

CLB Search

(C) Веретенников А. Б. 2010.

<http://www.veretennikov.org>

Версия 1.0.0.1

Copyright © Веретенников А. Б. 2010.
В данном документе описываются результаты
экспериментов
Последнее обновление документа 2010.05.22.
Версия программы 1.0.0.1.

В данном файле приведены результаты экспериментов. Для каждого эксперимента приведен конфигурационный файл индекса и дано описание используемых аппаратных и программных средств. Приведенные данные могут быть полезны для определения требуемой конфигурации индекса.

Далее перечислены результаты экспериментов.

Время эксперимента: 2010.05.31 18:51:18

Обрабатываемые данные

Название	Значение
Всего файлов	421 111
Проиндексировано файлов	417 612
Суммарный размер	91 524 793 627
Суммарный размер текста	91 524 381 549

Информация о системе

Название	Значение
ОС	Microsoft Windows Server 2008 Enterprise Edition Service Pack 2 (build 6002), 64-bit
Оперативная память	12 882 354 176
Процессор	Intel(R) Core(TM) i7 CPU 920 @ 2.67GHz

Время

Название	Значение
Время чтения файлов	03:06:39:034
Время создания индекса	02:14:19:000
Общее время	05:20:58:036
Общее время операции	06:01:09:740
Время определения формата файлов	00:02:43:259
Время морфологического анализа	00:35:40:848
Время поиска похожих документов	00:40:11:514

Слова

Название	Значение
Всего слов	12 860 142 374
'Известных' слов	11 802 039 154

Форм 'известных' слов	17 106 426 020
'Неизвестных' слов	1 058 103 220
Новых 'известных' слов	214 747
Новых 'неизвестных' слов	81 760 425
Суммарная длина слов	68 763 189 396

Размеры файлов

Название	Значение
Размер индекса известных слов	53 055 989 035
Размер индекса неизвестных слов	13 157 709 903
Размер дополнительного индекса часто встречающихся слов	8 199 126 417
Суммарный размер индексов	74 412 825 355
Размер репозитория	60 046 737 732
Размер индекса последовательностей	4 690 969 124
Размер индекса похожих документов	10 435 408
Размер списка файлов	163 194 048

Конфигурационный файл:

Название	Значение
NAME	1\1
BUFFER	1024*1024*1024
APPEND_BUFFER	1024*1024*64
CACHE_MULTIPLIER	15
GROUP_BUFFER_SIZE	1024*1024
DATA_DETAILS	WORD
REPOSITORY_ENABLE	1
REPOSITORY_COMPRESSION	Huffman
STORE_STOP_WORDS	0
STORE_STOP_WORDS_ONE_PER_FILE	1
MAKE_STOP_WORDS_INDEXES	1
FORMATS	rtf, pdf, djvu, msword
ARCHIVE_FORMATS	rar\rar.dll, cab\cab.dll, 7Z\7Z.dll
USE_TRANSFORMATION_SERVICE	1
KNOWN_WORDS.CLUSTER_SIZE	1024*16
UNKNOWN_WORDS.CLUSTER_SIZE	1024*4
UNKNOWN_WORDS.TREE_PAGE_SIZE	1024*4
SEQUENCE_REPOSITORY_ENABLE	1
SIMULAR_INDEX_ENABLE	1
SEQUENCE_BUFFER_SIZE	1024*1024*1024
SIMULAR_REQUIRED_PERCENT	0

TEMP_PATH	%INDEX_PATH%\Temp\
-----------	--------------------

Время эксперимента: 2010.06.01 18:17:17

Как в предыдущем случае, но с отключенными индексом похожих документов и репозитарием.

Обрабатываемые данные

Название	Значение
Всего файлов	421 111
Проиндексировано файлов	417 612
Суммарный размер	91 524 793 627
Суммарный размер текста	91 524 381 549

Информация о системе

Название	Значение
ОС	Microsoft Windows Server 2008 Enterprise Edition Service Pack 2 (build 6002), 64-bit
Оперативная память	12 882 354 176
Процессор	Intel(R) Core(TM) i7 CPU 920 @ 2.67GHz

Время

Название	Значение
Время чтения файлов	02:27:17:197
Время создания индекса	02:17:43:000
Общее время	04:45:00:199
Общее время операции	04:45:00:237
Время определения формата файлов	00:02:39:839
Время морфологического анализа	00:35:01:284
Время поиска похожих документов	00:00:00:006

