

РАЗРАБОТКА АЛГОРИТМА ДЛЯ ОБРАБОТКИ ДАННЫХ ИЗ ТЕКСТОВОГО ФАЙЛА

Бабаш А. В., Квашнин В. О.

Рассмотрены проблемы обработки больших массивов данных, содержащихся в строках текстовых файлов. Разработан алгоритм получения массивов данных из строк для их дальнейшей обработки. Приведено графическое описание алгоритма обработки строковых данных. Приведен интерфейс разработанного Windows приложения, которое позволяет получить массивы данных из текстовых файлов. Описаны основные возможности разработанного приложения. Представлен типовой IMDS текстовый файл с данными в виде строк. Приведен алгоритм анализатора строк (парсинга). Рассмотрены ограничения стандартных прикладных пакетов программ, таких как Matlab, Excel, Mathcad, при обработке больших массивов данных.

Розглянуто проблеми обробки великих масивів даних, що містяться в рядках текстових файлів. Розроблено алгоритм отримання масивів даних з рядків для їх подальшої обробки. Наведено графічний опис алгоритму обробки строкових даних. Наведено інтерфейс розробленого для Windows додатку, який дозволяє отримати масиви даних з текстових файлів. Описано основні можливості розробленого додатка. Представлений типовий IMDS текстовий файл з даними у вигляді рядків. Наведено алгоритм аналізатора рядків (парсинга). Розглянуто обмеження стандартних прикладних пакетів програм, таких як Matlab, Excel, Mathcad, при обробці великих масивів даних.

The problems of processing data bulks contained in lines of text files are considered here. The algorithm of data acquisition from lines for their further processing is developed. The graphic description of an algorithm of line data processing is given here. The interface of the developed Windows application which allows to receive data arrays from text files is represented in the article. The main features of the developed application are presented here. Typical IMDS text file with data in the form of lines is given. The algorithm of the analyzer of lines (parsing) is also demonstrated. Restrictions of standard applied software packages such as Matlab, Excel, Mathcad, when processing large data arrays are considered in this article.

Бабаш А. В.

аспирант ДГМА

Квашнин В. О.

канд. техн. наук, доц. каф. ЭСА ДГМА
tm@dgma.donetsk.ua

ДГМА – Донбасская государственная машиностроительная академия, г. Краматорск.

УДК 621.313

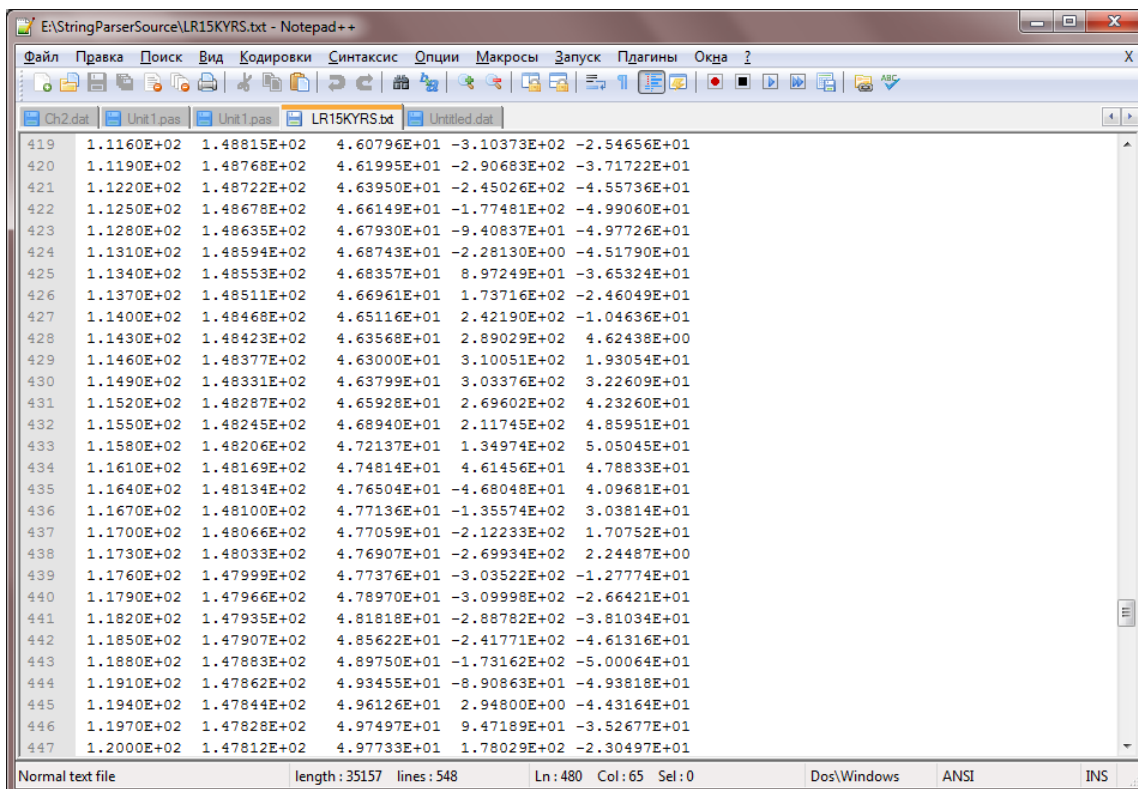
Бабаш А. В., Квашнин В. О.

