



RECAPTURE FOR RECFIND

OVERVIEW



ISO 9001:2000
Lic No.
OEC 5844

RecCapture Overview Paper

Introduction	3
RecCapture Objective.....	3
Rules And Filters - Introduction	3
Rules.....	4
Attributes	4
Keywords	4
TAGS.....	5
Document Profile Tags	6
Building An Abstract.....	6
Mapping	7
Summary	8
Retention and Workflow.....	9
Organizing Your Electronic Document Working Store.....	9
How Many Versions Is Too Many?	9

Introduction

Within this paper I have tried to describe how RecCapture works. I would like you to review the processes and features and try to visualize how you would configure and use RecCapture within your organization. Will the RecCapture described herein satisfy your needs for a fully automatic, rules-driven EDMS?

In particular, I would like you to review the way we build rules and determine what to capture and how to store and index it in RecFind.

We are not asking you to review the 'internals' of RecCapture. That is, all the fancy and complex ways we process directories of electronic documents (and images) and capture new and changed documents. This is the low level stuff where we have to code to Windows Events and manage complex asynchronous activities and do it in the most efficient way possible so we do not place an undue load on your LAN, WAN or Servers.

We are asking you to review the application level stuff; especially the rules processes and the options we offer. These processes are the key to RecCapture being a success in your organization. Please examine these aspects of RecCapture very carefully and try to visualize how you would apply them in your organization to capture all of the electronic documents you need to capture and ignore those you do not wish to capture.

The Rules are the core of RecCapture. We have tried to give you every conceivable option. We have tried to give you the ability to define to the nth degree which documents you want and which documents you don't want. We have also given you the ability to specify exactly what should happen to a document once it is stored in RecFind-Corporate. That is, what file folder, what document profile, what Metadata, what Retention Schedule, what Workflow, etc.

RecCapture Objective

The objective of RecCapture is to provide a totally automatic (no end user involvement), secure and rules-driven way to efficiently capture and store and index and action any electronic document. This is a brand new paradigm so we don't have anything to compare it against - you will have to use your imagination and your knowledge of your organization's business processes and compliance requirements.

Rules And Filters - Introduction

A rule is applied to a filter. A filter is composed of a specific file type, e.g., *.DOC, *.XLS, *.JPG, (yes we can also capture images with RecCapture) and a directory location, e.g., F:\WP\Frank.

1. First we specify a Rule, e.g., Rule01.
2. Then we specify a Filter, e.g., *.DOC and f:\wp\frank.
3. Then we link the Rule to the Filter.

For example: "Please search the directory location F:\wp\frank for all *.DOC electronic documents (new and changed) and then apply Rule01 against those documents to determine which ones we want to save in our corporate store (RecFind-Corporate) and which ones we want to ignore. Please ignore those (i.e., the same version) we already have stored in RecFind-Corporate and please automatically version (i.e., as version 1, version2, version3, etc) new versions. In addition, once you have stored it please apply whatever Retention Schedule and Workflow I have specified and initiate both processes."

You can have as many filters as you like. Each filter is linked to a Rule, (and a Rule, as we shall see, may have many parts).

