

**Московский государственный технический университет им. Н.Э. Баумана  
кафедра "Системы обработки информации и управления"**

**ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ИССЛЕДОВАНИЯ ЯЗЫКА  
ПЕЧАТНЫХ ИСТОЧНИКОВ XVIII – НАЧ. XIX ВВ.**

Расчетно-пояснительная записка  
курсовой работы по дисциплине  
"Семиотика информационных технологий"  
студент группы ИУ5-92  
**Григорьева Светлана Сергеевна.**

Шифр:92\_05

Преподаватель: к.т.н., доц. А.Ю.Филиппович

**Москва, 2010 г.**

## **Введение**

### **Цель курсовой работы**

Целью предлагаемых учебно-практических занятий является изучение современных информационных технологий и инновационных разработок для сохранения исторических и культурных ценностей России на примере задачи исследования языка печатных источников XVIII – нач. XIX вв. Выполнение заданий позволяет на практике изучить особенности рассматриваемых текстов, проанализировать факторы, влияющие на эффективность их распознавания с помощью современных OCR-систем.

### **Задачи курсовой работы**

1. Подготовка материалов для выполнения задания: установка ПО, анализ и фиксация параметров ПО. Анализ характеристик источника.
2. Ввод и распознавание текстового фрагмента. Предварительная оценка эффективности работы OCR-системы.
3. Формирование шрифтовых эталонов фрагментов, используя технологию обучения. Оценка эффективности использования технологии распознавания, включающей шрифтовые эталоны.
4. Оценка эффективности использования технологии распознавания, включающей дополнительный словарь языка текста.
5. Корректурa текста и анализ лексики и типов ошибок.
6. Квантитативные исследования текста, формирование словарей фрагментов, построение функций распределения частот. Выводы. Формирование отчета.

## Основная часть

### Задача 1

#### **Основные настройки программного и аппаратного обеспечения:**

Windows 7 64-bit

ABBYYFineReader, Выпуск 10.0.102.95, 2009г.

#### **Программное обеспечение:**

Так как текст был дан в формате pdf, то сканер не понадобился. Поэтому приведем только характеристики компьютера.

Процессор -IntelCorei3

Частота - 240Гц

ОП - 3 Гб

#### **Характеристики изображений страниц**

Многие страницы сохранились хорошо, на отсканированных страницах нет желтизны, большого количества лишних пятен. Только на некоторых изображениях присутствуют темные куски текста, образовавшиеся видимо из-за пометок в тексте или возможного сдвига бумаги. Но в целом строки в тексте идут очень ровно, что способствует хорошему распознаванию текста.

#### **Библиографические данные источника:**

Краткое физическое и топографическое описание Таврической области, сочиненное на Французском языке ПетромЪПалласомЪ, СтатскимЪСоветникомЪ, АкадемѣиНаукЪчленомЪ Ордена Св: Владимира КавалеромЪ, и переведенное ИваномЪРижскимЪ

Тип издания: научное издание.

Формат издания: Книга, 75 страниц.

Используемые шрифтовые гарнитуры: Обыкновенный или брусковый, с засечками, контрастный, пропорциональный.

Качество бумаги: большинство страниц – очень хорошее, оставшиеся страницы – приемлемое.

Качество чернил: хорошее

## Задача 2

### Схема технологического процесса ввода текстовой информации:

1. Получить необходимый текст в pdf-виде.
2. Открыть текст в OCR-системе (FineReader)
3. В OCR-системе отобразить нужные страницы (3-28 по нумерации в книге, 5-30, если смотреть нумерацию страниц в FineReader-е, учитывая титульные страницы).
4. Удалить ненужные страницы.
5. Настроить язык распознавания документа. Для лучше распознавания рекомендуется ставить дополнительный язык Русский(Старая орфография).
6. Дополнительно можно настроить пользовательские эталоны. Для этого в меню Сервис-Опции необходимо выбрать вкладку "Распознать". Там нужно найти кнопку Эталоны, нажать на нее. Откроется Редактор эталонов. С помощью кнопки Новый необходимо создать новый эталон. Далее необходимо вернуться к окну Опции и там выбрать пункт Распознавание с обучением. Далее, выйдя из меню, стоит выбрать страницу, нажать на ней правой кнопкой и выбрать Распознать страницу. Весь документ распознавать не стоит, потому что основным эталонам OCR-система может обучиться за несколько страниц. При распознавании текста пользователю будет предложен вариант распознавания символа, и если этот вариант правильный, пользователь оставляет его, если неверный, то пользователь вводит правильный вариант распознавания символа.
7. Также дополнительно можно подключить пользовательский словарь.

### Точность распознавания страниц:

В следующей таблице приведем статистические данные распознавания текста без подключения дополнительных словарей и пользовательских эталонов.