РАЗРАБОТКА АЛГОРИТМА ДЛЯ ОБРАБОТКИ ДАННЫХ ИЗ ТЕКСТОВОГО ФАЙЛА

При обработке экспериментальных данных зачастую возникает проблема обработки данных, полученных при помощи программного обеспечения L-Card и различных других программ (пакет моделирования IMDS) [1]. Эти данные представлены в виде большого файла массива строк. Различные исследуемые параметры представлены в строке через пробелы или другие разделители.

Данные в таком представлении очень трудно, а зачастую и невозможно обрабатывать при помощи стандартных программных продуктов, таких как Microsoft Excel и даже MathCad или MathLab [2,3]. Простое ручное копирование отдельных элементов массива из исходного файла в другие среды для дальнейшей обработки будет неэффективным, так как приходится работать с большим числом массивов данных. Для эффективной работы с массивами данных нужно использовать синтаксический анализатор строк, который можно реализовать в виде программного продукта. Синтаксический анализатор строк или парсер извлекает из строки значения параметров экспериментальных данных и позволяет их сохранять и представлять в удобном для их дальнейшей обработки виде.

Целью данной работы является разработка алгоритма для обработки больших массивов данных из текстового файла при помощи анализа строк (парсинга). Текстовый файл, который содержит большой набор данных приведен на рис. 1.



```
419 1.1160E+02 1.48815E+02 4.60799E+01 -3.10373E+02 -2.54656E+01
420 1.1190E+02 1.48768E+02 4.61995E+01 -2.90683E+02 -3.71722E+01
421 1.1220E+02 1.48722E+02 4.63950E+01 -2.45026E+02 -4.55736E+01
422 1.1250E+02 1.48678E+02 4.66149E+01 -1.77481E+02 -4.99060E+01
423 1.1280E+02 1.48635E+02 4.67930E+01 -9.40837E+01 -4.97726E+01
424 1.1310E+02 1.48594E+02 4.68743E+01 -2.28130E+00 -4.51790E+01
425 1.1340E+02 1.48553E+02 4.68357E+01 8.97249E+01 -3.65324E+01
426 1.1370E+02 1.48511E+02 4.66961E+01 1.73716E+02 -2.46049E+01
427 1.1400E+02 1.48468E+02 4.65116E+01 2.42190E+02 -1.04636E+01
428 1.1430E+02 1.48423E+02 4.63568E+01 2.89029E+02 4.62438E+00
429 1.1460E+02 1.48377E+02 4.63000E+01 3.10051E+02 1.93054E+01
430 1.1490E+02 1.48331E+02 4.63799E+01 3.03376E+02 3.22609E+01
431 1.1520E+02 1.48287E+02 4.65928E+01 2.69602E+02 4.23260E+01
432 1.1550E+02 1.48245E+02 4.68940E+01 2.11745E+02 4.85951E+01
433 1.1580E+02 1.48206E+02 4.72137E+01 1.34974E+02 5.05045E+01
434 1.1610E+02 1.48169E+02 4.74814E+01 4.61456E+01 4.78833E+01
435 1.1640E+02 1.48134E+02 4.76504E+01 -4.68048E+01 4.09681E+01
436 1.1670E+02 1.48100E+02 4.77136E+01 -1.35574E+02 3.03814E+01
437 1.1700E+02 1.48066E+02 4.77059E+01 -2.12233E+02 1.70752E+01
438 1.1730E+02 1.48033E+02 4.76907E+01 -2.69934E+02 2.24487E+00
439 1.1760E+02 1.47999E+02 4.77376E+01 -3.03522E+02 -1.27774E+01
440 1.1790E+02 1.47966E+02 4.78970E+01 -3.09998E+02 -2.66421E+01
441 1.1820E+02 1.47935E+02 4.81818E+01 -2.88782E+02 -3.81034E+01
442 1.1850E+02 1.47907E+02 4.85622E+01 -2.41771E+02 -4.61316E+01
443 1.1880E+02 1.47883E+02 4.89750E+01 -1.73162E+02 -5.00064E+01
444 1.1910E+02 1.47862E+02 4.93455E+01 -8.90863E+01 -4.93818E+01
445 1.1940E+02 1.47844E+02 4.96126E+01 2.94800E+00 -4.43164E+01
446 1.1970E+02 1.47828E+02 4.97497E+01 9.47189E+01 -3.52677E+01
447 1.2000E+02 1.47812E+02 4.97733E+01 1.78029E+02 -2.30497E+01
```

