Manual for
Retriever

Created 10-17-2021 T. Milius
Changed 12-01-2021 T. Milius

Content

Overview

Please read this manual carefully and in full. In case that you do not understand parts of it please desist from buying and using Retriever.

Note: The computer program Retriever operates only in conjunction with the computer operating system RISC OS! Operating this program requires a fundamental knowledge about operating of programs on RISC OS. These basics are not explained in this manual, they can be found in the RISC OS users guide.

Retriever is a program that helps you to find documents containing desired terms or fulfilling other criteria quickly.

To be able to do so Retriever must first analyse the documents it is to search which may take a while. Therefore Retriever needs parsers which are able to parse a certain type of document.

Warning: Due to the concept of Retriever there is always the risk that some changes of documents are not recognized and the data is therefore not properly synchronized. Additionally the content of some documents may not be analysed properly. You must not use Retriever in environments where a failure of the software could have deadly or negative consequences on health or negative consequences on data, finances, machines etc.!

Installation

Requirements

Retriever requires RISC OS 5.2X which includes MD5 checksum calculation inside !Internet. The program has been tested on an IYONIX pc, a BeagleBoard xM, a Raspberry Pi B, a Raspberry Pi 2B and a Raspberry Pi 3B+ but should work on other RISC OS computers running RISC OS 5.2x or newer too.

Additionally, a web browser must be available on the machine to show the search results. Any web browser should be able to cope with the HTML result files but as Retriever is internally making use of UTF-8 encoding it is recommended to use a web browser that can cope with UTF-8 encoded HTML-documents to display certain special characters correctly.

Performing the installation

Copy the !Retriever application to an arbitrary place on a storage medium. Now Retriever should be ready to use. However, you still have to configure it.

For removal delete the !Retriever application from the storage medium. If you stored your Retriever database elsewhere you have to remove the files separately. Additionally, you may delete the Retriever directory containing the configuration file inside the Choices: directory.

Regarding the location of the Retriever database please take into account that Retriever performs lots of writes to the database during analysis. Storage media based on flash memory technology like SD cards dislike much write accesses pretty much. It is thus recommended to locate the Retriever database on a conventional hard disc or to setup a copy of the database on RAM during analysis and copy it back to its original place afterwards.

Manual

Start !Retriever by a select double click on the !Retriever application. On the right side of the icon bar the symbol of the program will appear.

With a menu click over the Retriever symbol on the icon bar you will obtain a menu which allows you to display the copyright (menu entry "Info"), this help (menu entry "Help"), some configuration files (menu entry "Configuration") and to stop (menu entry "Quit") the program.
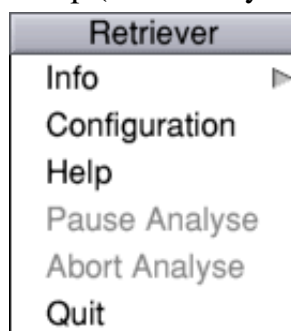
In case that an analysis is running additionally you can pause it (menu entry "Pause Analyse") and continue it afterwards (also menu entry "Pause Analyse") or try to abort the analysis (menu entry "Abort Analyse"). Whilst pausing will take place quite quickly an abort of an analysis may take a long time as Retriever tries to save the results of the analysis which have been obtained until the point of time at which the abort request occurred. A ticked menu entry "Pause Analyse" is indicating that the analyse is paused and a ticked menu entry "Abort Analyse" is the indication that an abort of the analyse has been requested. You can pause an analyse whilst an abort has been requested but in such a case the abort won't be finished until you are continuing the analyse.

When the program is quit all related windows will be closed. **Note: If an analysis is running Retriever will first try to abort this before quitting. This may take a while.**

General concept

To be able to locate documents quickly Retriever needs to analyse the documents first. The kind of a document is determined by its type (this means RISC OS file type). For each type of document a

certain so called parser is required which parses the document, extracts parts out of it and passes these parts which are called elements to one or more so called archivers which note the occurrence of such an element for a document inside an internal database. Elements are of a certain type. Element types are Retriever internal types. Every parser can handle one or more RISC OS file types and every archiver can cope with one or more element types. Parsers are allowed to extract documents out of documents or to convert them. These subdocuments are also analysed by Retriever in case that an according parser is available. The content of the subdocuments is logically assigned to the document out of which it was extracted.

After documents have been analysed you are able to search for the documents by entering search criteria. All entered criteria must be fulfilled by a document so that it is shown inside the result list. There is a separate input mask for every element type to enter the desired criteria. The desired criteria are passed internally to the archivers which are trying to find the related documents. The result list is a HTML-document which shows a list of the documents fulfilling all of the requested criteria.

The archivers of Retriever are storing their information inside a directory. The path of this directory is called the database path. Inside this directory there are subdirectories which are called groups. Analysis is done using a certain group to store the results. Searching for documents can be done on one or several groups. This allows you to optimize the analysis and a preselection of the results.

Retriever can only either analyse or search at the same time. It is not possible to analyse documents and search for documents in parallel.

Retriever internally distinguishes documents by an internal ID which is unique for every group. It also assigns a reference to every ID which allows you to identify the document during a search. This reference is often but not always the absolute path to the file containing the document on the disc. In case that a file is moved or renamed Retriever won't get aware of this and will show you an invalid reference as search result. Also, if documents are changed Retriever won't get aware of this and will show you an invalid reference as search result. However, in case that you are reanalysing the changed document Retriever will detect that the document has been changed, mark old database entries as deleted and create new entries. **However, the detection of a document change may fail due to Retrievers internal concept of detecting changes by checksums and checksums are not unique for arbitrary contents even it is quite unlikely that two different contents for the same document will generate the same checksum.** In doubt delete the whole group (by deleting all of its files) and reanalyse all related documents.

In the moment Retriever knows the following element types:

- Term
- Document data
    - Checksum (MD5)
    - Size of document
    - Date of the last change (GMT)
- Email header fields
- File (internal usage to handle subdocuments)

Analysis of documents

Starting the analysis of one or more documents is quite simple. Simply drag the files or folders containing the files onto the icon bar symbol of Retriever.

Afterwards the analysis of the files will start and the symbol of Retriever on the icon bar will change:

Whenever the analysis of a document/folder you dragged onto the Retriever icon bar symbol has been finished you will be informed. After all documents have been analysed the symbol of Retriever on the icon bar will switch back to its usual shape.

Whilst analysis is running you are able to drag additional files on the Retriever icon bar symbol which will be queued for analysis. However any attempt to start a search or an analysis configuration while an analysis is running will result in an error message.