*Таблица Статистические данные распознавания*

Страница	Количество символов	Количество неуверенно распознанных символов	Точность распознавания
3	625	94	85%
4	1047	181	83%
5	1063	207	81%
6	1035	80	92%
7	1056	86	92%
8	1017	100	90%
9	919	95	90%
10	1059	85	92%
11	1051	103	90%
12	1049	60	94%
13	1044	80	92%
14	1056	124	88%
15	1047	74	93%
16	1052	72	93%
17	1068	108	90%

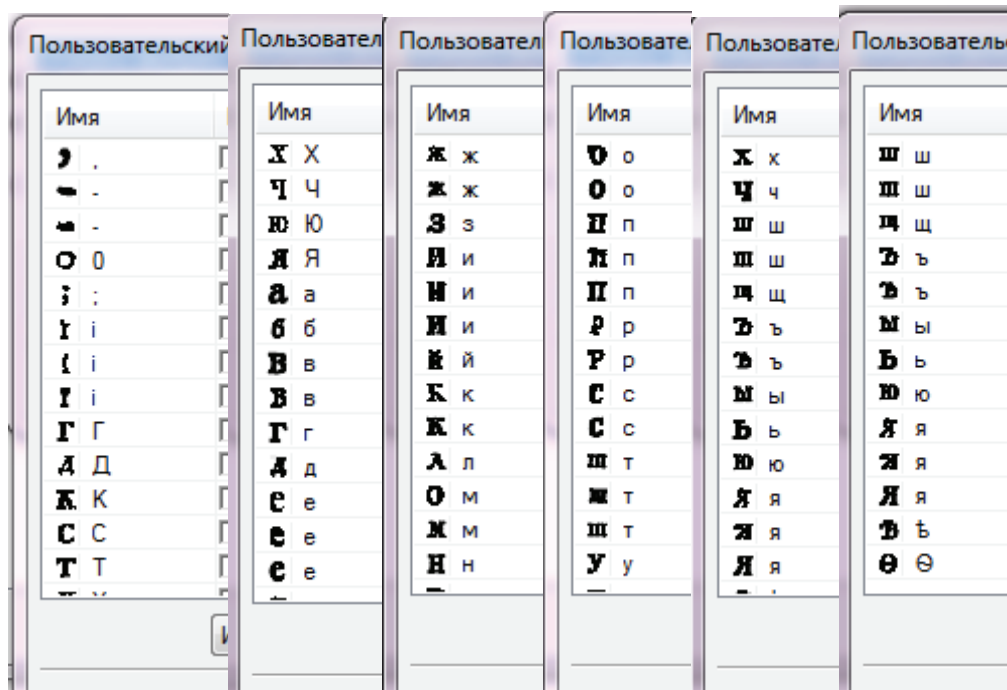
18	1031	78	92%
19	1033	75	93%
20	1062	59	94%
21	1038	59	94%
22	1062	59	94%
23	1045	73	93%
24	1020	88	91%
25	1064	95	91%
26	941	94	90%
27	1018	64	94%
28	1088	73	93%

### **Временный затраты**

Затраты на время можно рассматривать только с точки зрения распознавания, потому что все подготовительный работы проходят достаточно быстро и занимают разное время в зависимости от того, насколько пользователь хорошо знаком с данной OCR-системой.

Общие затраты на распознавание текста, состоящего из 26 страниц составляет около 30 секунд. Так как это время можно считать достаточно малым, то я считаю, что можно не приводить точное время, затраченное на распознавание.

Так как качество бумаги, чернил, и книги в целом были достаточно нормальные, то книга распозналась достаточно хорошо. Проблемы в основном встречались на первых трех страницах, что видно из процентов неуверенно распознанных символов. Эти проблемы возникли из-за того, что именно на этих страницах встречаются некоторые помарки или пометки в книге и немного замаявшиеся листы.

**Задача 3.****Пользовательские эталоны:****Алфавит:**

а, б, в, г, д, е, ж, з, и, к, л, м, н, о, п, р, с, т, у, ф, х, ц, ч, ш, щ, ъ, ы, ь, ю, я, ѐ, ё

А, Б, В, Г, Д, Е, Ж, З, И, К, Л, М, Н, О, П, Р, С, Т, У, Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ы, Ь, Ю, Я, Ё

**Эффективность распознавания с эталоном**

В данной таблице будут представлены статистические данные распознавания символов использованием пользовательских эталонов.

Таблица

*Статистические данные распознавания с использованием пользовательских эталонов*

Страница	Количество символов	Количество неуверенно распознанных символов	Точность распознавания
3	625	89	86%
4	1047	159	85%
5	1063	181	83%
6	1035	58	94%
7	1056	38	96%
8	1017	43	96%
9	919	40	96%
10	1059	64	94%
11	1051	65	94%

12	1049	80	92%
13	1044	45	96%
14	1056	63	94%
15	1047	106	90%
16	1052	60	94%
17	1068	59	94%
18	1031	83	92%
19	1033	59	94%
20	1062	69	94%
21	1038	55	95%
22	1062	55	95%
23	1045	57	95%
24	1020	72	93%
25	1064	79	93%
26	941	79	92%
27	1018	69	93%
28	1088	73	93%

С пользовательским эталоном система быстрее распознает текст, при тех же настройках, но подключенных пользовательских эталонах, система распознала текст за 20 секунд.