Рис. 1. Текстовый файл с данными (REP файл IMDS)

Алгоритм обработки данных при помощи парсинга строк и дальнейшего извлечения и обработки данных приведен на рис. 2, 3, 4.

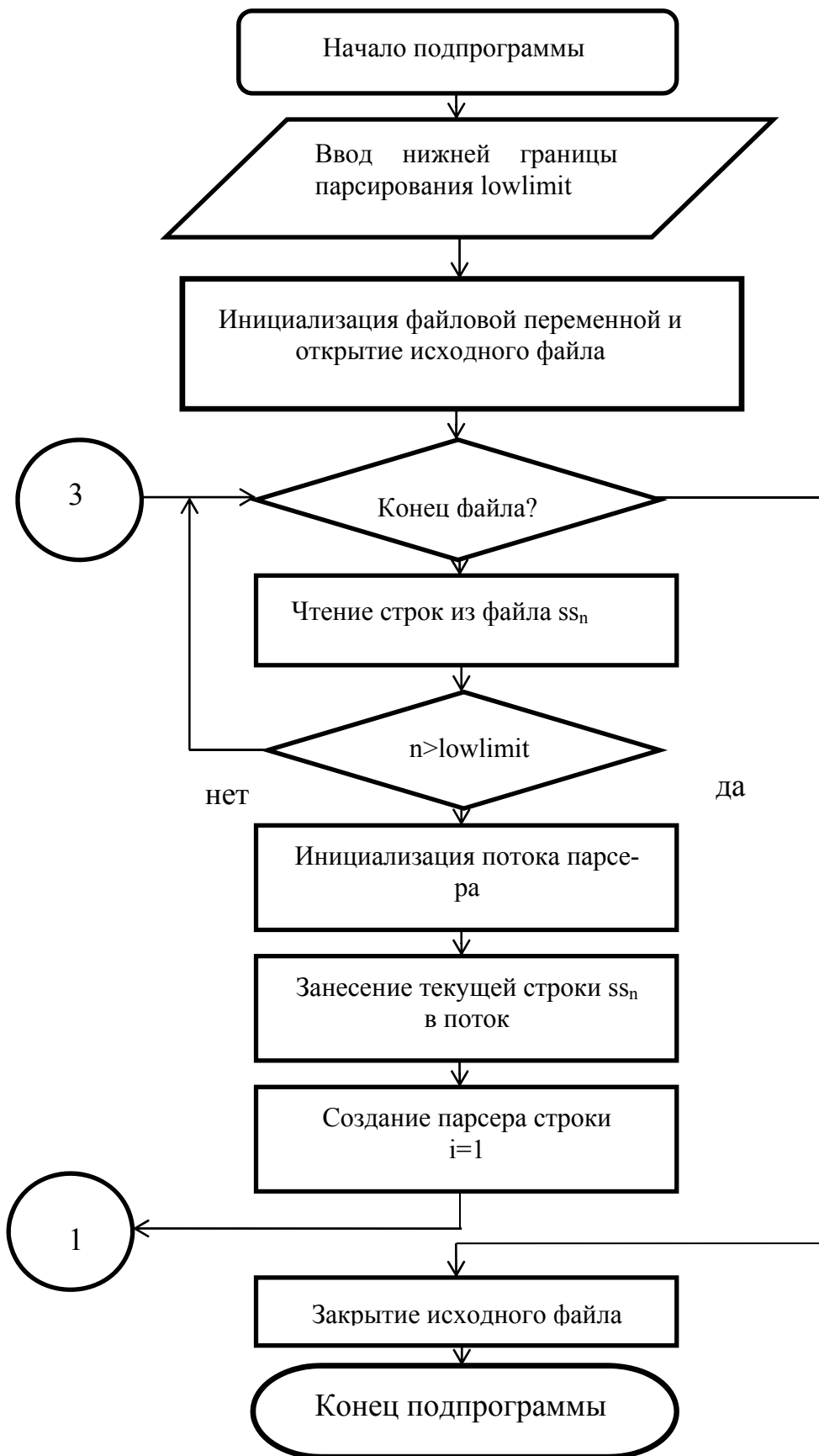


Рис. 2. Алгоритм обработки данных из текстового файла (часть 1)