Killing retriever during a running analyse will lead to a corrupted database which means that you have to delete all of the respective group files and analyse all related files again. You have the possibility to abort an analyse. This should leave the database in good order even not all parts of the document which is actually analysed may be analysed.

As documents may contain embedded documents which may be also analysed during the analysis process it is wise to activate image filing systems before starting the analysis. Eg. a zip file attachment inside an email will be handled as an unknown document if there is no matching image file handler available. But in case that an according image file handler is available Retriever will try to examine the documents contained inside it.

Analysis may take a long time. Retriever tries to multitask but this is limited.

The following parsers exist inside Retriever:

Parser	RISC OS File Types	Elements	Remarks
Retriever internal	Any real file	Document data: Checksum Document data: Size of document Document data: Date of the last change (GMT)	
Text et al. -> Term	&FFF - Text &FFE - Command &FEB - Obey &FD7 - TaskObey	Term	Note that the listed RISC OS files types are the default. You can add or exclude file types.
Directory -> File	&1000 - Directory	File	

HTML -> Term, Files	&FAF - HTML	Term	
Email -> File, Header	&B86	File Email header field: To (Default, may be disabled) Email header field: From (Default, may be disabled) Email header field: Subject (Default, may be disabled) Email header field: Arbitrary fields as defined	Files are getting the timestamp of the email header field "Date". As they are analysed immediately afterwards this information will find its way into the document data date. Type &B86 is used for single emails but also for a list of emails. A list of emails is recognized by the presence of "From ", Character value 1 (CTRL-A), "! rmail " or "#! rnews " in the first line of the file. In case that the file contains a single email without this tag the reference will be a file reference. In case of lists the message ID of the email will be the reference. The body parts of the emails are treated as separate files.
File -> Converted file	&FFB - BBC BASIC &FCA - Squash &ADF - PDF (with !PDFUtils installed) &B8B - TextEase	File &FFB - BBC BASIC -> Text &FCA - Squash -> unsquashed file &ADF - PDF -> Text (UTF-8) &B8B - TextEase -> Text	Note that the listed RISC OS files types are the default. You can add or exclude file types.

The following archivers exist inside Retriever:

| Archiver	Element Types	Remarks
Document data	Checksum Size of document Date of the last change (GMT)	
Terms	Term	Terms may be optionally standardized.
Email header	Email header field	Fields may be defined

Configuration of the analysis

You are able to configure some parts of the analysis. This includes general settings, settings for parsers and setting for archivers. Some of the options can be stored permanently other are only valid during the next analysis.

There is a general analysis configuration window which contains a section where the parsers and

archivers can be configured.

General

With an adjust click on the Retriever symbol on the icon bar you are opening the analysis configuration window.
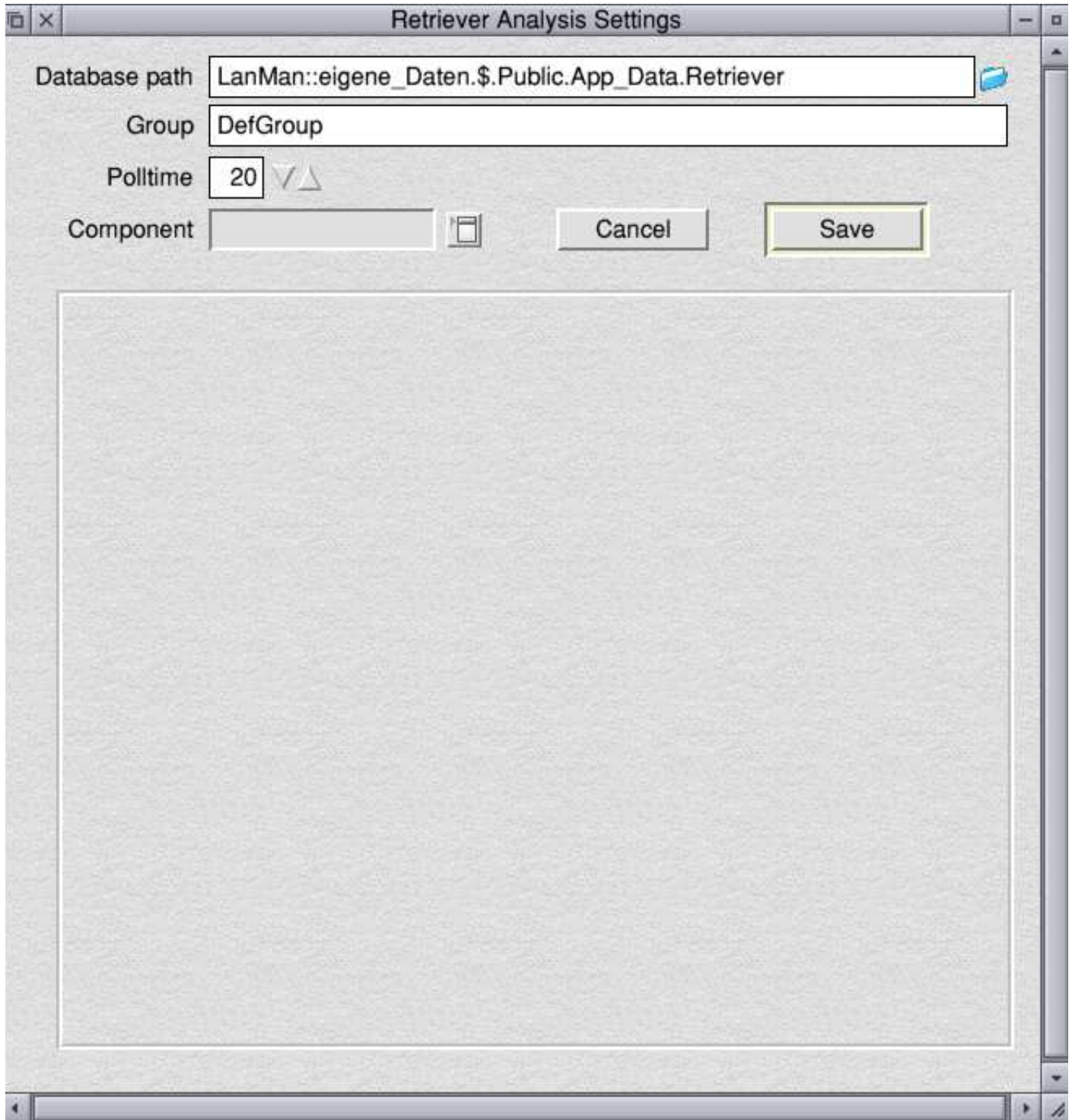
In the "Database path" field you set the location of the Retriever database on your disc. You can drag an according folder onto the field. You are able to create several databases in parallel even though Retriever can handle only one of them at the same time.

