



contentACCESS Analysis Tool Manual

OCTOBER 13, 2015 TECH-ARROW A.S. Kazanská 5, Bratislava 821 05, Slovakia, EU



Contents

Description of contentACCESS Analysis Tool	2
contentACCESS Analysis Tool requirements	2
Installing contentACCESS Analysis Tool	3
Structure of the Analysis Tool's user interface	6
Source folder and filtering criteria settings	7
Storage and archive speed testing settings	11
Analysis progress status pane	12
Detailed results of the analysis	13



Description of contentACCESS Analysis Tool

contentACCESS Analysis Tool is a software solution for analyzing file shares before archiving them with contentACCESS File Archive into a desired storage. It helps the user to plan the resources before starting the archiving process. The application calculates the size of all files to archive and it gives a time estimation for the user about the duration of the archiving process. It measures the performance of the file shares to be archived and also the speed of the desired target store. This helps to plan the storage capacities for the future, and prevent the user from accumulating of useless and old data, which ultimately would lead to a crisis of the file system. Using the filtering options of the tool, the user may analyze the documents according to age, size and file type. He may also specify the archiving methods (if shortcuts should be used to replace originals, if files will be compressed etc.) to estimate the archive time as precisely as possible. For estimation the user selects a representative number of files, and the tool will calculate the archive time according to the partial results obtained by analyzing this set of files.

The tool may be also used for collecting data of which analysis failed, with the goal of determining corrective actions before starting the archive process.

contentACCESS's Analysis Tool is the right software for users, who would like to plan the archiving process just-in-time.

contentACCESS Analysis Tool requirements

The following prerequisite has to be installed on the machine(s) where you would like to install the tool: **.NET Framework 4.5**

Another requirement is to run the tool under a system administrator user account.



Installing contentACCESS Analysis Tool

The tool can be downloaded from the TECH-ARROW's website. To download it, click here.

Download and run the setup package.



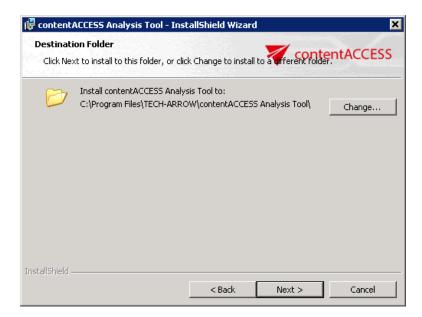
In the wizard's welcome window click on Next.

🙀 contentACCESS Analysis To	ool - InstallShield Wizard	×
~	Welcome to the InstallShield Wizard for contentACCESS Analysis Tool	
	The InstallShield(R) Wizard will install contentACCESS Analysis Tool on your computer. To continue, click Next.	;
	WARNING: This program is protected by copyright law and international treaties.	
	< Back Next > Cancel	

Read the software's license agreement and accept them by selecting the **I accept the terms in the license agreement** radio button.



Select the destination folder of the tool. By default it installs to C:\Program Files\TECH-ARROW\contentACCESS Analysis Tool, but you may change it with locating a custom folder with clicking on the **Change** button.



Now you are ready to install the software. Click on **Install** to start the installation process, and wait until it is finished.



🛱 contentACCESS Analysis Tool - InstallShield Wizard	🖶 contentACCESS Analysis Tool - InstallShield Wizard
Ready to Install the Program The wizard is ready to begin installation.	Installing contentACCESS Analysis Tool The program features you selected are being installed.
Click Install to begin the installation. If you want to review or change any of your installation settings, click Back. Click Cancel to exit the wizard.	Please wait while the InstallShield Wizard installs contentACCESS Analysis Tool. This may take several minutes. Status: Registering product
Instali5hield	Installshield

Click **Finish** to end the installation process.