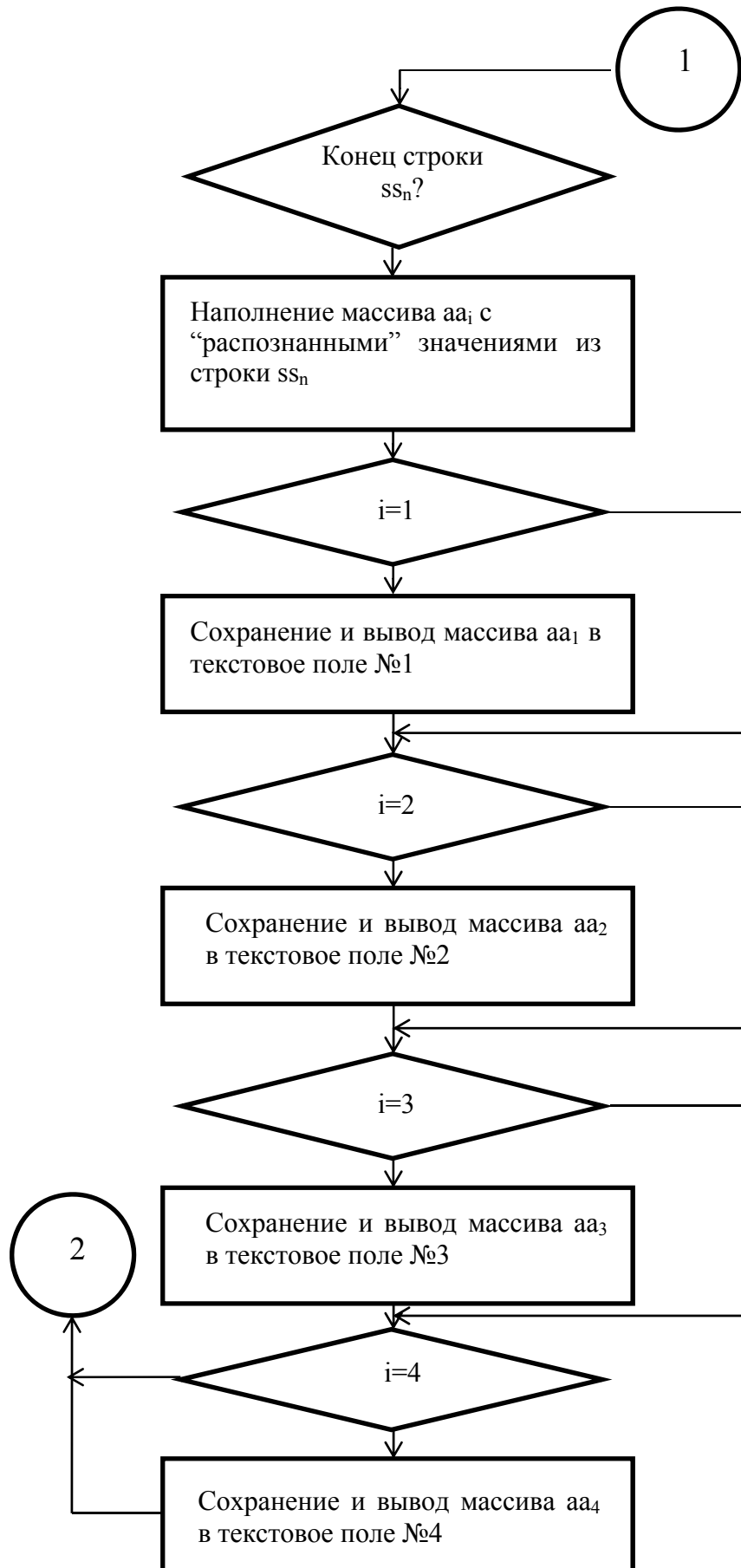


Рис. 3. Алгоритм обработки данных из текстового файла (часть 2)