Clicking on the folder symbol will open the directory belonging to the "Database path" field in the Filer so that you can create group folders in it.

In the "Group" field you are defining to which group the data of the next analysis will be written. You can drag an according folder onto the field.

"Polltime" determines the span of time after which Retriever will try to poll. However it there is no warranty that Retriever will really poll after this span of time.

All of the settings above are stored inside Choices:Retriever.config_ana when you click on the "Save" button.

Clicking on the "Cancel" button will close the window without saving anything but the settings will be still active.

With "Component" you can select sub windows allowing you to configure certain components in more details. The windows are displayed inside the frame below "Component".

Parser Text et al.
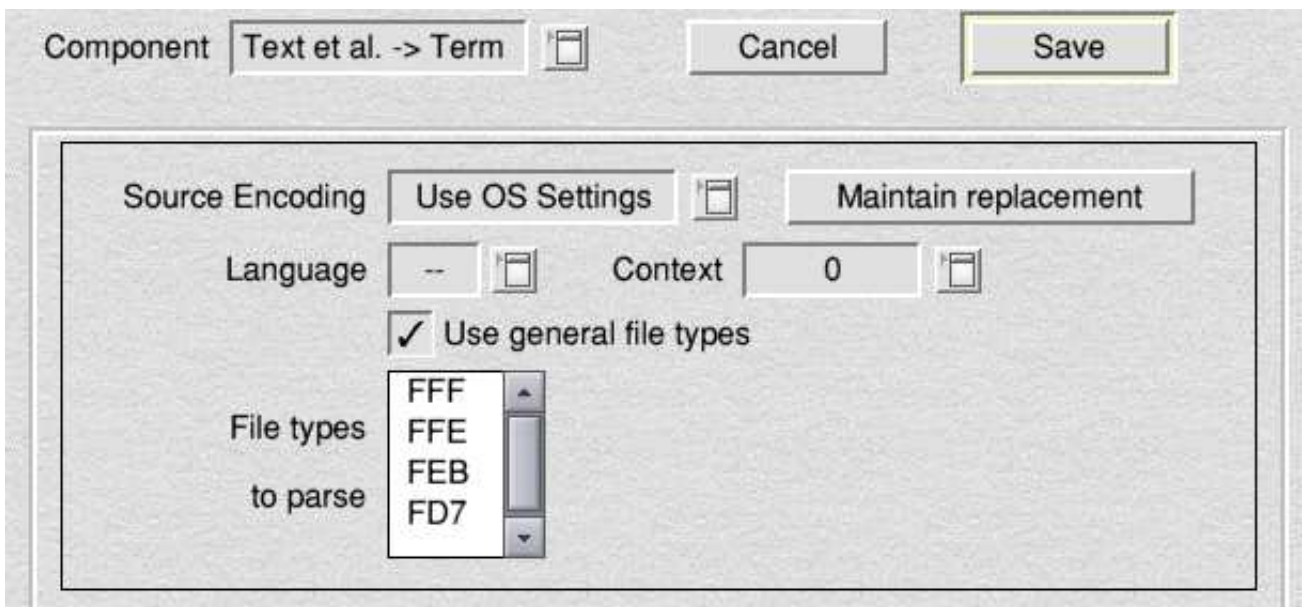
Simple text documents are in most cases not containing any hint about their encoding that means which computer alphabet is used to display their content correctly. This relates to special characters. You can tell Retriever which alphabet shall used for the next analysis by setting this inside "Source Encoding". **In case that texts are analysed using the wrong alphabet you won't be able to search for words containing special characters properly.**

As most texts on your disc will be self-written it is likely that they are encoded using the alphabet configured for RISC OS. So, this is the default setting.

Parsing of texts which are UTF-16 or UTF-32 encoded isn't supported in the moment.

"Maintain replacement" button, "Language" and "Context" are explained below inside the chapter Replacements. None of these settings will be stored and will have only an effect on the next analyse.

The text parser may be used for various files. However you may be interested only in analyse of files of a certain RISC OS file type. There is a list of RISC OS file types which are handled in general by the text parser. You can change this list. This is described inside the chapter Parser Text General RISC OS File Types. For a certain group you are allowed to define a subset of this list and store it within the group. To change the list you first need to deselect "Use general file types". Afterwards you can drag and drop files of the desired RISC OS file type onto the list with file types. In case that the file type is already in the list it will be removed otherwise it is added to this list but only if it is in the general list.

Parser HTML
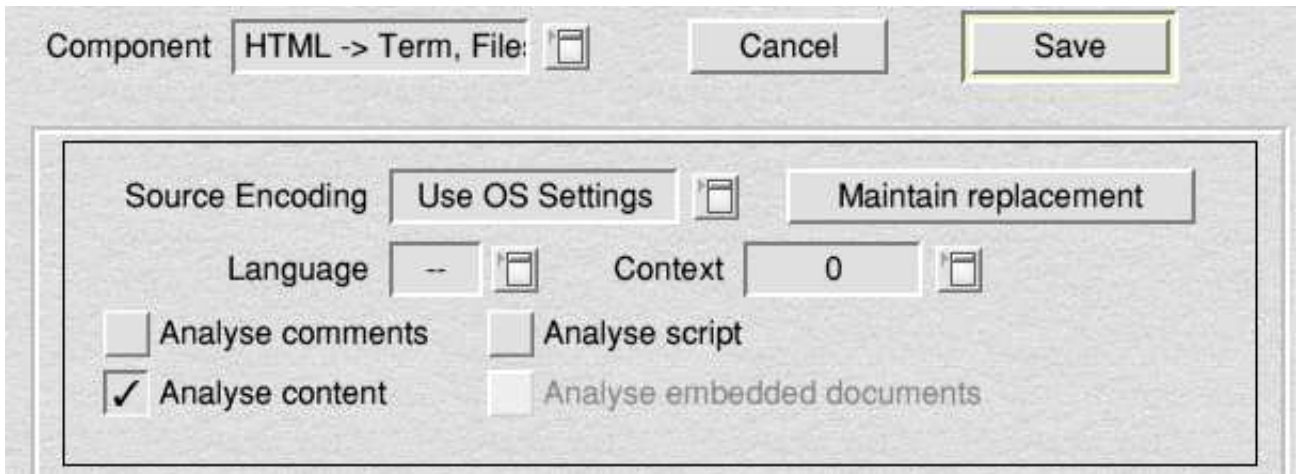
In contrast to text documents HTML documents _may_ contain information about their encoding. However, this is not mandatory and the HTML parser is not evaluating this information in the moment. So, the field "Source Encoding" has been provided for HMTL documents too. For more information on this see chapter Parser Text etc. above. In general this is not so important as HTML offers its own special character encoding which is handled properly.