По точности распознавания можно сказать следующее – пользовательские эталоны уменьшают процент неуверенно распознанных символов, но не на много. Существенно уменьшается процент неуверенно распознанных символов тех страниц, на которых велось распознавание, и совсем не на много улучшается качество распознавания страниц с дефектами чернил, пятнами или помятостью страниц.

Время, ушедшее на обучение эталону – приблизительно 10 минут.

Соответственно можно сделать вывод, что если пользователю важно наиболее точное распознавание и не так важно время, потраченное на обучение эталона, то этот метод очень полезен. Чем на большем количестве страниц будет обучен эталон, тем точнее будет распознавание. Также этот метод полезен, если необходимо обработать большие объемы текста, тогда потери на обучение будут незначительными в соотношении к общему времени, а точность увеличится.

**Задача 4****Точность распознавания со словарем**

В таблице приведены статистические данные распознавания текста с использованием пользовательского словаря.

*Таблица  
Статистические данные распознавания с использованием дополнительного словаря*

Страница	Количество символов	Количество неуверенно распознанных символов	Точность распознавания
3	625	97	84%
4	1047	170	84%
5	1063	199	81%
6	1035	72	93%
7	1056	70	93%
8	1017	89	91%
9	919	59	94%
10	1059	67	94%
11	1051	79	92%
12	1049	92	91%
13	1044	57	95%
14	1056	76	93%
15	1047	138	87%
16	1052	52	95%
17	1068	70	93%
18	1031	108	90%
19	1033	60	94%
20	1062	64	94%
21	1038	65	94%
22	1062	55	95%
23	1045	69	93%
24	1020	90	91%
25	1064	73	93%
26	941	70	93%
27	1018	62	94%
28	1088	66	94%

По сравнению с распознаванием без словаря, точность улучшается, но не намного. В основном это касается символов, которых нет в современном алфавите (Это показано в пункте 5 на примере символа ъ). Но на некоторых страницах точность распознавания даже становится хуже. Время распознавания сильно не изменилось – тоже порядка 20 секунд. Поэтому по большей части целесообразно распознавать с использованием пользовательского словаря, в том случае, если в тексте очень много символов, которые уже не используются в данном современном алфавите, так как без словаря эти символы



заменяются аналогами в современном языке и заметить ошибку становится достаточно трудно.

## Задача 5

### Анализ основных ошибок

После корректировки текста источника были выявлены те символы, которые чаще всего распознавались неверно. Больше всего ошибок в букве “т”. Это связано с тем, что в тексте используется шрифт с большими засечками, и так как качество печати в 18 веке было не очень высоким, и по написанию буквы “т” и “ш” остаточного часто путаются и одна буква принимается за другую. Частота появления буквы “т” выше, чем буквы “ш”, поэтому замена “т” на “ш” встречалась гораздо чаще, чем “ш” на “т”.

Пример: “ошъ Балаклавы до теодосїи” вместо “отъ Балаклавы до теодосїи”.

Еще одна часто встречаемая ошибка – замена “я” на “л”. Такого рода распознавание символа я происходит по той же причине, что и “т” на “ш” – похожесть символов не очень хорошее качество печати в 18 веке. То, что шрифт с засечками на этот раз не играет большой роли.

Пример: “огнл” вместо “огня”.

Также популярная замена OCR-системы при распознавании – замена “їи” на “ш”. Это происходит из-за того, что у шрифта большие засечки и пробел между “ї” и “и” становится практически незаметен.

Пример: “разсужденш” вместо “разсужденїи”

Достаточно часто происходило нечеткое распознавание символа ъ либо на е, либо на другие символы. Это возникает при распознавании текста без подключения дополнительных словарей. Причиной этого являлось неупотребление символа ъ в современном русском языке, соответственно слова с этим символом не содержатся в словаре. Тогда OCR-система вставляла символ е на место распознаваемого символа, потому что он больше всего подходил к слову, ну и между символами есть некоторое сходство.

Судя по пункту 3 точность распознавания текста при распознавании с эталоном выше, чем у распознавания без эталона. Способ распознавания с использованием словаря показывает свою эффективность в основном на страницах, где очень много символов “ъ”, а весь остальной текст распознается даже несколько хуже. Поэтому проведем анализ основных ошибок на более эффективном способе распознавания – распознавании с эталоном.