Слова

Название	Значение
Всего слов	12 860 142 374
'Известных' слов	11 802 039 154
Форм 'известных' слов	17 106 426 020
'Неизвестных' слов	1 058 103 220
Новых 'известных' слов	214 747

Новых 'неизвестных' слов	81 760 425
Суммарная длина слов	68 763 189 396

Размеры файлов

Название	Значение
Размер индекса известных слов	52 949 272 899
Размер индекса неизвестных слов	13 145 618 479
Размер дополнительного индекса часто встречающихся слов	8 149 863 825
Суммарный размер индексов	74 244 755 203
Размер репозитория	0
Размер индекса последовательностей	0
Размер индекса похожих документов	0
Размер списка файлов	67 921 664

Конфигурационный файл:

Название	Значение
NAME	1\1
BUFFER	1024*1024*1024
APPEND_BUFFER	1024*1024*64
CACHE_MULTIPLIER	15
GROUP_BUFFER_SIZE	1024*1024
DATA_DETAILS	WORD
REPOSITORY_ENABLE	0
REPOSITORY_COMPRESSION	Huffman
STORE_STOP_WORDS	0
STORE_STOP_WORDS_ONE_PER_FILE	1
MAKE_STOP_WORDS_INDEXES	1
FORMATS	rtf, pdf, djvu, msword
ARCHIVE_FORMATS	rar\rar.dll, cab\cab.dll, 7Z\7Z.dll
USE_TRANSFORMATION_SERVICE	1
KNOWN_WORDS.CLUSTER_SIZE	1024*16
UNKNOWN_WORDS.CLUSTER_SIZE	1024*4
UNKNOWN_WORDS.TREE_PAGE_SIZE	1024*4
SEQUENCE_REPOSITORY_ENABLE	0
SIMULAR_INDEX_ENABLE	0
SEQUENCE_BUFFER_SIZE	1024*1024*1024
SIMULAR_REQUIRED_PERCENT	0
TEMP_PATH	%INDEX_PATH%\Temp\

Время эксперимента: 2010.06.02 17:12:28

Обрабатываемые данные

Название	Значение
Всего файлов	421 111
Проиндексировано файлов	417 612
Суммарный размер	91 524 793 627
Суммарный размер текста	91 524 381 549

Информация о системе

Название	Значение
ОС	Microsoft Windows XP Professional Service Pack 2 (build 2600)
Оперативная память	1 054 355 456
Процессор	Intel(R) Celeron(R) CPU 430 @ 1.80GHz

Время

Название	Значение
Время чтения файлов	07:26:52:186
Время создания индекса	04:04:58:000
Общее время	11:31:50:203
Общее время операции	15:36:09:875
Время определения формата файлов	00:06:05:902
Время морфологического анализа	01:49:04:840
Время поиска похожих документов	04:04:17:906

Слова

Название	Значение
Всего слов	12 860 142 374
'Известных' слов	11 802 039 153
Форм 'известных' слов	17 106 426 019
'Неизвестных' слов	1 058 103 221
Новых 'известных' слов	214 747
Новых 'неизвестных' слов	81 760 425
Суммарная длина слов	68 763 189 396

Размеры файлов

Название	Значение
----------	----------

Размер индекса известных слов	53 820 951 825
Размер индекса неизвестных слов	14 467 392 609
Размер дополнительного индекса часто встречающихся слов	8 199 773 633
Суммарный размер индексов	76 488 118 067
Размер репозитория	60 046 737 732
Размер индекса последовательностей	4 690 969 124
Размер индекса похожих документов	10 435 408
Размер списка файлов	165 282 108

Конфигурационный файл:

Название	Значение
NAME	1\1
BUFFER	1024*1024*512
APPEND_BUFFER	1024*1024*64
CACHE_MULTIPLIER	15
GROUP_BUFFER_SIZE	1024*1024
DATA_DETAILS	WORD
REPOSITORY_ENABLE	1
REPOSITORY_COMPRESSION	Huffman
STORE_STOP_WORDS	0
STORE_STOP_WORDS_ONE_PER_FILE	1
MAKE_STOP_WORDS_INDEXES	1
FORMATS	rtf, pdf, djvu, msword
ARCHIVE_FORMATS	rar\rar.dll, cab\cab.dll, 7Z\7Z.dll
USE_TRANSFORMATION_SERVICE	1
KNOWN_WORDS.CLUSTER_SIZE	1024*16
UNKNOWN_WORDS.CLUSTER_SIZE	1024*4
UNKNOWN_WORDS.TREE_PAGE_SIZE	1024*4
SEQUENCE_REPOSITORY_ENABLE	1
SIMULAR_INDEX_ENABLE	1
SEQUENCE_BUFFER_SIZE	512*1024*1024
SIMULAR_REQUIRED_PERCENT	0
TEMP_PATH	%INDEX_PATH%\Temp\