🔂 contentACCESS Analysis To	ol - InstallShield Wizard	×
~	InstallShield Wizard Completed	
	The InstallShield Wizard has successfully installed contentACCESS Analysis Tool. Click Finish to exit the wizard.	
	< Back Finish Cancel	



Structure of the Analysis Tool's user interface

The user interface of the analysis tool is divided into the following main sections:

		Source folder and filtering criteria settings				Storage an	d archive speed testing settings	. 0
ler: * account (domain\user): * word: *	\\diskstation\Software\TestData\Enron_files tech\eba				-	 No archiving speed tests Archiving performance will not be ter Test archiving speed 		
not use mapped dhives, enter the full path	nsteed. If no account is entered, the current user will be used. The	enterned social releasing the second of the operating system publicage.	Age filter Size filter	Number of threads: 2 Refresh after evens: 5	* * * file(3)	estimated. Store path: \\tanews\DiskSti User account (domain\user): Password:	I be processed. The time needed for archiving the w oneTA\FileArchive\ tanews\administrator	Hole folder will be
			File type filter	Full refresh after every: 10	file(s)	Compress files	Process: 250	tile(s
peration running.		Yocessing time: 00:08:25.0648880				Create shortcut		Anahze
	natic folders and files Archiving statistics	hocessing time: 00.08.25.0648880	Progress status pane			Create shortcut		Analyze Save repor
		Nocessing time: 00:00:25.5643000 Processing	Progress status pane			Create shortcut		
iew Charts Statistics Proble Counters Number of folders: Number of files:	natic folders and files Archiving statistics 154 43233	Processing All time needed for processing Stop: 1. Time spent spacing the spacer charter of the statistics	00:08:25.0648880 00:00:05:6164928 00:07:57.32:31:38			Create shortout		

- Source folder and filtering criteria settings located at the left side of the tool's starting page
- Storage and archive speed testing settings located at the right side of the tool's starting page
- 3) **Analysis progress status pane** the progress can be started from here; it is located in the middle part of the tool's user interface
- 4) **Detailed results of the analyzes** divided into 5 main tabs:
 - Overview
 - Charts
 - Statistics
 - Problematic folders and files
 - Archiving statistics

The following sections of this documentation will provide the user with the information about the structure and usage of these individual user interface section.



Source folder and filtering criteria settings

	tACCESS				0 4
vlder: * ser account (domain\user): * issword: * Do not use mapped drives, enter the full pat itters	\\diskstation\Software\TestData\& tech\eba 	Enron_files will be used. The entered acount needs the 'Act as part of the Number of threads: 2 4 Refresh after every: 5 4 Full refresh after every: 10 4 file(s)	No archiving speed tests Archving performance will not be tests Test archiving speed A representative sample set of files will be estimated. Store path: \\tanews\DiskStor User account (domain\user): Password: Compress files Store metadata	e processed. The time needed for archi]
oparties supplies		Processing times 00.09/3E 06/0990	Create shortcut		
operation running. rview Charts Statistics Probl	matic folders and files Archiving sta	Processing time: 00:08:25.0648880	Create shortcut		alyze report

The user is required to specify the folder settings before he starts the analysis. For specifying a source folder execute the next steps:

1) Select the source folder, where the files to be archived are saved. Click into the Folder:* field and specify a desired folder, or locate the folder with the help of the triple dot (...) button:

Folder: *	
-----------	--

\\diskstation\Software\TestData\Enron_files

Note: Do not use the mapped drives, enter the full folder path instead.

2) In the next step the user is required to specify the domain user credentials, under which the analysis tool will connect to the source folder specified in step above. If no account is specified here, the currently logged on user logins will be used. The entered account needs the "Act as part of the operating system" privilege.

....



User account (domain\user): *	TECH\eba
Password: *	*****

* Do not use mapped drives, enter the full path instead. If no account is entered, the current user will be used. The entered acount needs the 'Act as part of the operating system' privilege.

Note: This privilege can be set in \Computer Configuration\Windows Settings\Security Settings\Local Policies\User Rights Assignment.

3) **Specify the files to archive with the help of the available filtering criteria**. These settings are optional. There are 3 filtering options available in contentACCESS Archive analysis tool: age filter; size filter and file type filter. To activate any of these filters check the checkbox next to it and then click on the filter button.

Filters	
Filters:	Age filter
	Size filter
	File type filter

✤ Age filter

To filter out items to analyze based on age, activate the age filter button and click on it to open the **Age filter** dialog. In case of age filter the modification date is determinative. It is possible to filter out the items

a) either by selecting specific dates from the calendar — in this case check the **Use Dates** checkbox to activate the calendar, and specify a date. The age filter set on Screenshot A will filter out all files of which modification date is older than 13rd of October, 2015.

b) or by specifying an exact number of days/months/years in the appropriate textbox — in this case leave the **Use Dates** checkbox unchecked and write the number of days/months/years into the appropriate textbox. The age filter set on Screenshot B will filter out files that are older than 1 year.