You can determine which parts of a HTML document will be analysed. By default this is only the usual content. But e.g. in case that you want to analyse your self-written web pages containing script to be able to find certain terms quickly enable the according options.

"Maintain replacement" button, "Language" and "Context" are explained below inside the chapter Replacements. None of the settings will be stored and will have only an effect on the next analyse.

Parser Email
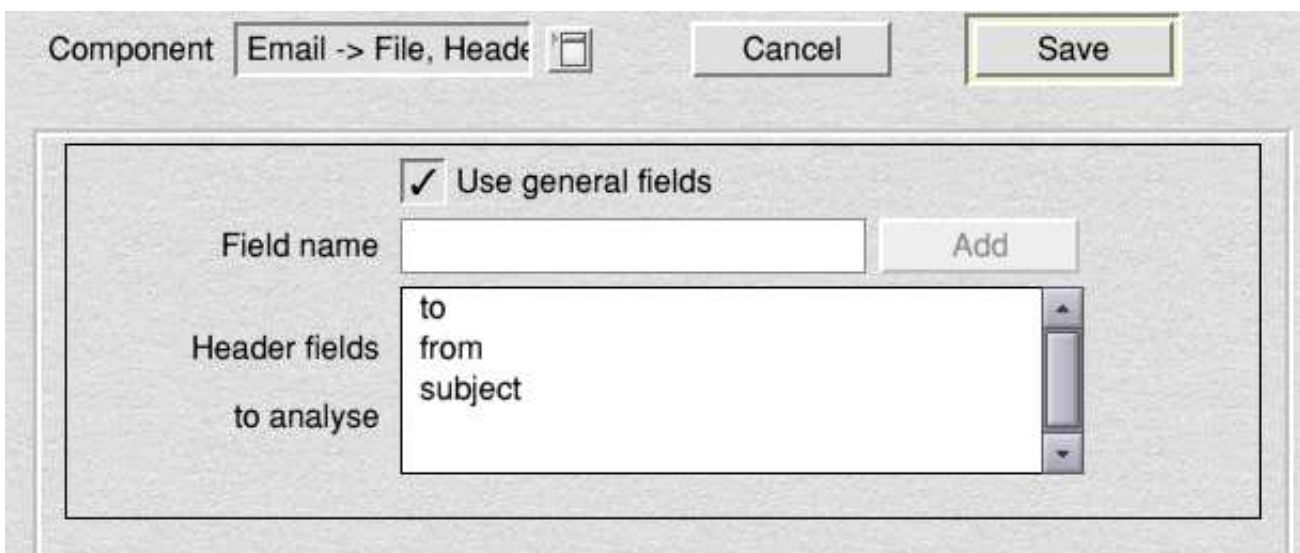
If "Use general fields" is selected a couple of default fields will be analysed. You can change this default list. This is described inside the chapter Parser Email General Header Fields. For a certain group you are allowed to change the list and store it within the group. Deselect "Use general fields" to modify the list of fields. Fields can be be added by entering their name in lower case letters into "Field name" and clicking on the "Add" button afterwards. A field can be removed from the list by

selecting it inside the list and performing a Menu click on the sub window afterwards.
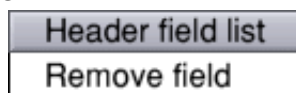
Selecting "Remove field" will remove the selected field from the list.

Converter Parser
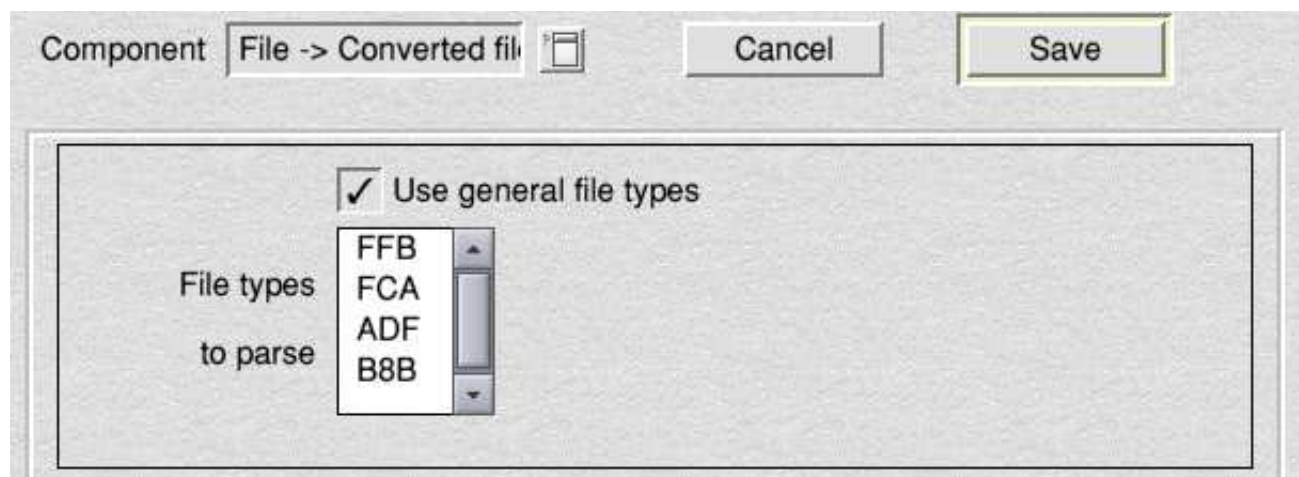
The converter parser converts files of certain RISC OS file types into temporary subdocuments of a RISC OS file type which can be analysed by Retriever. When such a file is encountered Retriever stops analyse and tells a converter program to convert the file. After this has been done Retriever continues its analyse. Some conversion are requiring programs which are not part of Retriever. They must have been installed and "seen" before starting an analyse making usage of them.

You can add your own converters to Retriever. This is described in the chapter Converter Parser inside the section Technical details.

The converter parser may be used for various files. However you may be interested only in conversion of files of a certain RISC OS file type. There is a list of RISC OS file types which are handled in general by the converter parser. You can change this list. This is described inside the chapter Converter Parser General RISC OS File Types. For a certain group you are allowed to define a subset of this list and store it within the group. To change the list you first need to deselect "Use general file types". Afterwards you can drag and drop files of the desired RISC OS file type onto the list with file types. In case that the file type is already in the list it will be removed otherwise it is added to this list but only if it is in the general list.