Rules

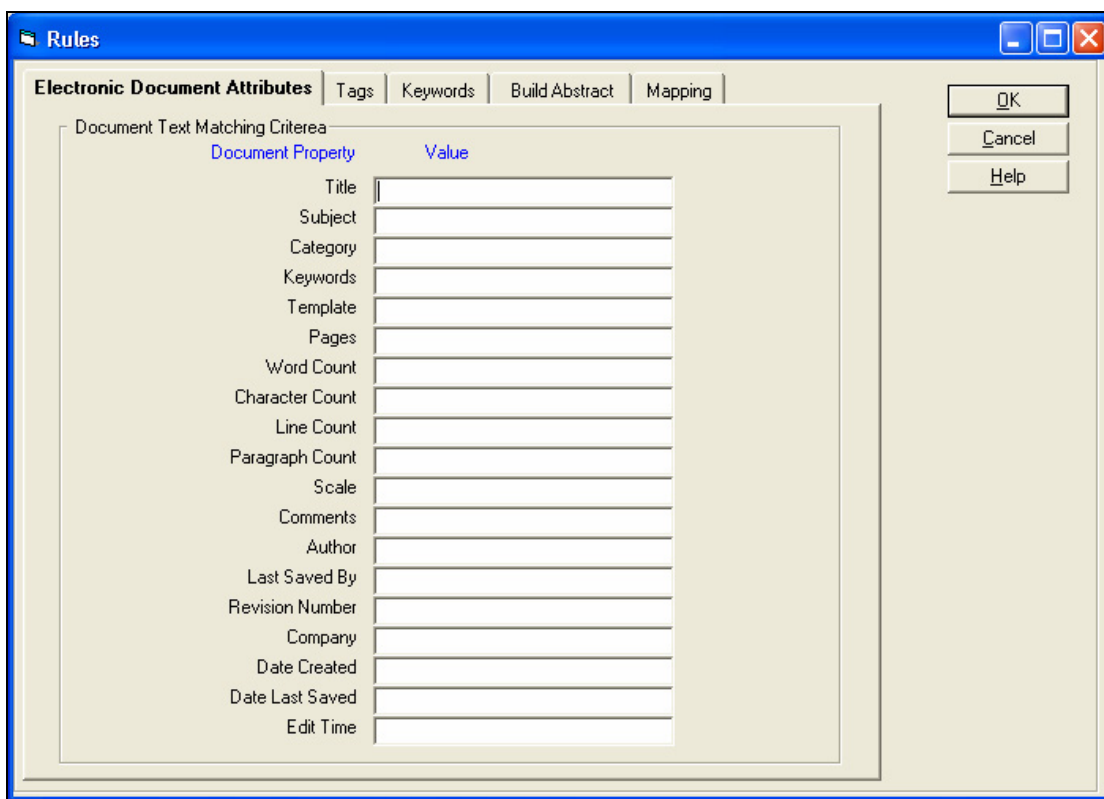
This is the core of RecCapture - the ability to build and apply a set of rules against your store of electronic documents.

We use rules to compare the attributes of an electronic document against a set of attributes you specify. Basically, if the attributes match we capture the electronic document; if they do not match we ignore it.

Attributes

There are many kinds of attributes. Some are automatically allocated by the operating system and some are entered by the user and some are a result of the source application used to create the document. We also call this collection of attributes "METADATA".

For example, below is one of the tabs from the RecCapture Rules module where you specify which attribute you wish to match on and what value you want to equal.



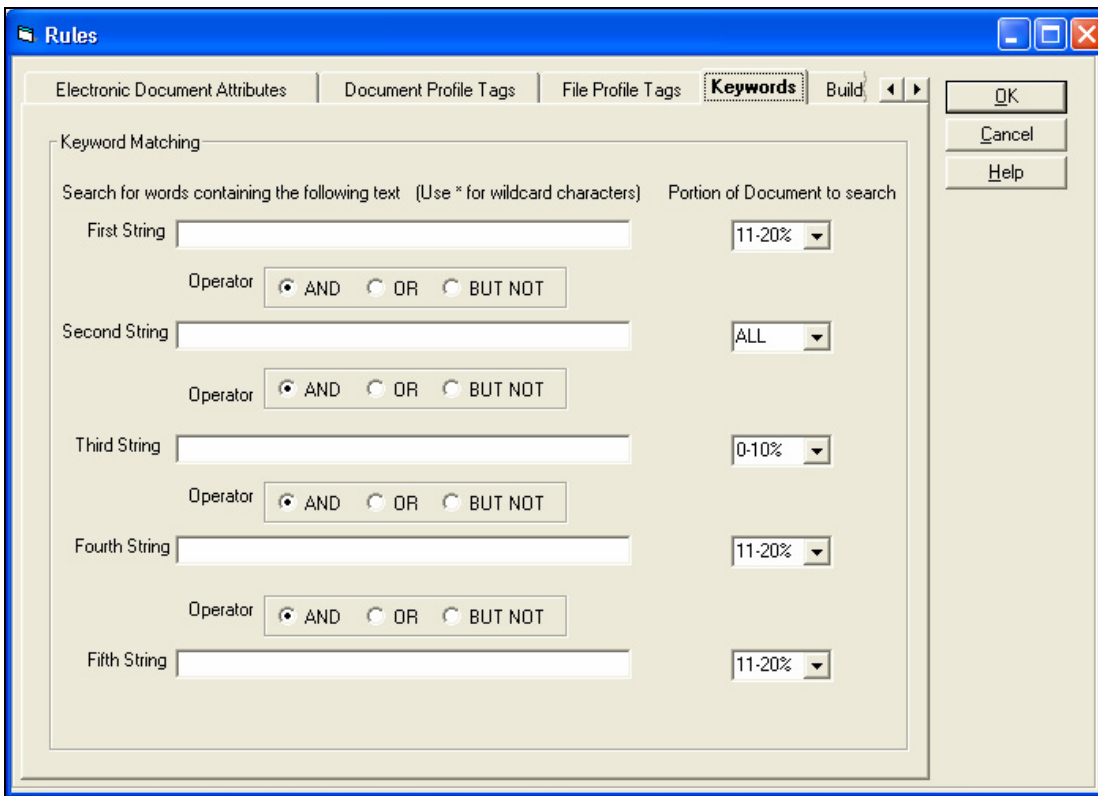
The screenshot shows a window titled "Rules" with a tabbed interface. The active tab is "Electronic Document Attributes". Other tabs include "Tags", "Keywords", "Build Abstract", and "Mapping". The main area is titled "Document Text Matching Criteria" and contains a table with two columns: "Document Property" and "Value". The table lists various document attributes, each with an empty text input field for its value. On the right side of the dialog, there are three buttons: "OK", "Cancel", and "Help".