Y Age filter	X	Y Age filter	X
Filter items based on age Use Dates		Filter items based on age	
Younger than	Tuesday , October 13, 2015 - Days O Months @ Years	Younger than	Tuesday , October 13, 2015 • O Days O Months O Years
Older than	Tuesday , October 13, 2015 Cotober, 2015 Trs	Older than	Tuesday , October 13, 2015 - 1 O Days O Months O Years
From the time interval from	Sum Mon Tue Wed Thu Fri Sat 27 28 29 30 1 2 3 4 5 6 7 8 9 10 11 11 2 13 14 15 16 17 ys Months ⊚ Years 18 19 20 21 22 22 42	From the time interval from	Tuesday , October 13, 2015 ▼ till Tuesday , October 13, 2015 ▼ till ◯ Days ◯ Months ⓒ Years
	25 26 27 28 29 30 31 1 2 3 4 5 6 7 Today: 10/13/2015		OK Cancel

Screenshot A: Filtering items older than October 13, 2015

Screenshot B: Filtering items older than 1 year

✤ Size filter

To filter out items to analyze based on file size, activate the size filter button and click on it to open the **Size filter** dialog.

V Size filter	Number of Streams	1		
Filter items based on s	ize			
Smaller than	0	КВ		
Larger than	64	кв		
With size from	0	KB to	0	KB
			ОК	Cancel

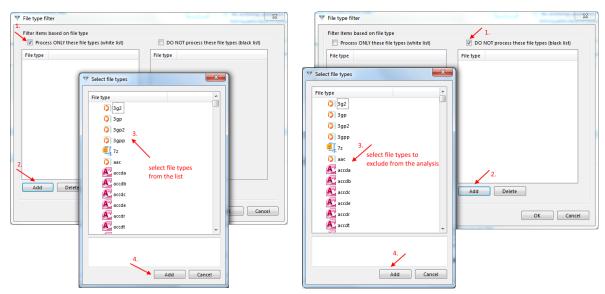
Specify a size that the size filter will use and click on **OK**. On the screenshot above the size filter will filter out all items larger than 65 KB.

✤ File type filter

This filter is used to

- a) filter out only certain file types to analyze in this case check the **Process only these file types (white list)** checkbox and add the file types to be analyzed. (Refer to Screenshot A.);
- b) or to exclude certain file types from the analysis process in this case check the DO NOT process these file types (black list) checkbox and add the file types to be excluded. (Refer to Screenshot B.)





Screenshot A

Screenshot B

Note: All file types will be automatically crawled, if no file types are specified in this filter.

The filters can be deactivated by unchecking the checkboxes of the particular filter buttons.

Note: It is possible to exclude from the analysis files already archived with the shortcutting method. This may be configured by excluding the LNK file type from processing (select it in the **File type filter**'s black list).

4) **Specify the number of threads and set the auto refresh intervals**. These values are by default, but the user is allowed to modify them if he prefers another settings.

Number of threads:	2	*	
Refresh after every:	5	▲ ▼	file(s)
Full refresh after every:	10	*	file(s)

Number of threads: the user may define here the number of simultaneously analyzed items; Refresh after every file(s): the Overview tab is refreshed in time intervals specified here; Full refresh after every file(s): information included in further analysis result tabs is refreshed in time intervals specified here.





Storage and archive speed testing settings

In the **Analysis Tool** the user may analyze the time frame needed to archive the contents of a folder that has been set in the folder settings. This configuration is optional. It helps to plan the time interval, when the archiving process will run in an organization. The time needed for archiving depends on the performance of the file shares to be archived, the number and size of the files included in the file shares and on the speed of the desired target store itself. The speed of the archiving process also depends on the further actions that are performed when analyzing the files (file compressing with the intention to save more storage space, metadata store or shortcut creation). Should the user test the archiving speed, he should configure the following settings:

					0
lder: * er account (domain\user): * ssword: *	\\diskstation\Software\TestData tech\eba	\Enron_files		 No archiving speed tests Archiving performance will not be tested. Test archiving speed 	
Xo not use mapped drives, enter the full pat	h instead. If no account is entered, the current us	er will be used. The entered acount needs th	he 'Act as part of the	A representative sample set of files will be processed. The time needed for archiving the whole folde estimated. Store path: \\tanews\DiskStoreTA\FileArchive\	er will be
lters	Age filter	Number of threads:	2	User account (domain\user): tanews\administrator	
	Size filter	Refresh after every:	5 🚔 file(s)	Password:	
	File type filter	Full refresh after every:	10 file(s)		file(s)
peration running.		Processing time:	00:08:25.0648880		nalyze
	ematic folders and files Archiving st		00:08:25.0648880		nalyze e report
Counters			Processing	A. Sav	
view Charts Statistics Probl	ematic folders and files Archiving st 154 43253 43253		Processing All time needed Steps: 1-Time spent 11. Time spent	A. Sav	
view Charts Statistics Probl Counters Number of folders: Number of files:	154 43253		Processing All time needed 1 Steps: I -Time spent II - Sample ar Archiving statis Number of proc Average size off Success / failure:	A for processing: 00:00:25.0648880 crawling the source: 00:00:08.6164928 t generating file statistics: 00:00:07.5233128 crawling the source: 00:00:01.90.500896 stics 250 files to archive: 246.34 kB	

 Should the user analyze the speed of the archiving process, he has to check the Test archiving tool radio button at the right side of the tool's start page. The time needed for archiving the source folder contents will be estimated – filtering criteria set on the source side will be considered as

well.

contentACCESS Analysis Tool | ALL RIGHTS RESERVED



0	No archiving speed tests Archiving performance will not be tested.
0) Test archiving speed
	representative sample set of files will be processed. The time needed for archiving the whole folder will be stimated.

2) Specify the store path, where the binaries will be stored. This can be either a local, or a remote share store. It is possible to locate it by using the triple dot (...) action button as well:

Store path: \\tanews\DiskStoreTA\FileArchive ...

3) Enter the user credentials, under which the store folder can be accessed:

User account (domain\user):	TANEWS\Administrator
Password:	*****

4) Specify the processing settings that will be used by the real archive job.

When archiving a file share with contentACCESS, the user may set, if the original files will be replaced by shortcuts or not, if the archived items will be compressed, if the extended metadata will be saved into the storage database. To simulate the real archiving time as far as possible, these values must be selected according to the planned archiving job settings in contentACCESS.

According to the screenshot below we would not like to compress files and store extended metadata when archiving, but will use shortcuts to retrieve the archived files easily:

- Compress files
- Create shortcut

5) **Specify a representative number of source files that will be analyzed.** For estimation the user selects a representative number of files, and the tool will calculate the archive time according to the partial results obtained by analyzing this set of files.

Process: 250 🚔 file(s)

Analysis progress status pane

The status pane showing the actual state of the analysis progress can be viewed in the middle of the tool's user interface. The analysis can be started manually from here, by clicking on the **Analyze** button.



The currently processed folder path can be seen beneath the progress line. The actual processing time can be seen in this pane, too. Once the tool finished the analysis, the progress report can be saved via **Save report** button in PDF format.

	tACCESS			0
Ider: * er account (domain\user): * ssword: * Joo not use mapped drives, enter the full path Iters		ser will be used. The entered acount needs the 'Act as part of the	No archiving speed tests Archiving performance will not be test Test archiving speed A representative sample set of files will estimated. Store path: [\\tanews\DiskSto User account (domain\user):	ed. De processed. The time needed for archiving the whole folder will be
	Age filter Size filter File type filter	Number of threads: 2 Refresh after every: 5 Full refresh after every: 10 file(s)	Password: Compress files Store metadata Create shortcut	Process: 250 Å
	ion\Software\TestData\Enron_files\ matic folders and files Archiving s			Save report
		Processing		
Counters Number of folders: Number of files: Number of files to archive: Sizes Size of all files:	154 0 0	All time needs Steps: I -Time spe II - Time sp	nt crawling the source: 00:0 ent generating file statistics: archiving done in: tistics	6:03.7012117)0:00.7970456

Detailed results of the analysis

The detailed results of the analysis can be viewed by clicking on the following tabs:

- Overview
- Charts
- Statistics
- Problematic folders and files
- Archiving statistics



The **Overview** tab gives an overview for the user about the:

- Total number of source files and folders, and the number of files to archive;
- Size of all files and size of files to be archived;
- Average file size;
- All time needed for processing of the selected set of files;
- Number of processed (representative) set of files;
- Average size of files to archive;
- Successfully processed / failed item count;
- Average time spent archiving 1 file;
- Estimated time needed for archiving the whole source folder. This value is calculated based on the results obtained by analyzing the archive time, size etc. of the representative set of files.