Время эксперимента: 2010.05.30 12:46:03

Обрабатываемые данные

Название	Значение
Всего файлов	421 110
Проиндексировано файлов	417 728

Суммарный размер	93 503 724 931
Суммарный размер текста	93 503 312 853

Информация о системе

Название	Значение
ОС	Microsoft Windows XP Professional Service Pack 2 (build 2600)
Оперативная память	1 054 355 456
Процессор	Intel(R) Celeron(R) CPU 430 @ 1.80GHz

Время

Название	Значение
Время чтения файлов	07:57:12:000
Время создания индекса	04:17:27:000
Общее время	12:14:39:000
Общее время операции	12:14:39:219
Время определения формата файлов	00:06:40:851
Время морфологического анализа	01:58:52:857
Время поиска похожих документов	00:00:00:016

Слова

Название	Значение
Всего слов	13 128 343 279
'Известных' слов	12 047 135 311
Форм 'известных' слов	17 450 649 440
'Неизвестных' слов	1 081 207 968
Новых 'известных' слов	215 362
Новых 'неизвестных' слов	82 625 078
Суммарная длина слов	70 246 528 149

Размеры файлов

Название	Значение
Размер индекса известных слов	54 748 231 988
Размер индекса неизвестных слов	14 658 069 716
Размер дополнительного индекса часто встречающихся слов	8 085 338 236
Суммарный размер индексов	77 491 639 940
Размер репозитория	61 340 581 387

Размер индекса последовательностей	0
Размер индекса похожих документов	0
Размер списка файлов	161 506 005

Конфигурационный файл:

Название	Значение
NAME	1\1
BUFFER	1024*1024*512
APPEND_BUFFER	1024*1024*64
CACHE_MULTIPLIER	15
GROUP_BUFFER_SIZE	1024*1024
DATA_DETAILS	WORD
REPOSITORY_ENABLE	1
REPOSITORY_COMPRESSION	Huffman
STORE_STOP_WORDS	0
STORE_STOP_WORDS_ONE_PER_FILE	1
MAKE_STOP_WORDS_INDEXES	1
FORMATS	rtf, pdf, djvu, msword
ARCHIVE_FORMATS	rar\rar.dll, cab\cab.dll, 7Z\7Z.dll
USE_TRANSFORMATION_SERVICE	1
KNOWN_WORDS.CLUSTER_SIZE	1024*16
UNKNOWN_WORDS.CLUSTER_SIZE	1024*4
UNKNOWN_WORDS.TREE_PAGE_SIZE	1024*4
SEQUENCE_REPOSITORY_ENABLE	0
SIMULAR_INDEX_ENABLE	0
SEQUENCE_BUFFER_SIZE	512*1024*1024
SIMULAR_REQUIRED_PERCENT	0
TEMP_PATH	%INDEX_PATH%\Temp\

В следующем разделе приведены результаты экспериментов создания и последующего обновления индекса.

Структура эксперимента:

- 1) Создание индекса для базового массива документов.
- 2) Добавление в индекс небольшого файла.
- 3) Добавление в индекс 1, 2, 3, ..., 9 Мб данных.
- 4) Добавление в индекс 10, 20, 30, ..., 90 Мб данных.
- 5) Добавление в индекс 100, 200, 300, ..., 900 Мб данных.
- 6) Добавление в индекс 1000, 2000, 3000, ..., 10000 Мб данных.

При добавлении в индекс каждой порции данных после добавления сохраняется весь кеш и программа создания индекса завершает работу, затем она снова запускается и добавляет следующую порцию данных. При работе с основным файлом индекса кеширование операционной системы отключается. Таким образом мы проверяем скорость обновления индекса для данных самого разного размера.