Document Property	Value
Title	
Subject	
Category	
Keywords	
Template	
Pages	
Word Count	
Character Count	
Line Count	
Paragraph Count	
Scale	
Comments	
Author	
Last Saved By	
Revision Number	
Company	
Date Created	
Date Last Saved	
Edit Time	

We can call these the natural or 'system' attributes of any electronic document. RecCapture will compare the set of attributes you enter in a rule against the attributes of a document to determine if it is a match. Alphanumeric fields (e.g., Title) are parsed on a "CONTAINS" basis and numeric and date fields are parsed on an "EQUALS" basis. All values are "ANDed". This means if you enter 4 values then all 4 values must be matched in the document for the document to be saved under this part of the rule.

Keywords

You can also use the Keywords Tab of the Rule module to search for any Keywords within the text of the document. Obviously, this Tab does not apply to image files (e.g. TIF, JPG, etc) because they are bit-mapped files, not text files and there is no text to search.



The above is the Keywords Tab from the RecCapture Rules module. It allows you to specify Keywords to search for within the body of the electronic document. You can specify up to 5 Keywords and you can combine them in a BOOLEAN selection using AND, OR and BUT NOT.

You can also specify what part of the electronic document to search in for each Keyword. This is handy if you structure your documents (e.g., using templates) and 'know' where key items of information should be for specific types and classifications of documents.

Note that all Keywords are "ANDed" and all must be present in the electronic document for it to meet this part of the rule and be captured.

TAGS

In order to give you greater control over the capture and classification process we allow you to embed 'Tags' within the electronic document and then search for these. You can also use these same Tags to provide the Metadata to be used when creating the File folder and Document profiles within RecFind-Corporate.

The range of available Tags includes all the standard properties (Metadata) of the RecFind File Folder profile and the Document profile. For example:

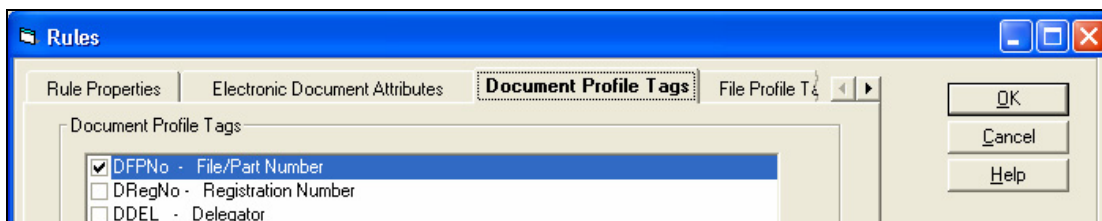
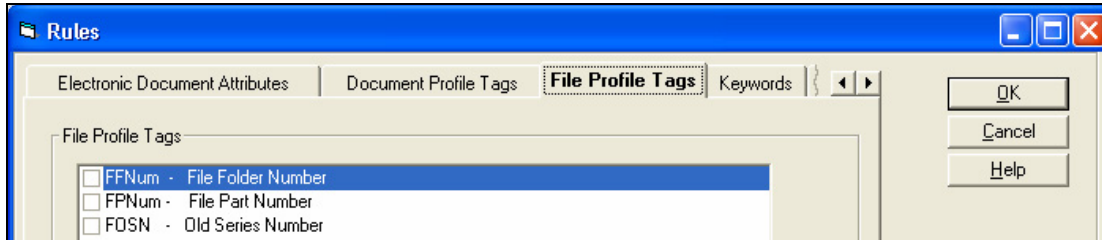
Document Profile Tags

RecFind Field	TAG
File/Part number	DFPNo
Registration number	DregNo
Delegator	DDel
Senders ref	DSR
Action Officer	DAO
Author	DAuth
Location/Note	Dloc
Abstract1	DAb1
Abstract2	DAb2
Abstract3	DAb3
Abstract4	DAb4
Abstract5	DAb5
major	DM
Folio	DF
Document type	DT

You enter a starting and ending delimiter (similar to XML syntax) with an embedded tag and then add the Tag anywhere in the body of your documents (you can for example put all Tags at the beginning of the document and make them invisible). If wanted to imbed the RecFind Document profile registration number in your document it would look as follows:

`<DregNo>04/6758</DregNo>`

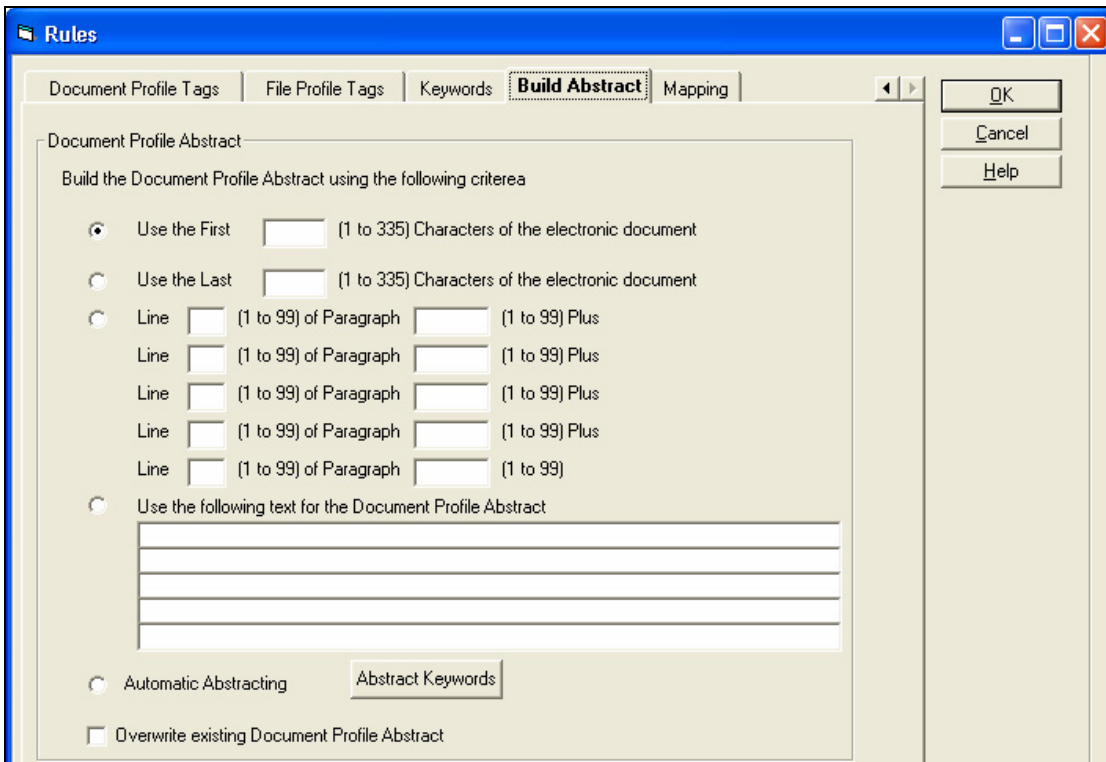
For this part of the Rule you tell RecCapture what Tags to search for, see below:



If would be easy to automate this process (the embedding of Tags with electronic documents) using templates and macros so you have much greater control over the classification and processing of your electronic documents.