Charts tab:

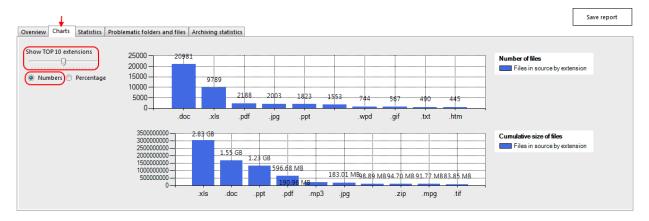
A simple chart in a tool says more than a sheet of full numbers. The **Charts** tab shows a graphical representation of the frequency of most common file types. It contains two charts:

contentACCESS Analysis Tool | ALL RIGHTS RESERVED

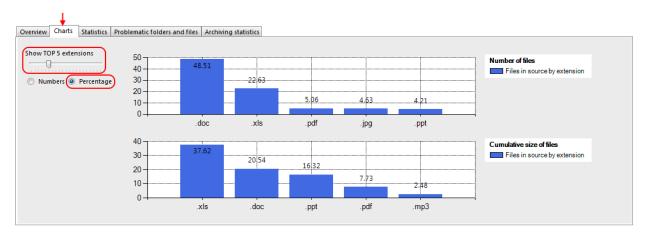


- The 1st chart shows the exact number of analyzed files based on file types. The user may also switch to "Percentage" mode — in this case the number of file types will be specified in percentages.
- The 2nd chart shows the cumulative size of analyzed file types (specified either in units or in percentages).

It is also possible to display the most recent (TOP 3, TOP 5, TOP 10 etc.) file extensions by moving the ruler in the **Show TOP [value] extensions** section.



Screenshot A: Number of files based on file types (10 most recent extensions selected)



Screenshot B: Number of files based on file types specified in percentages (5 most recent extensions selected)

The **Statistics** tab lists all discovered file types. You can see their amount, the cumulated and average size of the files, and their percentual share in the set of all files.

The items displayed in the **Statistics** table can be also sorted by all of these columns.

contentACCESS Analysis Tool | ALL RIGHTS RESERVED

TECH-ARROW, a. s. | Kazanská 5B, 821 06 Bratislava, Slovakia, EU | E-mail: sales@tech-arrow.com | www.tech-arrow.com



	_						
Dvervi	ew Charts Statistics Proble	ematic folders and files Archiving	g statistics				
Statis	stics						
	Extension		Count 🤜	, Count (%)	Cumulated size	Cumulated size (%)	Average size
Þ	.doc		20981	48.51	1.55 GB	20.54	77.32 kB
	.xls		9789	22.63	2.83 GB	37.62	303.55 kB
	.pdf		2188	5.06	596.68 MB	7.73	279.25 kB
	jpg		2003	4.63	183.01 MB	2.37	93.56 kB
	.ppt		1823	4.21	1.23 GB	16.32	707.08 kB
			1553	3.59	98.89 MB	1.28	65.21 kB
	.wpd		744	1.72	45.67 MB	0.59	62.86 kB
	.gif		567	1.31	8.16 MB	0.11	14.73 kB
	.txt		490	1.13	53.24 MB	0.69	111.26 kB
	.htm		445	1.03	7.54 MB	0.10	17.36 kB
	.rtf		275	0.64	30.99 MB	0.40	115.40 kB
	.exe		228	0.53	28.56 kB	0.00	128.26 bytes

Screenshot A: Statistical data of all analyzed files are displayed in the grid

Problematic folders and files tab

The failure analysis process of the tool relies on collecting failed data with the goal of determining corrective actions. If the analysis of certain data fails, the user may face with technical troubles when trying to archive the files in real, too. This tab contains the list of files or folders that could not be analyzed due to some kind of failures, or of which archiving may hit an obstacle. The problems can be related to insufficient permissions to the files, for example. The grid contains 3 columns:

- **Type**: can be either file or folder,
- Path: contains the full path of the problematical file or folder,
- **Error**: describes the error that occurred during the analysis/might occur during the archive process.

The items of the grid can be sorted by these 3 columns. **Before starting the archiving process, it is** highly recommended to examine the files/folder of which analysis failed.