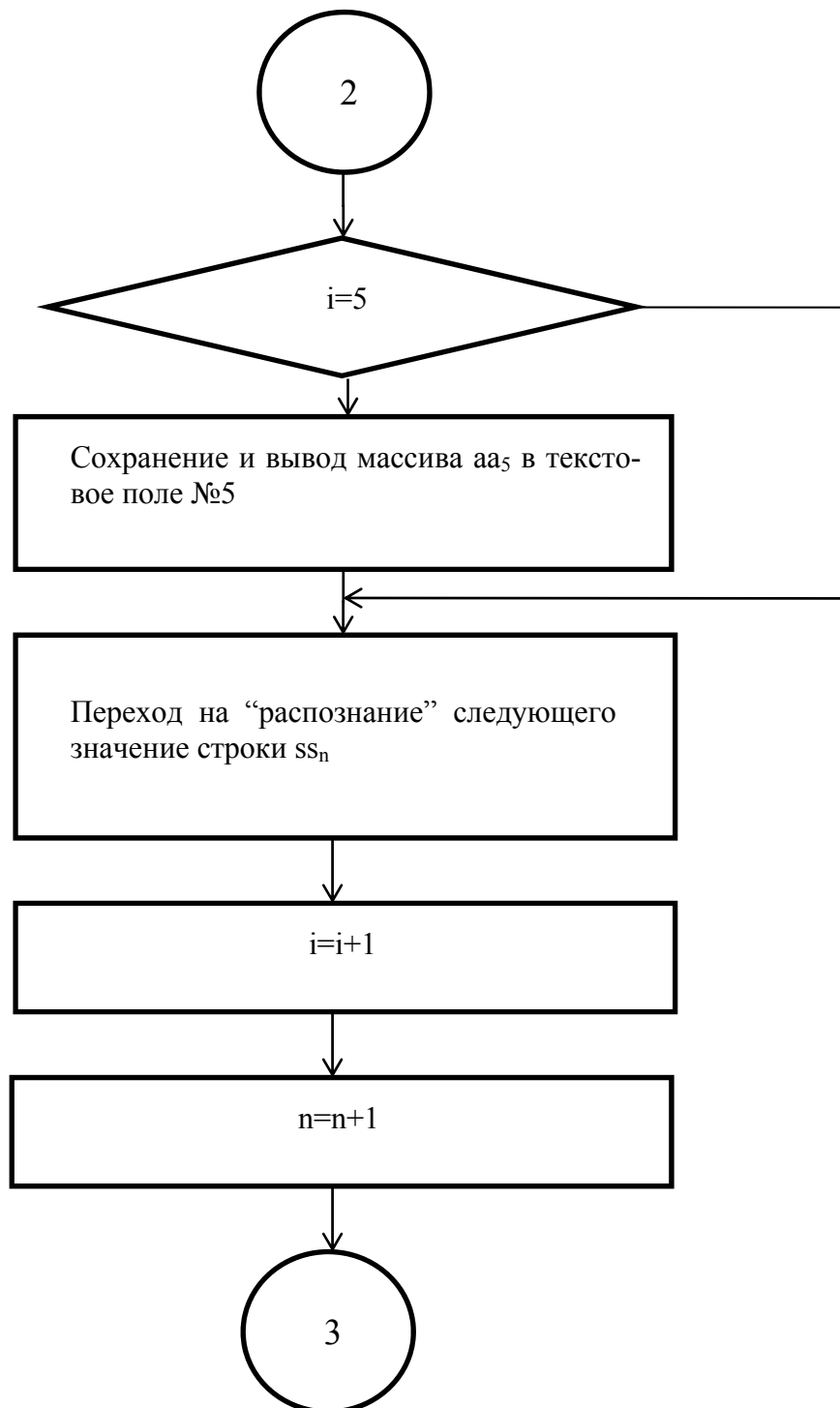


Рис. 4. Алгоритм обработки данных из текстового файла (часть 3)