Представим наиболее встречаемые ошибки в виде таблицы:

### Символ “т”:

Таблица Статистические данные распознавания символов “т”

Страница	Количество символов	Количество ошибок (в символах)	Точность распознавания символов
3	44	0	100%
4	47	0	100%
5	45	2	96%
6	49	3	94%

7	57	1	98%
8	55	6	89%
9	41	3	93%
10	61	1	98%
11	50	3	94%
12	56	4	93%
13	58	1	98%
14	62	5	92%
15	55	5	91%
16	42	0	100%
17	48	7	85%
18	49	2	96%
19	44	2	95%
20	42	6	86%
21	42	5	88%
22	41	1	98%
23	45	3	93%
24	42	12	71%
25	48	3	94%
26	39	5	87%
27	53	2	96%
28	53	3	94%

**Символ ”я”:***Таблица Статистические данные распознавания символов “я”*

Страница	Количество символов	Количество ошибок (в символах)	Точность распознавания символов
3	17	2	88%
4	21	1	95%
5	16	2	88%
6	24	1	96%
7	31	4	87%
8	19	1	95%
9	28	2	93%
10	14	0	100%
11	19	0	100%
12	20	9	55%
13	22	0	100%
14	11	0	100%
15	22	2	91%
16	24	2	92%
17	21	4	81%
18	25	0	100%
19	16	0	100%
20	26	1	96%
21	28	0	100%
22	29	1	97%

23	15	0	100%
24	22	2	91%
25	25	1	96%
26	27	4	85%
27	26	1	96%
28	25	0	100%

### Символ Ъ (распознавание с помощью пользовательских эталонов)

Таблица Статистические данные распознавания символов “Ъ”

Страница	Количество символов	Количество ошибок (в символах)	Точность распознавания символов
3	11	2	82%
4	16	3	81%
5	20	5	75%
6	12	1	92%
7	15	5	67%
8	23	4	83%
9	13	3	77%
10	24	2	92%
11	15	12	20%
12	19	4	79%
13	13	3	77%
14	13	1	92%
15	17	7	59%
16	15	10	33%
17	13	5	62%
18	14	1	93%
19	16	13	19%
20	15	2	87%
21	10	3	70%
22	20	19	5%
23	18	15	17%
24	16	3	81%
25	17	3	82%
26	16	13	19%
27	16	11	31%
28	15	3	80%

### Символ Ъ (распознавание с использованием пользовательского словаря)

Для сравнения эффективности распознавания символа “Ъ” в тексте при распознавании с эталоном и распознавании со словарем, приведем дополнительно еще и статистику появления ошибок в этом символе при использовании пользовательского словаря.

Григорьева С.С. ИУ5-92

В таблице представлены только страницы, на которых были обнаружены ошибки в символах ъ.

Таблица Статистические данные распознавания символов “Ъ”

Страница	Количество символов	Количество ошибок (в символах)	Точность распознавания символов
4	16	2	88%
7	15	1	93%
12	19	1	95%
19	16	1	94%
20	15	3	80%
28	15	1	93%

**Лексический анализ источника**

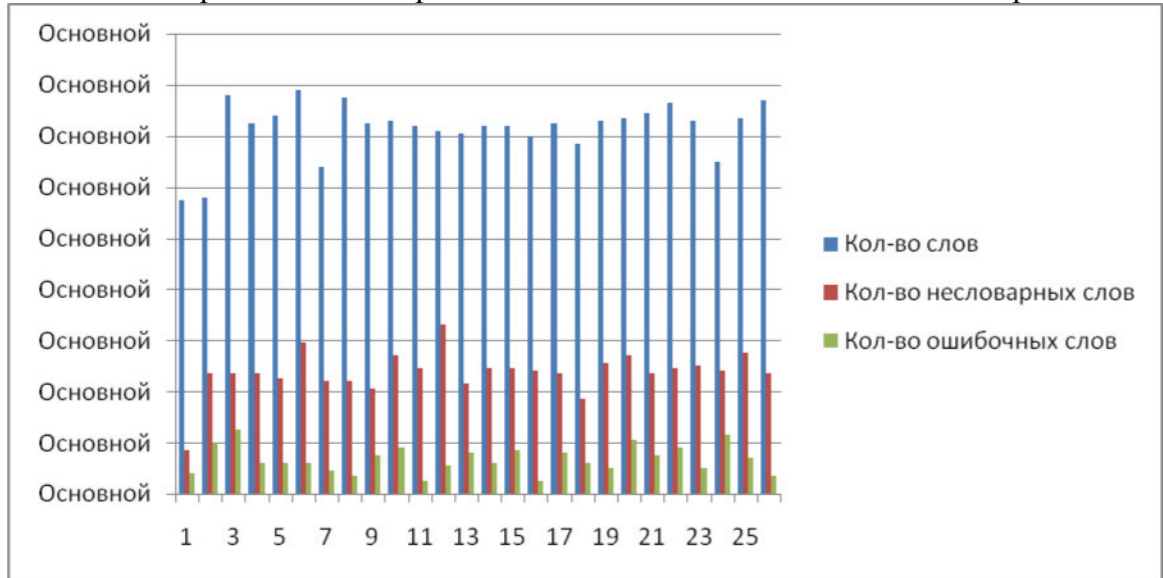
Проведем лексический анализ источника и рассмотрим несловарные слова. Результаты представим в виде следующей таблицы:

Таблица Статистические данные распознавания слов

Страница	Количество слов	Количество проверяемых несловарных слов	Количество ошибочных несловарных слов	Количество ошибочных слов	Точность распознавания несловарных слов	Точность распознавания слов
3	115	17	3	8	82%	93%
4	116	47	10	20	79%	83%
5	156	47	11	25	77%	84%
6	145	47	5	12	89%	92%
7	148	45	4	12	91%	92%
8	158	59	7	12	88%	92%
9	128	44	2	9	95%	93%
10	155	44	5	7	89%	95%
11	145	41	2	15	95%	90%
12	146	54	3	18	94%	88%
13	144	49	3	5	94%	97%
14	142	66	6	11	91%	92%
15	141	43	4	16	91%	89%
16	144	49	1	12	98%	92%
17	144	49	1	17	98%	88%
18	140	48	2	5	96%	96%
19	145	47	5	16	89%	89%
20	137	37	2	12	95%	91%
21	146	51	3	10	94%	93%
22	147	54	2	21	96%	86%
23	149	47	2	15	96%	90%
24	153	49	6	18	88%	88%
25	146	50	5	10	90%	93%
26	130	48	8	23	83%	82%

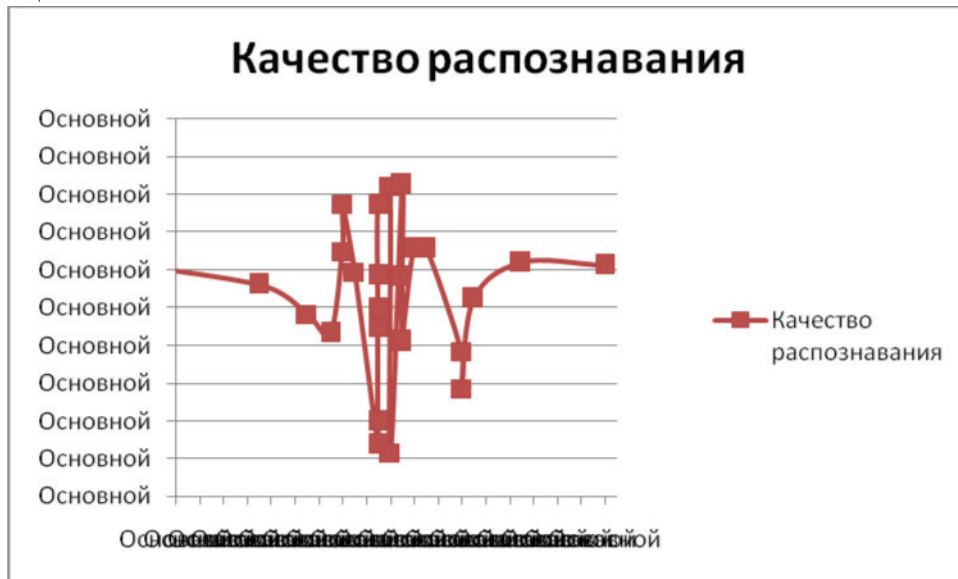
27	147	55	1	14	98%	90%
28	154	47	3	7	94%	95%

Рассмотрим на графике основные величины – количество слов на странице, количество несловарных слов на странице и количество ошибочных слов на странице:



При подробном рассмотрении графика, можно увидеть лишь некоторую зависимость между количеством слов на странице и количеством несловарных слов на странице, что достаточно логично, ведь в среднем на странице где-то треть слов не употребляются в современном русском языке либо их написание изменилось. Поэтому они помечаются как несловарные. В общем случае при увеличении количество слов на странице должно увеличиться и количество несловарных слов на странице. Но эта зависимость в среднем не оказывает влияния на количество ошибочных слов на странице. Это мы рассмотрим на другом графике.

Рассмотрим зависимость качества распознавания от количества несловарных слов на странице:



Зависимости между качеством распознавания и количеством несловарных слов не было выявлено, поэтому можно сделать вывод, что ошибочные слова присутствовали не только в несловарных словах.