Результаты экспериментов оформлены в виде таблицы со следующими колонками: размер - размер текста, добавляемый в индекс, общее время - суммарное время, составленное из времени чтения файлов и времени записи индекса. В последних двух колонках указан суммарный объем прочитанных и записанных блоков в основном файле индекса. Все обрабатываемые файлы представляли собой обычный текст.

Время эксперимента: 2010.06.04 18:53:50

Информация о системе

Название	Значение
ОС	Microsoft Windows Server 2008 Enterprise Edition Service Pack 2 (build 6002), 64-bit
Оперативная память	12 882 354 176
Процессор	Intel(R) Core(TM) i7 CPU 920 @ 2.67GHz

Конфигурационный файл:

Название	Значение
NAME	H:\Indexes\IndexF\Index
BUFFER	1024*1024*1024

APPEND_BUFFER	1024*1024*64
CACHE_MULTIPLIER	15
GROUP_BUFFER_SIZE	1024*1024
DATA_DETAILS	WORD
REPOSITORY_ENABLE	1
REPOSITORY_COMPRESSION	Huffman
STORE_STOP_WORDS	0
STORE_STOP_WORDS_ONE_PER_FILE	1
MAKE_STOP_WORDS_INDEXES	1
FORMATS	rtf, pdf, djvu, msword
ARCHIVE_FORMATS	rar\rar.dll, cab\cab.dll, 7Z\7Z.dll
USE_TRANSFORMATION_SERVICE	1
KNOWN_WORDS.CLUSTER_SIZE	1024*16
UNKNOWN_WORDS.CLUSTER_SIZE	1024*4
UNKNOWN_WORDS.TREE_PAGE_SIZE	1024*4
SEQUENCE_REPOSITORY_ENABLE	0
SIMULAR_INDEX_ENABLE	0
SEQUENCE_BUFFER_SIZE	1024*1024*1024
SIMULAR_REQUIRED_PERCENT	0
TEMP_PATH	%INDEX_PATH%\Temp\

Результаты эксперимента:

Размер	Общее время, ч.:м.:с.	Время записи индекса	Чтение кластеров	Запись кластеров
85.23 Гб	05:10:55	02:14:41	8.45 Гб	83.73 Гб
1.41 Кб	00:01:03	00:00:58	4.65 Гб	4.46 Мб
1.08 Мб	00:02:05	00:02:00	4.65 Гб	330.01 Мб
2.17 Мб	00:02:26	00:02:22	4.65 Гб	463.16 Мб
3.09 Мб	00:02:44	00:02:39	4.65 Гб	584.90 Мб
4.03 Мб	00:02:57	00:02:52	4.66 Гб	684.12 Мб
5.03 Мб	00:03:07	00:03:02	4.65 Гб	748.27 Мб
6.25 Мб	00:03:13	00:03:08	4.65 Гб	825.67 Мб
7.31 Мб	00:03:24	00:03:19	4.66 Гб	872.14 Мб
8.21 Мб	00:03:27	00:03:22	4.66 Гб	903.69 Мб
9.19 Мб	00:03:32	00:03:27	4.65 Гб	962.68 Мб
10.02 Мб	00:03:40	00:03:34	4.65 Гб	1004.10 Мб
10.02 Мб	00:03:36	00:03:31	4.66 Гб	1013.51 Мб
20.03 Мб	00:04:14	00:04:07	4.66 Гб	1.32 Гб
30.37 Мб	00:04:34	00:04:25	4.67 Гб	1.47 Гб
40.19 Мб	00:05:09	00:05:00	4.68 Гб	1.63 Гб
50.60 Мб	00:05:09	00:04:59	4.68 Гб	1.74 Гб
60.17 Мб	00:05:27	00:05:15	4.68 Гб	1.79 Гб
70.26 Мб	00:05:29	00:05:17	4.67 Гб	1.80 Гб