Building An Abstract

When you capture an electronic document using a Rule you can choose to attach it to an existing Document profile or you can choose to create a new Document profile. Part of the Metadata of any RecFind Document profile is the abstract (335 characters). You can either enter the abstract text you want or you can ask RecCapture to build the abstract based on a full text analysis of the actual electronic document. We give you multiple ways to do this so you can use the method most appropriate to each type of document; see below:



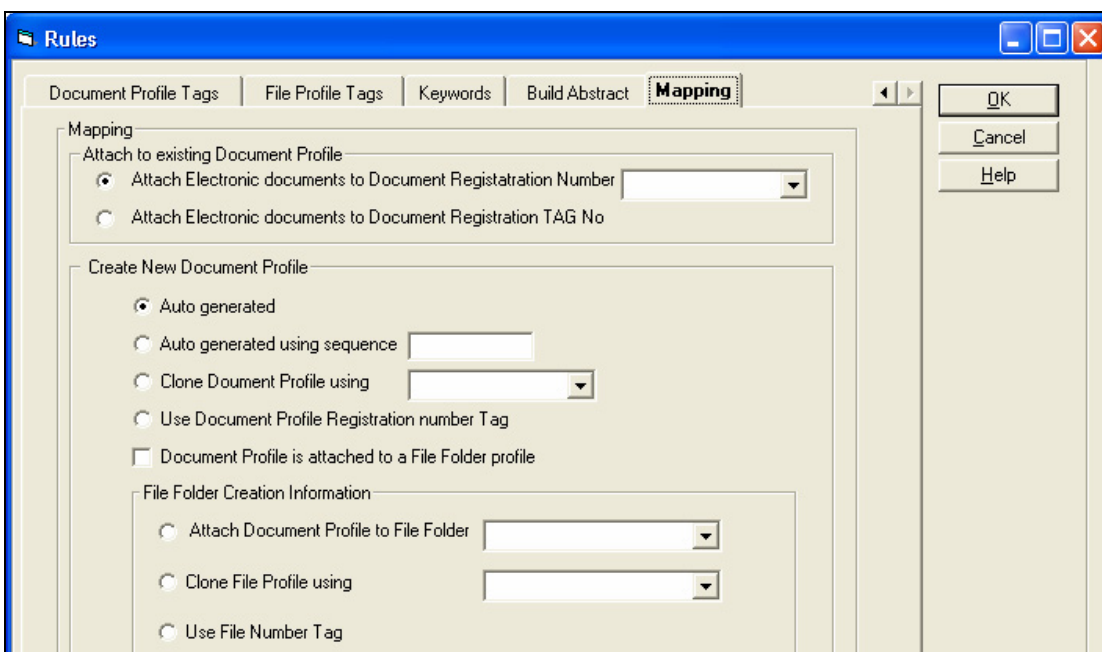
The first 4 choices are fairly obvious, e.g., Use the first nnn characters of the electronic document and "Use the following text...".

Option number 5, "Automatic Abstracting" is GMB's proprietary abstracting algorithm based on a set of Keywords. You can enter the Keywords, you can use RecFind's Keyword Thesaurus or you can import the Keyword set from a CSV file. You can also edit the set of Keywords at any time.

We suggest that you experiment with all methods until you find the one that best suits your data. If in doubt, use option 5 as the most appropriate 'general' solution.

Mapping

The last thing we have to do after deciding to capture a document is to decide how to store it in RecFind's corporate store. We call this "Mapping."



Once again we have given you many, many options. You can simply select a Document Profile (Doc Reg number) or you can use the one imbedded as a Tag or you can create a new one from scratch or you can use an existing Document profile template to 'clone' the new one.