Archiver term
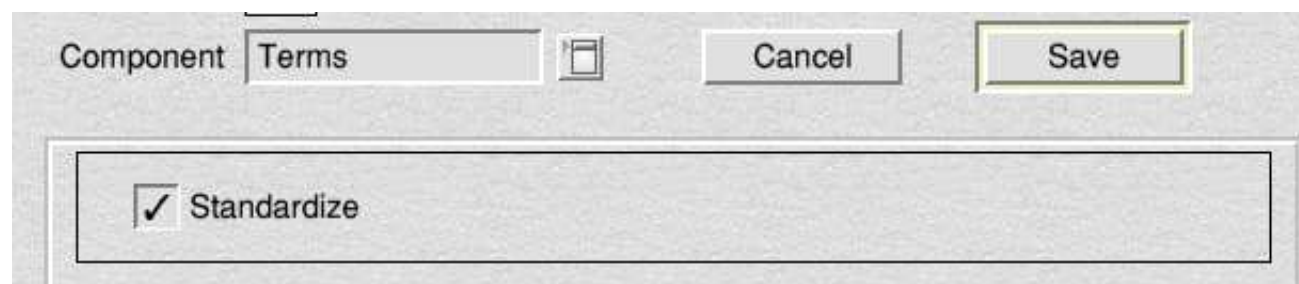
The term archiver provides a single option "Standardize" which can be activated. Once activated you

should save this setting within the respective group and leave it activated. "Standardize" will convert all terms to lower case letters and will replace German "Umlauts" in the following way:

- Ä-> ae
- Ö-> oe
- Ü-> ue
- ä-> ae
- ö-> oe
- ü-> ue
- ß-> ss

This avoids trouble during search in case that you can't be sure in which way a term has been written inside a document. E.g. terms like "the" can be located at beginning of a sentence and therefore are written there with a capital letter at beginning.

Internally the term archiver stores the original term from the document in addition to the standardized variant and you can search for it by using a special character if desired.

Replacements

Replacements are a general concept which are used optionally by some parsers which are generating elements of "Term" type. In the moment these are the Text parser and the HTML parser but not the Email one for header fields.

The concept behind replacement is to unify some terms. At certain areas of live people are using certain special abbreviations like for the name for a certain range of law. Unfortunately, they do not always use the abbreviation but sometimes the full name instead. Even worse the same abbreviations have an entirely different meaning in a different context. E.g. in information technology "OS" is nearly always used as abbreviation for "Operating System". Some months ago, in a German juristic business text I read something about an "OS" clause. In fact, the text referred to the "Online Settlement" of the EU :-(. This complicates the search a lot. Also in common life sometimes you write "8" and in other cases "eight".

So inside Retriever you can set up a table of terms which shall be replaced under certain circumstances by another term or a couple of terms. In general, the related window will be opened by clicking on the "Maintain replacement" button.
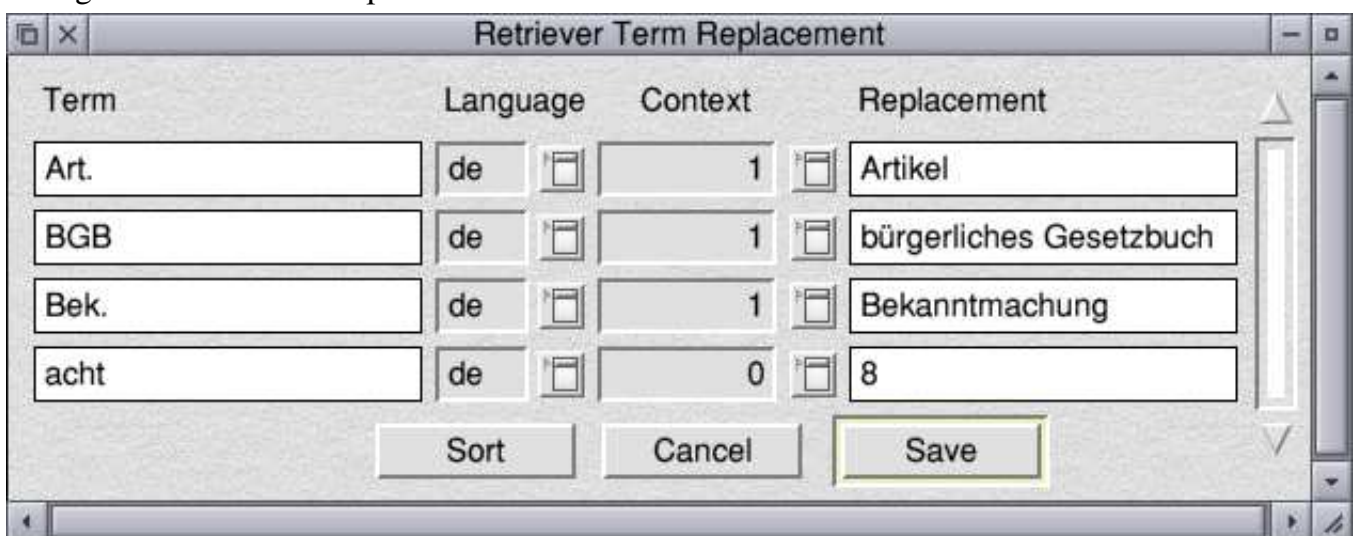

"Term" is the term which shall be replaced if it occurrs inside a text and "Replacement" defines the

Manual Retriever

list of terms (separated by a space) by which the term shall be replaced in the archivers. You can define additional circumstances under which the term shall be replaced. These are the languages
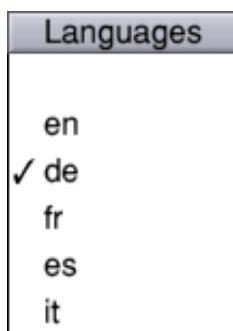
and the context. In case that no context has been selected this means that the replacement will take place independently of the context.
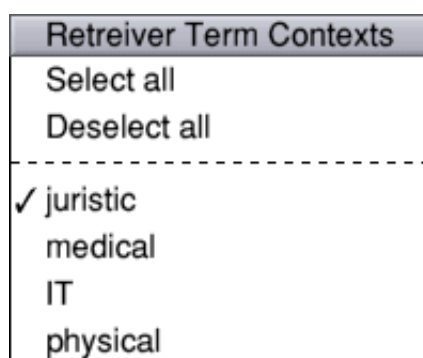
The same will happen in case that there is no language defined. To distinguish "any language" and "disable replacements" the language menu of the parsers is offering the additional option "--", which means "disable replacements",
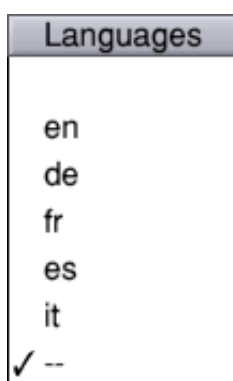
You can scroll through the list by making use of the slider or clicking on the arrows.