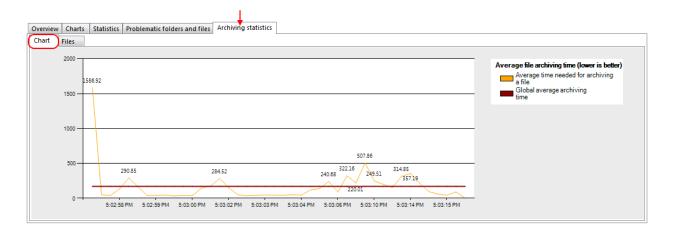
↓	Save report	
Overview Charts Statistics Problematic folders and files Archiving statistics		_
Type Path	Error	

Archiving statistics tab

The **Archiving statistics tab** is showing a real-time graph about the time needed for the archive analysis of the documents in the sample set. In this archiving statistics pane the user may select from 2 tabs:



1) **Chart**: This tab informs the user about the time needed for archiving. The yellow line displays the average time needed for archiving a file, while the red line displays the global average archiving time.



2) Files: This tab contains the times needed for archiving each file that has been selected. If the analysis tool is used very often, and a high number of files accumulate in the grid, the performance might be also effected. From this reason, it is highly recommended to clear the list after every 1000 items. This can be done automatically with checking the Clear list after every 1000 items checkbox at the right upper side of the grid.

\downarrow	Save report
verview Charts Statistics Problematic folders and files Archiving statistics	
hart (Files)	
UII Clear list after every 1000 items (more ite	ems might effect performance)
diskstation\Software\TestData\Enron files\full\arnold-j\Enron Wholesale UNIX - Storage.xls ; size: 37888 ; archiving time: 192.0109 ms	
diskstation\Software\TestData\Enron_files\full\allen-p\Phillip & Kieth Lender decision 6.19.01.doc ; size: 25600 ; archiving time: 112.0064 ms	
diskstation\Software\TestData\Enron_files\full\arnold-i\enron_33.exe ; size: 125 ; archiving time: 96.0054 ms	
diskstation\Software\TestData\Enron files\full\allen-p\Post offices.xls ; size: 20992 ; archiving time: 115.0065 ms	
diskstation\Software\TestData\Enron_files\full\arnold-j\fox-sports-nba-knicks[1].mov; size: 2100448 ; archiving time: 1413.0808 ms	
diskstation\Software\TestData\Enron_files\full\arnold-i\q051401.pdf; size: 47797; archiving time: 665.038 ms	
<u>diskstation\Software\TestData\Enron_files\full\arnold-i\q051501.pdf</u> ; size: 45602 ; archiving time: 288,0165 ms	[
diskstation\Software\TestData\Enron files/full\arnold-\Gas Swap Term Sheet2.doc ; size: 99840 ; archiving time: 659.0376 ms	
diskstation/Software/TestData/Enron files/full/amold-J/Guggenheim attendees.xis; size: 19968; archiving time: 382.0218 ms	
diskstation/Software/TestData/Enron files/full/arnold-J/Guqgenheim Opportunity.doc; size: 35328; archiving time: 338.0194 ms	
<u>diskstation\Software\TestData\Enron_files\full\allen-p\power_cap3.xls</u> ; size: 10747904 ; archiving time: 4880.2792 ms diskstation\Software\TestData\Enron_files\full\arnold-j\Homer.xls ; size: 773120 ; archiving time: 662.0378 ms	
diskstation/Software/TestData/Error files/ul/Jallen-p/Profit Staringu/s; size: 29184; archiving time: 23.0071 ms	
diskstation/Software/Festbala/Enron files/ul/ameri-pyrtoin_sharing/xis/size/size/wine/inter/inte	
diskstation/Software/FestData/Erron files/fullyalinger/mcGstOrtex/size/size/size/size/size/size/size/size	
diskstation/Software/TestData/Ernor files/fullmande/ip/rotoma.softri park.as, size: 100400, archiving time: 15:0051 ms	
diskstation/Software/TestData/Erron files/full/allen-p/Proforma.Waters At Bluff Springs.xk1 ; size: 165888 ; archiving time: 142.0081 ms	
disktation/Software/FestData/Ernon_files/full/arnoid-/invoice enron 9950.doc; size: 19968; archiving time: 102.0059 ms	
diskstation\Software\TestData\Enron files\full\allen-p\PROSPECTUS.doc; size: 144384 ; archiving time; 260.0149 ms	