80.46 Мб	00:05:39	00:05:25	4.71 Гб	1.90 Гб
90.01 Мб	00:05:47	00:05:33	4.69 Гб	1.90 Гб
100.09 Мб	00:05:50	00:05:35	4.71 Гб	1.95 Гб
200.00 Мб	00:07:18	00:06:39	4.81 Гб	2.32 Гб
300.00 Мб	00:08:16	00:07:22	4.85 Гб	2.48 Гб
400.44 Мб	00:10:33	00:09:19	4.96 Гб	2.74 Гб
500.13 Мб	00:11:06	00:09:43	5.00 Гб	2.90 Гб
600.01 Мб	00:11:44	00:10:10	5.06 Гб	3.07 Гб
700.90 Мб	00:12:22	00:10:36	5.14 Гб	3.29 Гб
800.62 Мб	00:13:09	00:11:12	5.19 Гб	3.41 Гб
900.24 Мб	00:13:29	00:11:21	5.26 Гб	3.48 Гб
1000.54 Мб	00:14:13	00:11:53	5.27 Гб	3.61 Гб
1.95 Гб	00:21:25	00:16:56	5.79 Гб	4.98 Гб
2.93 Гб	00:26:42	00:19:57	6.22 Гб	6.38 Гб
3.90 Гб	00:31:03	00:22:10	6.83 Гб	7.80 Гб
4.88 Гб	00:37:21	00:26:18	7.28 Гб	9.16 Гб
5.85 Гб	00:42:20	00:28:28	7.52 Гб	10.34 Гб
6.83 Гб	00:46:54	00:30:43	8.08 Гб	11.77 Гб
7.81 Гб	00:51:27	00:33:14	8.30 Гб	12.77 Гб
8.78 Гб	00:56:06	00:35:39	8.73 Гб	13.94 Гб
9.76 Гб	01:03:32	00:40:42	9.45 Гб	15.52 Гб

Время эксперимента: 2010.06.05 10:02:58

Информация о системе

Название	Значение
ОС	Microsoft Windows XP Professional Service Pack 2 (build 2600)
Оперативная память	1 054 355 456
Процессор	Intel(R) Celeron(R) CPU 430 @ 1.80GHz

Конфигурационный файл:

Название	Значение
NAME	E:\Indexes\IndexF\Index
BUFFER	512*1024*1024
APPEND_BUFFER	1024*1024*64
CACHE_MULTIPLIER	15
GROUP_BUFFER_SIZE	1024*1024
DATA_DETAILS	WORD
REPOSITORY_ENABLE	1
REPOSITORY_COMPRESSION	Huffman
STORE_STOP_WORDS	0

STORE_STOP_WORDS_ONE_PER_FILE	1
MAKE_STOP_WORDS_INDEXES	1
FORMATS	rtf, pdf, djvu, msword
ARCHIVE_FORMATS	rar\rar.dll, cab\cab.dll, 7Z\7Z.dll
USE_TRANSFORMATION_SERVICE	1
KNOWN_WORDS.CLUSTER_SIZE	1024*16
UNKNOWN_WORDS.CLUSTER_SIZE	1024*4
UNKNOWN_WORDS.TREE_PAGE_SIZE	1024*4
SEQUENCE_REPOSITORY_ENABLE	0
SIMULAR_INDEX_ENABLE	0
SEQUENCE_BUFFER_SIZE	1024*1024*1024
SIMULAR_REQUIRED_PERCENT	0
TEMP_PATH	%INDEX_PATH%\Temp\

Результаты эксперимента:

Размер	Общее время, ч.:м.:с.	Время записи индекса	Чтение кластеров	Запись кластеров
85.23 Гб	11:22:33	03:54:45	8.82 Гб	82.06 Гб
1.41 Кб	00:01:52	00:01:42	5.40 Гб	4.46 Мб
1.08 Мб	00:02:50	00:02:40	5.40 Гб	329.89 Мб
2.17 Мб	00:03:01	00:02:51	5.40 Гб	462.78 Мб
3.09 Мб	00:03:11	00:03:00	5.40 Гб	584.42 Мб
4.03 Мб	00:03:19	00:03:09	5.41 Гб	683.98 Мб
5.03 Мб	00:03:23	00:03:13	5.40 Гб	747.86 Мб
6.25 Мб	00:03:35	00:03:24	5.40 Гб	825.89 Мб
7.31 Мб	00:03:37	00:03:26	5.41 Гб	871.25 Мб
8.21 Мб	00:03:38	00:03:26	5.41 Гб	903.60 Мб
9.19 Мб	00:03:45	00:03:34	5.40 Гб	962.93 Мб
10.02 Мб	00:03:48	00:03:36	5.40 Гб	1004.58 Мб
10.02 Мб	00:03:43	00:03:31	5.41 Гб	1013.75 Мб
20.03 Мб	00:04:31	00:04:16	5.41 Гб	1.32 Гб
30.37 Мб	00:04:44	00:04:26	5.42 Гб	1.48 Гб
40.19 Мб	00:05:23	00:05:03	5.42 Гб	1.65 Гб
50.60 Мб	00:05:24	00:05:00	5.43 Гб	1.76 Гб
60.17 Мб	00:05:34	00:05:09	5.43 Гб	1.81 Гб
70.26 Мб	00:05:38	00:05:09	5.43 Гб	1.82 Гб
80.46 Мб	00:05:56	00:05:25	5.46 Гб	1.92 Гб
90.01 Мб	00:05:59	00:05:25	5.44 Гб	1.93 Гб
100.09 Мб	00:06:05	00:05:28	5.46 Гб	1.98 Гб
200.00 Мб	00:10:39	00:09:00	5.56 Гб	2.44 Гб
300.00 Мб	00:12:06	00:09:47	5.60 Гб	2.61 Гб
400.44 Мб	00:14:39	00:11:36	5.73 Гб	2.89 Гб
500.13 Мб	00:15:59	00:12:11	5.80 Гб	3.02 Гб