Данный алгоритм может быть реализован при помощи любого языка программирования высокого уровня [4, 5]. Для демонстрации обработки данных из файла и работоспособности разработанного алгоритма, было разработано Windows приложение в среде IDE Delphi [6]. Его интерфейс и результаты обработки представлены на рис. 5.

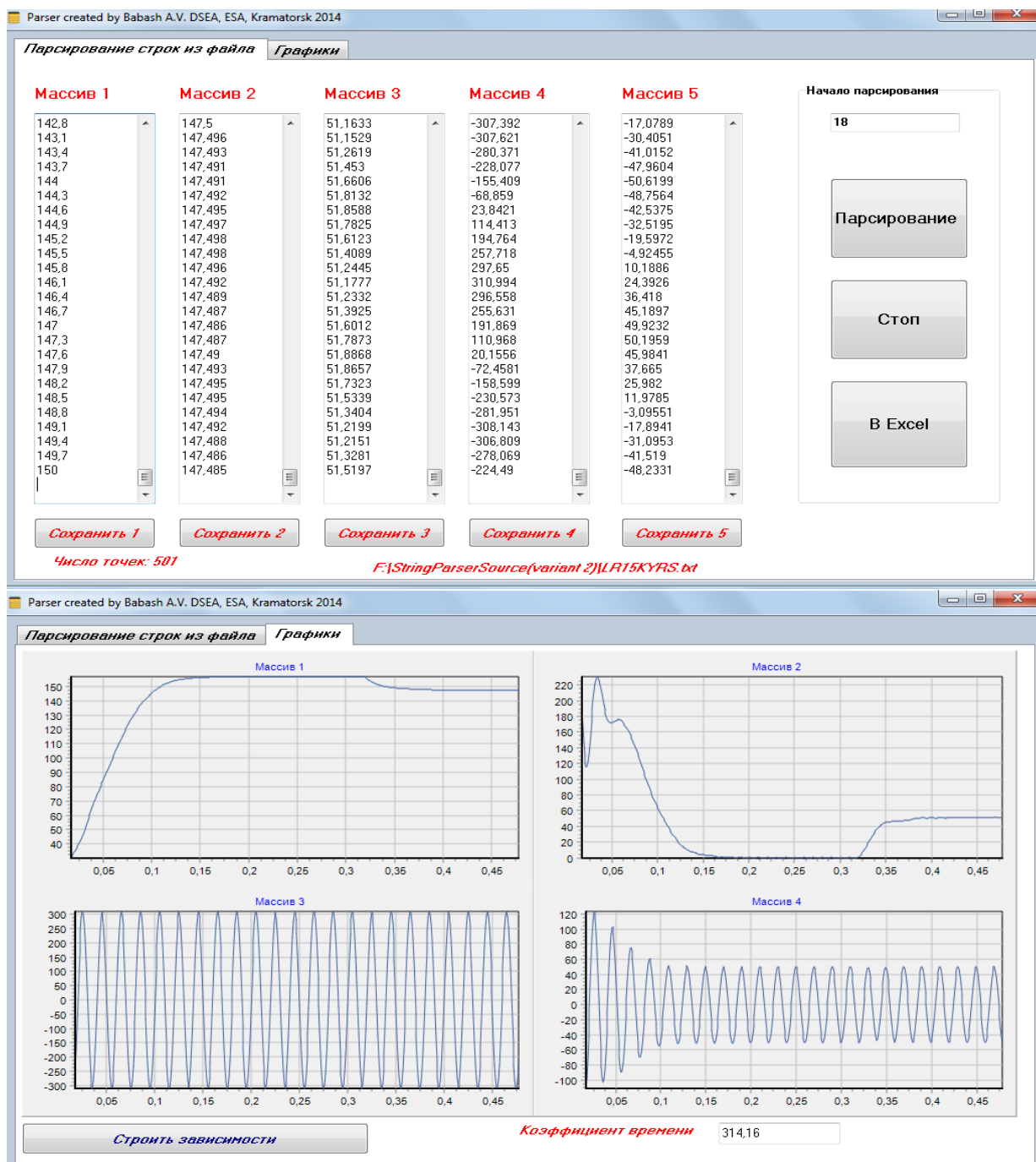


Рис. 5. Интерфейс приложения для обработки и визуализации данных при помощи парсинга строк

Разработанное приложение позволяет конвертировать массивы данных в удобочитаемый формат Excel файла с расширением .xls. Оно содержит пять полей для полученных массивов данных при помощи парсинга строк. Для запуска процесса парсинга необходимо нажать клавишу «Парсирование». Далее нужно выбрать исходный файл, данные из которого необходимо получить. Для определения границ обработки строк из файла нужно установить номер начальной строки в поле «Начало парсирования». Когда процесс анализа строк закончится, данные из пяти полей можно сохранить в excel-файле нажав на кнопку «В Excel». Получится удобный файл для дальнейшей обработки (рис. 6).

	A	B	C	D	E	F	G	H	I
1	0	0	0	0	0				
2	0,3	0,001123	0,588377	88,9311	12,5165				
3	0,6	0,028616	7,10838	173,028	40,2425				
4	0,9	0,164522	25,677	241,669	72,2514				
5	1,2	0,524068	57,7309	288,722	101,824				
6	1,5	1,21556	100,678	309,985	124,528				
7	1,8	2,31345	149,71	303,558	137,486				
8	2,1	3,84602	199,305	270,014	139,145				
9	2,3999	5,79431	244,297	212,351	129,211				
10	2,6999	8,09877	280,611	135,72	108,575				
11	2,9999	10,6708	305,668	46,9649	79,1742				
12	3,2999	13,4065	318,478	-45,9854	43,7648				
13	3,5999	16,2001	319,485	-134,828	5,62631				
14	3,8999	18,9548	310,214	-211,626	-31,7719				