**Анализ основных типов ошибок:***Таблица Типы ошибок*

Тип ошибки	Пример		Причина возникновения	Количество в тексте
	Ошибка	Исправление		
т-ш	ошъ Балаклары до Ѳеодосїи	отъ Балаклары до Ѳеодосїи	Схожесть написания символов, большие засечки	45
я-л	огнл	огня	Схожесть написания символов, не очень хорошее качество печати	33
с-е	вея	вся	Схожесть символов, не очень хорошее качество печати	2
їи-ш	разсужденш	разсужденїи	Большие засечки, маленькие интервалы между буквами	3
нї-ш	устроешя	устроенїя	Большие засечки, маленькие интервалы между буквами	2
ш-т	переметанных	перемешанных	Схожесть написания символов, большие засечки	4
е-с	окаменѣлостей	окаменѣлостей	Схожесть символов, не очень хорошее качество печати	2
ї-г	Минералогя	Минералогїя	Схожесть символов, отсутствие символа їв современном языке	3

## Задача 6

### Построение словника

В данной записке приведена только часть словника, потому что словник содержит около 1000 записей, что займет достаточно большое количества места, что будет несколько неоправданно.

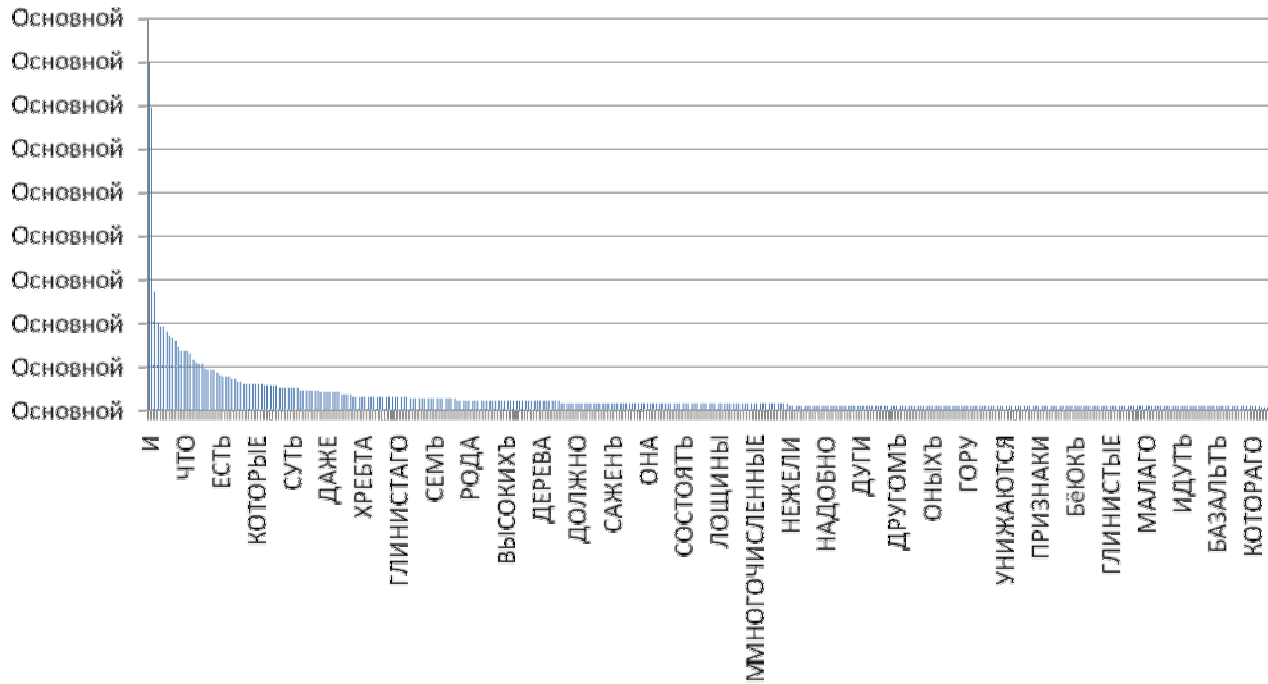
i	Слово	i/k	r	1/r	0,1*(1/r)
160	И	0,050972	1	1,000	0,100
139	ВЪ	0,044282	2	0,500	0,050
54	НА	0,017203	3	0,333	0,033
40	КЪ	0,012743	4	0,250	0,025
38	ПО	0,012106	5	0,200	0,020
38	СЪ	0,012106	5	0,200	0,020
36	ИЗЪ	0,011469	6	0,167	0,017
34	ИЛИ	0,010831	7	0,143	0,014
33	ГОРЪ	0,010513	8	0,125	0,013
32	ОТЪ	0,010194	9	0,111	0,011
29	ВЕСЬМА	0,009239	10	0,100	0,010
27	СЕГО	0,008601	11	0,091	0,009
27	ЧТО	0,008601	11	0,091	0,009
27	МЕЖДУ	0,008601	11	0,091	0,009
26	СЛОВЕЪ	0,008283	12	0,083	0,008
23	ТАКЪ	0,007327	13	0,077	0,008
22	НЕ	0,007009	14	0,071	0,007
21	СЛОИ	0,00669	15	0,067	0,007
21	КАКЪ	0,00669	15	0,067	0,007
19	ЖЕ	0,006053	16	0,063	0,006
18	ГОРЫ	0,005734	17	0,059	0,006
18	ТО	0,005734	17	0,059	0,006
18	ЧАСТО	0,005734	17	0,059	0,006
17	СЕЙ	0,005416	18	0,056	0,006
16	ЕСТЬ	0,005097	19	0,053	0,005
15	А	0,004779	20	0,050	0,005
15	КОТОРЫХЪ	0,004779	20	0,050	0,005
15	ОНЪ	0,004779	20	0,050	0,005
14	ИНОГДА	0,00446	21	0,048	0,005
14	КАМЕНЬ	0,00446	21	0,048	0,005
13	ИХЪ	0,004141	22	0,045	0,005
13	СТЛАНЦА	0,004141	22	0,045	0,005
12	ТАВРИЧЕСКОЙ	0,003823	23	0,043	0,004
12	ОБЛАСТИ	0,003823	23	0,043	0,004
12	ПРИ	0,003823	23	0,043	0,004
12	НИ	0,003823	23	0,043	0,004
12	КОТОРЫЕ	0,003823	23	0,043	0,004
12	РЯДА	0,003823	23	0,043	0,004
12	НАХОДЯТСЯ	0,003823	23	0,043	0,004
11	ЕГО	0,003504	24	0,042	0,004