600.01 Мб	00:16:45	00:12:41	5.85 Гб	3.19 Гб
700.90 Мб	00:18:15	00:13:13	5.94 Гб	3.42 Гб
800.62 Мб	00:19:14	00:13:47	6.01 Гб	3.53 Гб
900.24 Мб	00:20:28	00:13:56	6.06 Гб	3.63 Гб
1000.54 Мб	00:20:40	00:14:21	6.11 Гб	3.74 Гб
1.95 Гб	00:31:14	00:18:17	6.67 Гб	5.13 Гб
2.93 Гб	00:39:45	00:21:52	7.15 Гб	6.54 Гб
3.90 Гб	00:49:17	00:24:55	7.82 Гб	7.95 Гб
4.88 Гб	01:01:48	00:28:08	8.30 Гб	9.30 Гб
5.85 Гб	01:06:10	00:30:49	8.56 Гб	10.45 Гб
6.83 Гб	01:19:40	00:33:37	9.08 Гб	11.78 Гб
7.81 Гб	01:24:36	00:36:38	9.28 Гб	12.77 Гб
8.78 Гб	04:26:26	03:32:35	9.67 Гб	13.96 Гб
9.76 Гб	03:12:07	02:04:33	10.36 Гб	15.51 Гб

Время эксперимента: 2010.06.05 19:48:01

Информация о системе

Название	Значение
ОС	Microsoft Windows Server 2008 Enterprise Edition Service Pack 2 (build 6002), 64-bit
Оперативная память	12 882 354 176
Процессор	Intel(R) Core(TM) i7 CPU 920 @ 2.67GHz

Конфигурационный файл:

Название	Значение
NAME	H:\Indexes\IndexF\Index
BUFFER	1024*1024*1024
APPEND_BUFFER	1024*1024*64
CACHE_MULTIPLIER	15
GROUP_BUFFER_SIZE	1024*1024
DATA_DETAILS	WORD
REPOSITORY_ENABLE	0
REPOSITORY_COMPRESSION	Huffman
STORE_STOP_WORDS	0
STORE_STOP_WORDS_ONE_PER_FILE	1
MAKE_STOP_WORDS_INDEXES	0
FORMATS	rtf, pdf, djvu, msword
ARCHIVE_FORMATS	rar\rar.dll, cab\cab.dll, 7Z\7Z.dll
USE_TRANSFORMATION_SERVICE	1
KNOWN_WORDS.CLUSTER_SIZE	1024*16

UNKNOWN_WORDS.CLUSTER_SIZE	1024*4
UNKNOWN_WORDS.TREE_PAGE_SIZE	1024*4
SEQUENCE_REPOSITORY_ENABLE	0
SIMULAR_INDEX_ENABLE	0
SEQUENCE_BUFFER_SIZE	1024*1024*1024
SIMULAR_REQUIRED_PERCENT	0
TEMP_PATH	%INDEX_PATH%\Temp\

Результаты эксперимента:

Размер	Общее время, ч.:м.:с.	Время записи индекса	Чтение кластеров	Запись кластеров
85.23 Гб	03:33:44	01:34:19	6.50 Гб	57.15 Гб
1.41 Кб	00:00:59	00:00:54	4.44 Гб	4.46 Мб
1.08 Мб	00:01:54	00:01:49	4.44 Гб	322.72 Мб
2.17 Мб	00:02:14	00:02:10	4.44 Гб	461.57 Мб
3.09 Мб	00:02:33	00:02:29	4.44 Гб	583.10 Мб
4.03 Мб	00:02:43	00:02:38	4.44 Гб	681.53 Мб
5.03 Мб	00:02:51	00:02:46	4.44 Гб	746.05 Мб
6.25 Мб	00:03:00	00:02:55	4.45 Гб	822.55 Мб
7.31 Мб	00:03:09	00:03:04	4.44 Гб	866.07 Мб
8.21 Мб	00:03:12	00:03:07	4.44 Гб	890.28 Мб
9.19 Мб	00:03:20	00:03:14	4.44 Гб	957.13 Мб
10.02 Мб	00:03:24	00:03:18	4.45 Гб	998.44 Мб
10.02 Мб	00:03:20	00:03:16	4.44 Гб	996.63 Мб
20.03 Мб	00:03:55	00:03:50	4.45 Гб	1.25 Гб
30.37 Мб	00:04:12	00:04:06	4.45 Гб	1.40 Гб
40.19 Мб	00:04:39	00:04:32	4.46 Гб	1.53 Гб
50.60 Мб	00:04:53	00:04:45	4.46 Гб	1.60 Гб
60.17 Мб	00:04:59	00:04:51	4.47 Гб	1.64 Гб
70.26 Мб	00:05:04	00:04:56	4.47 Гб	1.68 Гб
80.46 Мб	00:05:17	00:05:07	4.46 Гб	1.71 Гб
90.01 Мб	00:05:26	00:05:16	4.47 Гб	1.74 Гб
100.09 Мб	00:05:29	00:05:18	4.48 Гб	1.78 Гб
200.00 Мб	00:07:01	00:06:30	4.55 Гб	2.02 Гб
300.00 Мб	00:07:54	00:07:12	4.61 Гб	2.15 Гб
400.44 Мб	00:10:06	00:09:14	4.68 Гб	2.34 Гб
500.13 Мб	00:10:39	00:09:37	4.74 Гб	2.54 Гб
600.01 Мб	00:11:11	00:10:02	4.77 Гб	2.55 Гб
700.90 Мб	00:11:51	00:10:35	4.87 Гб	2.75 Гб
800.62 Мб	00:12:09	00:10:45	4.87 Гб	2.79 Гб
900.24 Мб	00:12:37	00:11:07	4.94 Гб	2.88 Гб
1000.54 Мб	00:13:01	00:11:23	4.99 Гб	3.05 Гб
1.95 Гб	00:17:28	00:14:25	5.37 Гб	4.03 Гб

2.93 Гб	00:21:29	00:16:53	5.77 Гб	5.12 Гб
3.90 Гб	00:24:54	00:18:51	6.19 Гб	6.30 Гб
4.88 Гб	00:28:33	00:21:09	6.53 Гб	7.11 Гб
5.85 Гб	00:31:28	00:22:54	7.01 Гб	8.18 Гб
6.83 Гб	00:34:24	00:24:36	7.44 Гб	9.16 Гб
7.81 Гб	00:40:39	00:29:43	8.07 Гб	10.11 Гб
8.78 Гб	00:44:46	00:31:12	8.16 Гб	10.92 Гб
9.76 Гб	00:51:23	00:36:23	8.81 Гб	11.89 Гб

Время эксперимента: 2010.06.06 16:16:52

Информация о системе

Название	Значение
ОС	Microsoft Windows XP Professional Service Pack 2 (build 2600)
Оперативная память	1 054 355 456
Процессор	Intel(R) Celeron(R) CPU 430 @ 1.80GHz

Конфигурационный файл:

Название	Значение
NAME	E:\Indexes\IndexF\Index
BUFFER	512*1024*1024
APPEND_BUFFER	1024*1024*64
CACHE_MULTIPLIER	15
GROUP_BUFFER_SIZE	1024*1024
DATA_DETAILS	WORD
REPOSITORY_ENABLE	0
REPOSITORY_COMPRESSION	Huffman
STORE_STOP_WORDS	0
STORE_STOP_WORDS_ONE_PER_FILE	1
MAKE_STOP_WORDS_INDEXES	0
FORMATS	rtf, pdf, djvu, msword
ARCHIVE_FORMATS	rar\rar.dll, cab\cab.dll, 7Z\7Z.dll
USE_TRANSFORMATION_SERVICE	1
KNOWN_WORDS.CLUSTER_SIZE	1024*16
UNKNOWN_WORDS.CLUSTER_SIZE	1024*4
UNKNOWN_WORDS.TREE_PAGE_SIZE	1024*4
SEQUENCE_REPOSITORY_ENABLE	0
SIMULAR_INDEX_ENABLE	0
SEQUENCE_BUFFER_SIZE	1024*1024*1024
SIMULAR_REQUIRED_PERCENT	0
TEMP_PATH	%INDEX_PATH%\Temp\