15	4,1999	21,5902	292,858	-269,521	-65,102				
16	4,4999	24,0469	269,882	-303,341	-91,4748				
17	4,7999	26,2869	243,718	-310,063	-108,679				
18	5,0999	28,2926	216,572	-289,089	-115,354				

Рис. 6. Массив данных в формате Excel

ВЫВОДЫ

Таким образом, разработанный алгоритм позволяет извлечь массивы данных из строк различных текстовых файлов посредством их парсинга. Он позволяет существенно упростить выборку и распознавание и дальнейшую обработку больших массивов данных, которые представлены в виде строк исходного текстового файла. Полученный алгоритм может быть реализован с использованием любого языка программирования высокого уровня, например C#, Pascal и т.д.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Яблунь В.П. Прикладні паети машинного аналізу й синтезу систем електроприводу: методичні вказівки / В.П. Яблунь. – Алчевськ: ДГМІ, 2003. – 45 с.
2. Зеленов А.Б. Синтез та цифрове моделювання систем управління електроприводів постійного струму з вентильними перетворювачами: навч. посібник для студентів вузів/ А.Б. Зеленов, І.С. Шевченко, Н.І. Андрєєва. – Алчевськ: ДГМІ, 2002. – 400 с. – ISBN 966-7560-62-7
3. Копилов І.П. Математичне моделювання електричних машин / І.П. Копилов. – М.:Вища школа, 2001. –327 с. – ISBN 5-06-003861-0
4. Кальтин Н. Б. Delphi в задачах и примерах. – 2-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2008. – 288 с.: ил.
5. Дарахвелидае П. Г. Программирование в Delphi 7 / П. Г. Дарахвелидае, Е. П. Марков. – СПб.: БХВ-Петербург, 2003. – 784 с.: ил. – ISBN 5-94157-116-X
6. Хомоненко А.Д. Delphi 7 / Под общ. ред. А.Д. Хомоненко. – СПб.: БХВ-Петербург, 2008. – 1216 с.: ил.

Статья поступила в редакцию 25.05.2016 г.