You can delete certain replacements by blanking the respective "Term" field. In case of modifying, deleting or adding terms these will be not sorted automatically. You have to enforce this manually by clicking on the "Sort" button.

In case that you are changing the RISC OS alphabet settings terms may have been defined in the list containing characters which can't be displayed properly by the current alphabet. This characters are displayed as "—" and terms containing them can't be updated.

Note that you need to save your entries inside the replacement table by clicking on the "Save" button before leaving the window or all changes will be lost, The table data itself is stored inside Choices:Retriever.ReplcTerms. You can drop your changes by clicking on the "Cancel" button.

Of course, Retriever doesn't know e.g. that you want to analyse a couple of texts regarding German law the in next go. But you know (or so I hope) and can configure Text and HTML parser

accordingly before starting the analysis. After doing so the parsers will try to look up every term inside the replacement table regarding the configured conditions like language and context.

Search for documents

After you let your documents be analysed, you can search for documents which are matching given criteria. Whilst the analysis tries to multitask search doesn't do so as the search is fast.
With a select click on the Retriever symbol on the icon bar you open the search window.
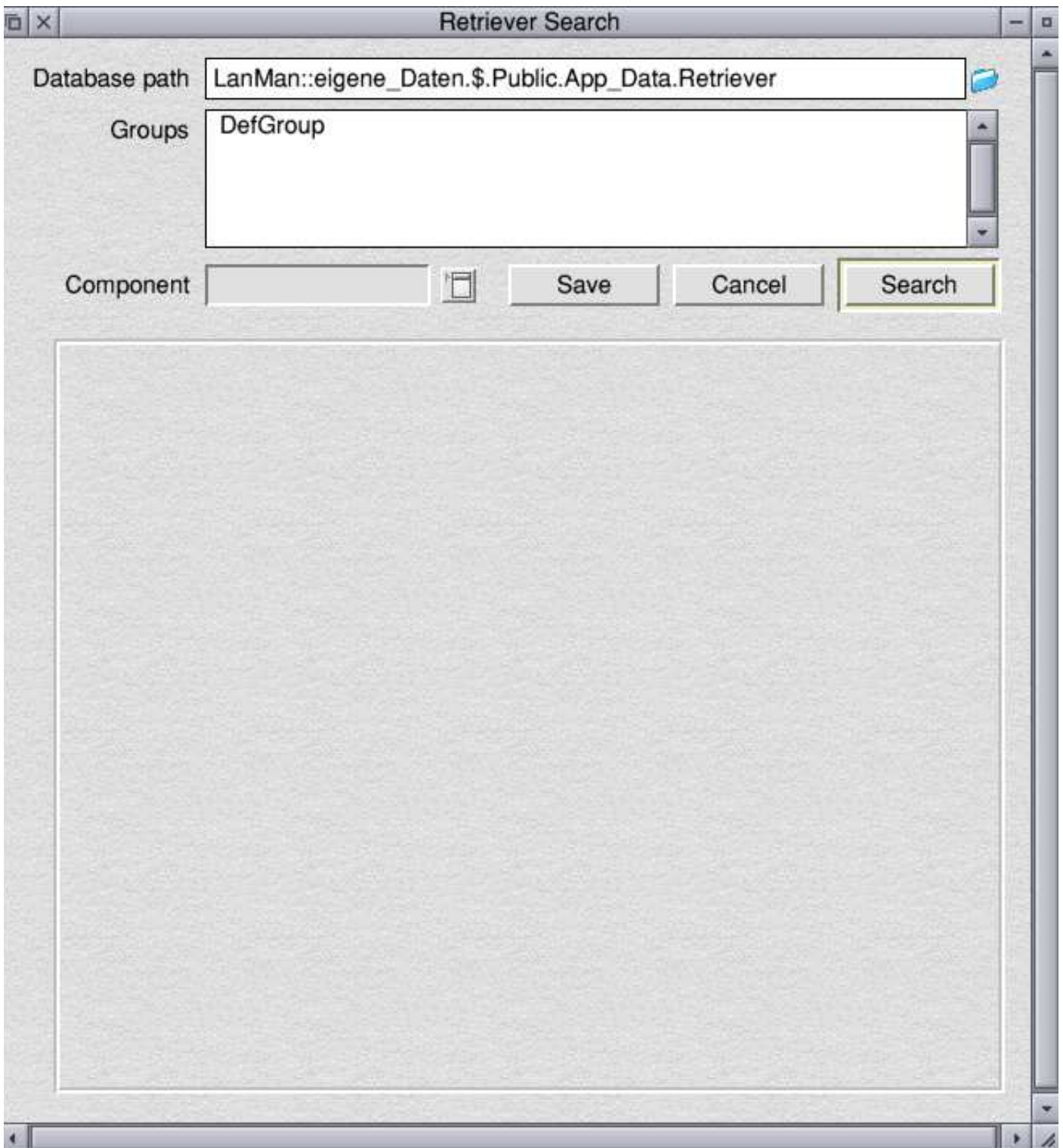
"Database path" tells Retriever where the internal data used for searching is located. This may differ from the same field inside the analysis configuration. However, the search field variant is not saved and at start filled with the value of its analysis configuration pendant. You can drag a folder onto the field.

Clicking on the folder symbol will open the directory belonging to the "Database path" field in the Filer. You will needs this to setup the groups.

The "Groups" list tells Retriever the groups where it has to search for documents. You can define up to 256 groups for searching. The list itself is changed by dragging group folders from the "Database path" directory onto the "Groups" field. In case that the group is already in the list it will be removed. If it is not in the list it will be added to the list.

The group list is stored inside Choices:Retriever.config_srch when you click on "Save" button. No other search settings are stored in the moment.

The search criteria can be entered in sub windows shown in the frame of the search window. One sub window can be shown at the same time but the search criteria entered are kept internally. The sub windows are selected by choosing one from the "Component" area.

After entering all desired criteria the search must be started by clicking on the "Search" button or pressing the "Enter/Return" key.

The result will be presented as a list inside an HTML file which will be displayed inside a web browser e.g. like the following

...

- Reference 123456789ab.Verw@Verwaltung.stader-softwareentwicklung-gmbh.com.de
  Group DefGroup
  ID51

  1. VAT
     Occurrence 2
     Occurrence Base 13674
     Classification 0
     Files 6
     Occurrence in Files 22
     Occurrence base in files 82044
     Variance 0.000000

...