II	ПОЧТИ	0,003504	24	0,042	0,004
II	СИХЪ	0,003504	24	0,042	0,004
II	ЧАСТИ	0,003504	24	0,042	0,004
II	ДО	0,003504	24	0,042	0,004
IO	НИХЪ	0,003186	25	0,040	0,004
IO	НО	0,003186	25	0,040	0,004
IO	ВО	0,003186	25	0,040	0,004
IO	ВИДНЫ	0,003186	25	0,040	0,004
IO	СУТЬ	0,003186	25	0,040	0,004
IO	СТОРОНЫ	0,003186	25	0,040	0,004
IO	КАМНЯ	0,003186	25	0,040	0,004
9	НАКОНЕЦЪ	0,002867	26	0,038	0,004
9	БЫ	0,002867	26	0,038	0,004
9	КОТОРОЙ	0,002867	26	0,038	0,004
9	ОНИ	0,002867	26	0,038	0,004
9	БЕРЕГУ	0,002867	26	0,038	0,004
9	Я	0,002867	26	0,038	0,004
9	СЛОЯМИ	0,002867	26	0,038	0,004
8	НАПРОТИВЪ	0,002549	27	0,037	0,004
8	ДРУГИХЪ	0,002549	27	0,037	0,004
8	ДАЖЕ	0,002549	27	0,037	0,004
8	НАХОДИТСЯ	0,002549	27	0,037	0,004
8	ГОРАЗДО	0,002549	27	0,037	0,004
8	ОБЫКНОВЕННО	0,002549	27	0,037	0,004
8	БЫВАЮТЬ	0,002549	27	0,037	0,004
7	СОСТОИТЬ	0,00223	28	0,036	0,004
7	КОТОРЫЯ	0,00223	28	0,036	0,004
7	СО	0,00223	28	0,036	0,004
7	ТАМЪ	0,00223	28	0,036	0,004
6	ПОВЕРЬХНОСТИ	0,001911	29	0,034	0,003
6	КОТОРАЯ	0,001911	29	0,034	0,003
6	ТОГО	0,001911	29	0,034	0,003
6	ХРЕБТА	0,001911	29	0,034	0,003
6	СОБОЮ	0,001911	29	0,034	0,003
6	ЛЕЖАЩИХЪ	0,001911	29	0,034	0,003
6	МОЖНО	0,001911	29	0,034	0,003
6	МОРЯ	0,001911	29	0,034	0,003
6	УЖЕ	0,001911	29	0,034	0,003
6	РЯДЪ	0,001911	29	0,034	0,003
6	МАЛО	0,001911	29	0,034	0,003
6	ПЕРВАГО	0,001911	29	0,034	0,003
6	ИЗВЕСТНОЙ	0,001911	29	0,034	0,003
6	НЕМЪ	0,001911	29	0,034	0,003
6	ИЗВЕСТНАГО	0,001911	29	0,034	0,003
6	ГЛИНИСТАГО	0,001911	29	0,034	0,003
6	ОДНАКО	0,001911	29	0,034	0,003
6	ПЕЩАНОЙ	0,001911	29	0,034	0,003
6	ОДНА	0,001911	29	0,034	0,003
5	БЕРЕГА	0,001593	30	0,033	0,003
5	НЕЧУВСТВТЕЛЬНО	0,001593	30	0,033	0,003

5	ВЫШЕ	0,001593	30	0,033	0,003
5	БЫТЬ	0,001593	30	0,033	0,003
5	ДРУГАГО	0,001593	30	0,033	0,003
5	ЯВСТВЕННО	0,001593	30	0,033	0,003
5	ОДИНЪ	0,001593	30	0,033	0,003
5	БОЛЬШЕЙ	0,001593	30	0,033	0,003
5	СЕМЬ	0,001593	30	0,033	0,003