Summary

We build Filters to point to the network location we want to search and to nominate the type of electronic document we are searching for, e.g., .DOC.

We can have any number of Filters pointing to the same network location.

```
*.DOC = j:\WP\Frank = Rule01  
*.XLS = j:\WP\Frank = Rule11  
*.JPG = j:\WP\Frank = Rule43
```

We can have multiple filters pointing to the same network location all specifying the same document type (but each one linked to a different Rule).

```
*.DOC = j:\WP\Frank = Rule01  
*.DOC = j:\WP\Frank = Rule02  
*.DOC = j:\WP\Frank = Rule03
```

We can have multiple Filters for different document types pointing to multiple network locations.

```
*.DOC = k:\Contracts\Acme Corp\Orders  
*.TIF = k:\Contracts\Acme Corp\invoices
```

We build Rules for each type or classification of document we wish to capture and process.

We assign a Rule to a Filter.

A Rule is made up using the values from 7 Tabs as follows:

1. Rule Properties Tab

This is where we define the name and description of the rule. *There is nothing in this Tab to determine if a document is to be captured or not.*

2. Electronic Document Attributes Tab

This is where we specify all 'natural' properties of an electronic document (Title, Author, Date Created, etc) and enter values to be matched against the document. If all values are matched then Capture is set to TRUE. If all values are not matched then Capture is set to FALSE.

3. Document Profile Tags Tab

This is where we select the Document Profile Tags to be searched for within the body of the document. Each Tag represents an attribute of the RecFind Document Profile (e.g., Registration number, Author). *There is nothing in this Tab to determine if a document is to be captured or not.*

4. File Folder Profile Tags Tab

This is where we select the File Folder Profile Tags to be searched for within the body of the document. Each Tag represents an attribute of the RecFind File Folder Profile (e.g., File folder number, first line of title, etc). *There is nothing in this Tab to determine if a document is to be captured or not.*

5. Keywords Tab

This is where we specify up to 5 keywords/phrases to be searched for within any part of the document. If all values are matched then Capture is set to TRUE. If all values are not matched then Capture is set to FALSE.

6. Build Abstract Tab

This is where we specify if and how the RecFind Document Profile abstract is to be built. *There is nothing in this Tab to determine if a document is to be captured or not.*

7. Mapping Tab

This is where we determine where and how the document will be stored; what document profile, what file folder profile. *There is nothing in this Tab to determine if a document is to be captured or not.*

Retention and Workflow

Retention and Workflow are both major functions of RecFind-Corporate. You will build/enter your Retention schedules in RecFind-Corporate and you will build your Workflows in RecFind-Corporate.

You assign both to a RecFind Document profile, either an existing one or one you create dynamically using the Rules Mapping Tab. You can dynamically assign either by either using the Tags feature or by using a clone.

Organizing Your Electronic Document Working Store

RecCapture continually searches your working store and captures copies of electronic documents and stores them as inviolate corporate records, together with all Metadata, in the RecFind-Corporate relational database. That is our job. We help you meet all EDMS and compliance requirements and guarantee an inviolate copy of each and every version of a corporate record.

One can envision any number of ways organizations organize their electronic document working store. We could simply have all documents in a single directory path, e.g., k:\WP or we could build the most complex structure so that each function and customer and type of document has a specific network location.

One thing we know for sure is that every customer will have a different structure in place.

RecCapture, because of its Filter feature, will easily apply to any structure no matter how simple or complex.

However, we anticipate that once an organization starts using RecCapture that it will begin to rethink and re-implement the way it saves electronic documents in its working store. We would also expect organizations to start making more use of templates and macros in order to explicitly tell RecCapture what and how to store electronic documents.

How Many Versions Is Too Many?

Let's say we are all running with 'AutoSave' set to 5 minutes. Let's also say we are all working on 50MB Word documents. Do we really want every saved 'version' to be also saved in RecFind-Corporate?

This is a tough question. Because there is no universal answer (it varies by organization) in RecCapture we solve this vexing problem by only acquiring a new or changed document if we can obtain exclusive access. This means we effectively ignore autosaves and multiple saves by a user when working on a document. We only obtain a copy of the new or modified document when the user has closed it allowing us to obtain exclusive access.