The following search sub windows exist inside Retriever:

Window	Element Types	Remarks
Document Data	Checksum Size of document Date of the last change (GMT)	Search ranges for size and date
Email Header	Email header field	Wildcarded search possible
Term	Term	Wildcarded search possible

Searching document data

You may enter a "Checksum". This is an MD5 32 character long hexadecimal sequence of lower-case letters and digits. Much easier is to simply drop a file of your choice for which you want to search onto the field. Retriever will calculate the checksum for you and fill size and date too. Note that checksums are not unique so if a document is shown in the result file matching the checksum this is not a warranty that it is exactly the same file. Why searching for checksums? This may tell you whether the document is located at different places on your disc or the email with which you obtained the document.

To search for "Size" of a document you have to check it and you must enter a minimal and a maximal value. So you are able to search for ranges.

To search for "Date GMT" of a document you have to check it and you must enter a minimal and a maximal value. The values must have the form Month(2 digits)-Day(2 digits)-Year(2 or 4 digits) optionally separated by a space followed by Hour(2 digits):Minute(2 digits):Second(2 digits). In case that the time is missing it is internally added 00:00:00 for the minimal date and 23:59:59 for the maximal date. Date is the GMT date of the last modification of a document. In emails this is the content of the header field Date:.

Searching email data

Select the desired email header field from the drop-down list or enter it manually into to the "Field" and enter your search terms into the "Search terms" field. Field is only used for search when "Search terms" is not empty. Up to two fields can searched in the same go (logical AND). Search terms may contain wildcards. These are explained inside the chapter Searching terms in more details.
Important: Contents of email header fields are archived as they are. No standardization will

take place. So you must enter the desired search term as-is e.g. by taking capital letter into account. Consequently, there is no " option as inside regular terms.

Note that you may get an error message in case you are searching a group for which no emails have been analysed. This is only a hint for you not to include the group into the search but it doesn't impact the result.

Searching terms

Searching documents by terms is the main purpose of Retriever. Simply enter the desired terms into the "Search terms" field. If it is left blank Retriever won't search for terms. The terms are separated by spaces. A document matches your requirements if all terms you entered are occurring in the document. However, you can use some additions to improve your search. You may remember that terms may have been standardized during analysis. The default is to search for the standardized variant. So the terms you enter for the search will be also internally standardized. However, this is unwanted under certain circumstances. E.g. a name may occur inside a couple of documents in some cases in capital letters to show its importance. If you want to find only the documents where the name occurs in capital letters you need to type it in capital letters into the search terms preceding it with a ". You may also want to ensure that a certain term is not occurring inside a document. Precede it with a - in such cases.

You are also able to use wildcards inside terms. * will match any sequence and \ a single arbitrary character. As \ is used to indicate wildcards you have to enter \\ in case you want to search for \ itself. So \east should match beast and least and dog* should match dog, dogs and dogfight. There are internal limitations so you may get an error in case that a wildcarded term matches too often.

Take also into account that terms might have been replaced during analysis and so you may not find them under the term you may expect. To avoid looking into the replacement table simply set "Language" and "Content" as desired and described inside the chapter Replacements before. Enter your terms and before starting the search click on the "Adapt search term" button. Retriever will scan your list of search terms and try to replace every term which is defined inside the replacement list. After doing so you can start your search. So you can use abbreviations etc. inside search terms.

Hints

Emails

To prepare analysis of emails export the emails first from your email program into temporary file.

Use "From ", "CTRL-A", "#! rmail" or "#! rnews" format. Ensure that the file has the RISC OS type &B86. Drag the file onto the Retriever symbol on the icon bar.

Finding emails in a search will not create a HTML file URI as usual documents will do. This is because emails are not stored (necessarily) inside a usual filing system but inside the email program. Therefore emails are identified by their message-id. Retriever is generating a message: URI for emails. Unfortunately I know of no RISC OS email program handling such URIs. So locate your pointer over the URI inside the result file in the web browser. In general web browsers allow you to save an URI reference using the menu. Save the URI reference as plain text. Remove "message:" at the beginning and put the remaining message-id into the search of your email program. If possible restrict search to "Message-ID". For Messenger a small BBC BASIC program has been added to Retriever which tries to perform this automatically in case that you are clicking on a message: URI. The program is emulating the according keystrokes and therefore may not work with all versions of Messenger.

Right of use

©Copyright 2021. All rights reserved.

The program Retriever including this manual is intellectual property of the

Stader Softwareentwicklung GmbH, Germany ([Contact](#))

named as manufacturer below. In case that you bought and got this software directly delivered from the

Stader Softwareentwicklung GmbH, Germany ([Contact](#))

this is named below as supplier.

These terms apply to

- the program Retriever
- updates of the program supplied by the manufacturer
- additions to the program supplied by the manufacturer

By the usage of this software you acknowledge these terms. In case that you disagree with these terms you are not allowed to use the software.

If you comply with the terms of license depending on the type of license you dispose of the following rights.

License Types

Test License

A test license has the same right of use as a single machine license but you are only allowed to use it within the 14 days after the granting of the license. Afterwards all installations of the program must be removed and the program must not be used any longer.

Single Machine License

A single machine license must not be used in the scope of cloud computing.

As private person you are allowed to install and use this software on exactly one RISC OS based computer which is owned by you, your spouse or your children below the age of 18 in the country

where you bought this license.

As company you are allowed to install and use this software on exactly one RISC OS based computer which is owned by company in the country where you bought this license.

Scope of the license

The software is licensed not sold. The act of sale grants only some rights to use the software to you. The manufacturer reserves itself all other rights. You are only allowed to use the software in such a way as explicitly permitted by type and terms of license unless the applicable law grants more comprehensive rights to you regardless of these limitations. You are obliged to comply with all technical limitations of the software which permit you only special usages. In case that the law is not explicitly permitting something different it is illegal:

- to circumvent technical restrictions of the software
- to reverse engineer, to decompile or to disassemble the software
- to publish the software so that others are able to copy it
- to rent, loan or distribute the software

You must not use Retriever in environments where a failure of the software could have deadly or negative consequences on health or negative consequences on data, finances, machines etc. especially in scope of so-called "high risk activities"! You resign explicitly any explicit or implicit warranty regarding the fitness of the software in scope of "high risk activities".

The software is subject to export laws and provisions of the Federal Republic of Germany. You are obliged to comply with all national and international export laws and provisions which apply to the software including limitations with regard to point of destination, end user and end use.

You are allowed to transfer the license inside the country where the license has been purchased if the one to which the license is assigned to complies with the terms of license and the export laws and provisions of the Federal Republic of Germany are not in opposition to the transfer. In case of a transfer of the license you must delete the software free of residues and must no longer use it. To avoid ambiguities about the license it is recommended to inform the manufacturer of the software about the transfer.

