38         print "$file_name : $ans\n";
39     }
40 }
41 }
```

## ПРИЛОЖЕНИЕ В

### Многостраничная таблица

Таблица В.1 – ГОСТ DIN ISO — Таблица соответствия стандартов

| Стандарт<br>ГОСТ | Наименование   | Стандарт<br>DIN      | Стандарт<br>ISO        |
|------------------|--|----------------------|------------------------|
| 1                | 2  | 3                    | 4                      |
| ГОСТ<br>397-79   | Шплинты  | DIN 94               | ISO 1234               |
| ГОСТ<br>1144-80  | Шурупы с полукруглой<br>головкой   | DIN<br>96DIN<br>7981 | ISO 7049               |
| ГОСТ<br>1145-80  | Шурупы с потайной головкой   | DIN<br>97DIN<br>7982 | ISO 7050               |
| ГОСТ<br>1146-80  | Шурупы с полупотайной<br>головкой  | DIN<br>95DIN<br>7983 | ISO 7051               |
| ГОСТ<br>1476-93  | Винты установочные с<br>коническим концом и прямым<br>шлицем классов точности А и<br>В             | DIN 553              | ISO 7434               |
| ГОСТ<br>1477-93  | Винты установочные с плоским<br>концом и прямым шлицем<br>классов точности А и В                   | DIN<br>438DIN<br>551 | ISO<br>4766ISO<br>7436 |
| ГОСТ<br>1478-93  | Винты установочные с<br>цилиндрическим концом и<br>прямым шлицем классов<br>точности А и В         | DIN 417              | ISO 7435               |
| ГОСТ<br>1481-84  | Винты установочные с<br>шестигранной головкой и<br>цилиндрическим концом<br>классов точности А и В | DIN 561              |                        |
| ГОСТ<br>1482-84  | Винты установочные с<br>квадратной головкой и<br>цилиндрическим концом<br>классов точности А и В   | DIN 479              |                        |

Продолжение таблицы В.1

| 1               | 2   | 3                    | 4  |
|-----------------|---|----------------------|--|
| ГОСТ<br>1485-84 | Винты установочные с<br>квадратной головкой и<br>засверленным концом классов<br>точности А и В          | DIN 479              |  |
| ГОСТ<br>1486-84 | Винты установочные с<br>квадратной головкой и<br>ступенчатым концом со сферой<br>классов точности А и В | DIN 480              |  |
| ГОСТ<br>1488-84 | Винты установочные с<br>квадратной головкой и<br>буртиком классов точности А и<br>В                     | DIN 478              |  |
| ГОСТ<br>1491-80 | Винты с цилиндрической<br>головкой классов точности А и<br>В  | DIN 84               | ISO 1207                                     |
| ГОСТ<br>3032-76 | Гайки-барашки   | DIN 315              |  |
| ГОСТ<br>3033-79 | Болты откидные  | DIN 444              |  |
| ГОСТ<br>3057-90 | Пружины тарельчатые   | DIN 2093             |  |
| ГОСТ<br>3070-88 | Канат стальной двойной<br>свивки типа ТК конструкции<br>6х19 (1+6+12)+1 о.с.                            | DIN 3060             |  |
| ГОСТ<br>3128-70 | Штифты цилиндрические<br>незакаленные   | DIN 7DIN<br>6325     | ISO<br>2338ISO<br>8734                       |
| ГОСТ<br>3129-70 | Штифты конические<br>незакаленные   | DIN 1                | ISO 2339                                     |
| ГОСТ<br>4751-73 | Рым-болты   | DIN 580              | ISO 3266                                     |
| ГОСТ<br>5915-70 | Гайки шестигранные стальные<br>класса точности В  | DIN<br>555DIN<br>934 | ISO<br>4032ISO<br>4033ISO<br>8673ISO<br>8674 |

Продолжение таблицы В.1

| 1               | 2   | 3                     | 4   |
|-----------------|---|-----------------------|---|
| ГОСТ<br>5916-70 | Гайки шестигранные низкие<br>класса точности В  | DIN<br>439DIN<br>936  | ISO<br>4035ISO<br>4036ISO<br>8675             |
| ГОСТ<br>5918-73 | Гайки шестигранные<br>прорезные и корончатые<br>класса точности В                               | DIN 935               | EN ISO<br>7035EN<br>ISO<br>7036EN<br>ISO 7037 |
| ГОСТ<br>5919-73 | Гайки шестигранные<br>прорезные и корончатые<br>низкие класса точности В                        | DIN 937               | EN ISO<br>7038                                |
| ГОСТ<br>5927-70 | Гайки шестигранные класса<br>точности А   | DIN<br>555DIN<br>934  | ISO<br>4032ISO<br>4034ISO<br>8673             |
| ГОСТ<br>5932-73 | Гайки шестигранные<br>прорезные и корончатые<br>класса точности А                               | DIN<br>935DIN<br>937  | EN ISO<br>7035EN<br>ISO<br>7036EN<br>ISO 7037 |
| ГОСТ<br>6393-73 | Гайки круглые с отверстиями<br>на торце "под ключ" класса<br>точности А                         | DIN 1816              |   |
| ГОСТ<br>6402-70 | Шайбы пружинные   | DIN 127               |   |
| ГОСТ<br>6958-78 | Шайбы увеличенные. Классы<br>точности А и С   | DIN<br>440DIN<br>9021 | ISO<br>7094ISO<br>7093-1ISO<br>7093-2         |
| ГОСТ<br>7786-81 | Болты с потайной головкой и<br>квадратным подголовком<br>класса точности С                      | DIN 608               |   |
| ГОСТ<br>7795-70 | Болты с шестигранной<br>уменьшенной головкой и<br>направляющим подголовком,<br>класс точности В |                       |   |