### Функция распределения частот:



## Литература

1. Филиппович Анна. Методические указания к выполнению курсовой работы по теме «Информационные технологии сохранения исторических и культурных ценностей России.»
2. Филиппович Анна. Исследование эффективности систем оптического распознавания текстов. // Интеллектуальные технологии и системы. Сборник учебно-методических работ и статей аспирантов и студентов. Выпуск 7 / Сост. и ред. Ю.Н. Филипповича. – М. Изд-во ООО «Эликс+» 2005. – С. 272-297
3. ABBYY® FineReaderВерсия 10Руководство пользователя© 2009

## Приложение

## Исходная страница текста

7

пихъ подробное поняшіе. Сей рядъ горъ простираешся въ длину 'отъ Георгіевскаго монастыря и отъ одного мыса на Гераклеопическомъ Херсонесѣ, называемаго шѣмъ же именемъ, даже до находящейся близъ Θεодосіи горы Карадагъ; а ширину имѣеть различную, отъ 20 до 30 верстъ и болѣе, разумѣя отъ берега во внутренность земли. Соспавляющія его горы суть самыя высокія во всей Таврической Области, и раздѣляются на три главныя возвышенности, то есть, на гору Чашырдагъ, почитаемую высочайшею изъ всѣхъ Таврическихъ горъ, такъ что прямоспоящую высоту ея можно положить въ тысячу сто сажень. Она находится почти противъ середины горнаго берега. Другія двѣ возвышенности суть двѣ Яйлли, копорыя соспавляютъ кряжи непрерывныхъ горъ, весьма высокихъ, утесистыхъ со стороны моря, но опускающихся уступами, соспавляющими великія ровнины, наклоненныхъ къ Сѣверу, и пресѣкаемыхъ лощинами. Сіи два кряжа непрерывныхъ горъ отдѣлены отъ Чашырдага, съ копорымъ почти равную имѣютъ высоту, двумя весьма глубокими и узкими долинами, копорыя пересѣкаютъ хребетъ отъ Сѣвера къ Югу,

## Распознанная страница текста

пихъ подробноепонятіе. Сей рїидъгоръ простираетслвъ длину 'отъГеорггевскаго монастыря и отъ одного мыса на ГераклеотическомъХерсонесѣ, называемаготѣмъ же именемъ, даже до находящейся близъ Θεодосіи горы Карадагъ; а ширину имѣеть

различную, отъ 20 до 30 верстъ и болізе\*, разумѣютъ берега во внутренность земли. Составляющія его горы суть самыявысокія во всей Таврической Области, и раздѣляются на три главныя возвышенности, то есть, на гору Чатырдагъ, почитаемую высочайшею изъвсѣхъТаврическихъгоръ, такъ что прямостоящую высоту ея можно положить въ тысячу сто сажень. Она находится почти противъ середины горнаго берега. Другіядвѣ возвышенности суть двѣЛйлли, которыясоставляютъ кряжи непрерывныхъгоръ, весьма высокыхъ, утесистыхъ со стороны моря, но опускающихся уступами, составляющими великіяровнины, наклоненныхъкъ Северу, и пресѣкаемыхъ лощинами. Сіи два кряжа непрерывныхъгоръ отделены отъЧатырдага, съкоторымъ почти равную имѣютъ высоту, двумя'весьма глубокими и узкими долинами, которыяпересекаютъхребетъотъСъверакъ Югу,

#### Отредактированная страница

нихъ подробное понятіе. Сей рядъгоръ простирается въ длину отъГеоргіевскаго монастыря и отъ одного мыса на ГераклеопическомъХерсонесѣ, называемагопѣмъжеименемъ, дажедо находящейсяблизъѲеодосіигорыКарадагъ; аширинуимѣетъразличную, отъ 20 до 30 верспѣиболѣе, разумѣяотъ берега во внутренность земли. Составляющія его горы суть самыявысокія во всей Таврической Области, и раздѣляютсяна три главныя возвышенности, то есть, на горуЧатырдагъ, почитаемую высочайшею изъвсѣхъТаврическихъгоръ, такъ что прямостоящую высоту ея можно положить въ тысячу сто сажень. Она находится почти противъ середины горнаго берега. Другіядвѣ возвышенности суть двѣЛйлли, которыясоставляютъ кряжи непрерывныхъгоръ, весьма высокыхъ, утесистыхъ со стороны моря, но опускающихся уступами, составляющими великіяровнины, наклоненныхъкъСеверу, и пресѣкаемыхъ лощинами. Сіи два кряжа непрерывныхъгоръ отделены отъЧатырдага, съкоторымъ почти равную имѣютъ высоту, двумя весьма глубокими и узкими долинами, которыяпересекаютъхребетъотъСъверакъ Югу,