In case of a move you are allowed to use the software only if the export laws and provisions of the Federal Republic of Germany are not in opposition to the move and if warranty deeds or warranties are not increasing from the legitimate point of view of manufacturer and the original supplier compared to the place at which the license was originally bought. Namely the local applicable law must remain the same one or liability risk of manufacturer or the original supplier must downgrade. To avoid ambiguities about the license it is recommended to inform the manufacturer of the software about the move.

Limited warranty and disclaimer

The software and the related documentation are supplied "as is". The manufacturer explicitly disclaims any warranty explicit or implied as to the usability of the software or its documentation or their fitness for any particular purpose. Neither the manufacturer nor its distributors will be liable for any direct, indirect, consequential or incidental damage (including loss of profits, business interruption or loss of data) which are based on the use of the software or its documentation or on the inability to use it or on its ownership or on the sale of the software no matter whether such claims are based on warranty, other contractual claims, illegal activities or other legal claims. This applies also if

the manufacturer or its distributors have been informed about the possibility of such a damage.

The disclaimers and limitations described above obligate you only in case the locally applicable law is not defining contrary terms.

If you purchased the software in Germany additionally the following clauses apply:

The manufacturer assumes no liability on selection, installation and usability of the combination selected by you or on the configuration of the software as well as on executability of the software or the results which you want to obtain with the software.

The manufacturer is not obliged to provide any liability with regard to the software, regardless of whether it is a matter of material or legal defects, as long as there is no fraudulent concealment of an error.

The manufacturer is liable in case of deliberate intention, gross negligence, titles according to the "Produkthaftungsgesetz" as well as in case of personal injuries according the legal requirements.

Any liability of the manufacturer due to the slightly negligent violation of essential contractual duties is limited to typical and foreseeable damages.

Severability clause

In case that individual terms of this terms of license are ineffective or void by law in full or in part or will become ineffective or void by law in full or in part due to a change of the legal situation or by high court decisions or in other ways or in case that this terms of license have gaps the other provisions of this terms of license remain unaffected of this and applicable.

Contact

Manufacturer:
Stader Softwareentwicklung GmbH
Tilsiter Straße 2
D-21680 Stade
Germany
Telephone: +49 4141 7891060
E-Mail: Verwaltung@stader-softwareentwicklung-gmbh.com.de
Handelsregister Registerblatt: HRB 207721
Registergericht: Amtsgericht Tostedt
Umsatzsteueridentifikationsnummer conformable to § 27 a Umsatzsteuergesetz: DE 327715351
Homepage:
http:///www.stader-softwareentwicklung-gmbh.com.de/

History

Version 0.03

Release date:

- 11-29-2021

Remarks:

- None

Corrected errors:

- Time problems with certain RISC OS territory modules
- No direct analysis of Pluto email export possible
- Bug in HTML/text parser memory allocation
- Term archiver: Wrong calculation of occurrence in case of mass analyse

New errors/problems:

- None

New features:

- Text parser: File types supported by text parser configurable
- Text parser: = handled as separator
- HTML parser: = handled as separator
- Email parser: Analyse of arbitrary fields added
- Email header fields search: Search of arbitrary fields added
- Analyse pause, countine, abort added
- Converter parser added
- Converter parser: Squash to unsquashed files added (if !Squash is available)
- Converter parser: BBC BASIC to text added
- Converter parser: PDF to text added (if !PDFUtils is available)
- Converter parser: TextEase to text added

History

Version 0.02

Release date:

- 10-25-2021

Remarks:

- None

Corrected errors:

- None

New errors/problems:

- Time problems with certain RISC OS territory modules
- No direct analysis of Pluto email export possible
- Bug in HTML/text parser memory allocation
- Term archiver: Wrong calculation of occurrence in case of mass analyse

New features:

- None because it is the first official version.

Technical details

Appearance of result file

The appearance of the result file is controlled by the SSI file <u>\<Retriever$Dir></u>
<u>.Resources.Results.Results/htm</u> and the BBC BASIC program \<Retriever$Dir>
.Resources.Results.EntryData. You may adapt them if you are knowing what you are doing. So keep
a copy before modifying the files.

Parser Text General RISC OS File Types

The list of the general RISC OS files types handled by the text parser is defined inside a configuration
file named <u>Choices:Retriever.config_p_txt</u>.
The list can be edited by a text editor. A change of this list requires a restart of Retriever to let the
changes take place.

Parser Email

You can enable or disable the claiming of message: URIs for Messenger by adding or removing the
respective Set Alias$URLOpen_message line inside the file <u>Choices:Retriever.config_div</u>.

Parser Email General Header Fields

The list of the general email header fields handled by the email parser is defined inside a
configuration file named <u>Choices:Retriever.config_p_email</u>.
The list can be edited by a text editor. A change of this list requires a restart of Retriever to let the
changes take place.

Converter Parser

All files which are handling the conversions are located inside <u>\<Retriever$Dir></u>
<u>.Resources.Parser.Converter directory</u>.
Retriever calls Adapter with RISC OS file type and the names of source and target file as parameters.
The source file is the original file which shall be converted. The target file is a temporary one which
will be automatically removed by Retriever after it has been analysed. Note that it must be ensured
that the target file has the same time stamp as the source file. Squash ensures this. For other programs
a simple BBC BASIC program named TimeCopy can be used to adapt the time accordingly after the
file has been generated. Adapter checks whether it can handle the desired RISC OS file type and calls
the according converter program by using the StartDeskTask command in conjunction with a
TaskWindow command. This ensures that the converter can act entirely independently of Retriever
which waits until conversion has been performed. To indicate the end of conversion the system
variable Retriever$ParserConvWait is used. As long as it is existing Retriever assumes that the
conversion is still working. The converter has to Unset Retriever$ParserConvWait when it finishes.
In case that a converter crashes you can emulate the end of conversion manually by typing
```
Unset Retriever$ParserConvWait
```

at a command prompt e.g. inside a task window. However you should only do so if you are knowing what you are doing. Keep a copy of the original Adpater file before you start modifying it.

So you can set up your own converters by adjusting Adapter accordingly.

Special requirements

Conversion of squashed files requires that !Squash has been booted before an according analyse takes place.

Conversion of PDF files requires that !PDFUtils has been booted before an according analyse takes place.

Converter Parser General RISC OS File Types

The list of the general RISC OS files types handled by the conveter parser is defined inside a configuration file named Choices:Retriever.config_p_conv.

The list can be edited by a text editor. A change of this list requires a restart of Retriever to let the changes take place.