Результаты эксперимента:

Размер	Общее время, ч.:м.:с.	Время записи индекса	Чтение кластеров	Запись кластеров
85.23 Гб	08:52:22	03:15:13	6.65 Гб	56.43 Гб
1.41 Кб	00:01:46	00:01:35	5.17 Гб	4.46 Мб
1.08 Мб	00:02:53	00:02:44	5.17 Гб	322.18 Мб
2.17 Мб	00:03:05	00:02:54	5.17 Гб	461.81 Мб
3.09 Мб	00:03:25	00:03:15	5.17 Гб	583.26 Мб
4.03 Мб	00:03:27	00:03:17	5.17 Гб	682.73 Мб
5.03 Мб	00:03:31	00:03:21	5.17 Гб	746.15 Мб
6.25 Мб	00:03:41	00:03:30	5.17 Гб	824.73 Мб
7.31 Мб	00:03:54	00:03:42	5.17 Гб	867.62 Мб
8.21 Мб	00:03:57	00:03:45	5.17 Гб	890.75 Мб
9.19 Мб	00:03:45	00:03:34	5.17 Гб	959.75 Мб
10.02 Мб	00:03:56	00:03:44	5.17 Гб	1001.07 Мб
10.02 Мб	00:04:00	00:03:48	5.17 Гб	1002.48 Мб
20.03 Мб	00:04:33	00:04:19	5.17 Гб	1.27 Гб
30.37 Мб	00:04:42	00:04:27	5.18 Гб	1.44 Гб
40.19 Мб	00:05:06	00:04:49	5.18 Гб	1.58 Гб
50.60 Мб	00:05:05	00:04:46	5.19 Гб	1.66 Гб
60.17 Мб	00:05:16	00:04:55	5.19 Гб	1.70 Гб
70.26 Мб	00:05:20	00:04:57	5.20 Гб	1.76 Гб
80.46 Мб	00:05:33	00:05:07	5.19 Гб	1.80 Гб
90.01 Мб	00:05:32	00:05:06	5.19 Гб	1.83 Гб
100.09 Мб	00:05:34	00:05:05	5.21 Гб	1.86 Гб
200.00 Мб	00:09:58	00:08:37	5.29 Гб	2.20 Гб
300.00 Мб	00:11:21	00:09:30	5.35 Гб	2.32 Гб
400.44 Мб	00:13:46	00:11:24	5.43 Гб	2.52 Гб
500.13 Мб	00:14:34	00:11:45	5.46 Гб	2.71 Гб
600.01 Мб	00:15:27	00:12:13	5.53 Гб	2.73 Гб
700.90 Мб	00:16:25	00:12:51	5.65 Гб	2.92 Гб
800.62 Мб	00:17:04	00:13:05	5.62 Гб	2.93 Гб
900.24 Мб	00:17:55	00:13:35	5.74 Гб	3.07 Гб
1000.54 Мб	00:18:31	00:13:48	5.80 Гб	3.22 Гб
1.95 Гб	00:25:40	00:17:00	6.18 Гб	4.19 Гб
2.93 Гб	00:32:57	00:19:39	6.61 Гб	5.31 Гб
3.90 Гб	00:39:48	00:22:13	7.14 Гб	6.51 Гб
4.88 Гб	00:46:21	00:24:43	7.56 Гб	7.37 Гб
5.85 Гб	00:52:38	00:27:03	8.10 Гб	8.36 Гб
6.83 Гб	00:58:32	00:29:10	8.44 Гб	9.23 Гб
7.81 Гб	01:04:50	00:31:40	8.99 Гб	10.23 Гб
8.78 Гб	01:10:44	00:33:43	9.05 Гб	11.04 Гб
9.76 Гб	01:19:11	00:38:01	9.73 Гб	12.02 Гб