



# OpenText RightFax Searchable PDF

## Creating Searchable Content from Fax Images

OpenText Fax and Document Distribution Group  
March 2011



## Contents

<b>Executive Summary .....</b>	<b>3</b>
<b>Introduction.....</b>	<b>4</b>
Challenges of traditional fax capture, routing and storage .....	4
Providing support for searchable faxes .....	6
<b>Technology Overview: Searchable Faxes.....</b>	<b>8</b>
Specifications .....	9
<b>Summary .....</b>	<b>9</b>
<b>About OpenText.....</b>	<b>10</b>



## Executive Summary

Large enterprise wide storage systems combined with Enterprise Content Management (ECM) applications have become common investments for organizations looking to manage the large volumes of content and data pertaining to their business transactions. The return on investment is staggering, as efficiency gains, reduction of costs through consolidation (storage and paper reduction for example) and risk avoidance due to compliance mandates are huge drivers for a company's long term cost-controlled IT strategy.

At the core of the ECM and Storage Management system are a company's business documents and data. Alongside this core of critical information are systems with the ability to apply discovery and retrieval methods such that information can be searched and retrieved easily.

Not all documents or information that ends up in an ECM system are the same. Faxes for example, are difficult to search since after all, they are mere images of a document, albeit legally binding and acceptable in its photographic form. Mining for the true content of a fax requires various methods of intervention, some manual and some more automated. Companies have standardized the managing of faxes on enterprise fax server systems to do just that. Today's fax server has various complex methods in which to accept and route captured faxes into their business workflows. A logjam occurs however when faxes are archived or stored and at some point require a future search and retrieval. Because it's an image, the information on the fax is rarely able to be identified and thus properly indexed and searched upon. The overhead involved in managing faxes in an ECM then affects the bottom line as productivity and efficiencies are challenged.

The objective is to then raise the level of content that can be gleaned from a fax. This paper introduces OpenText RightFax and its ability to convert incoming faxes into searchable documents. This paper highlights a new method to convert captured business faxes into searchable content which improves efficiency, productivity and leverages ECM investments.



## Introduction

Although fax servers have been and continue to be a robust and growing means by which companies send and receive critical business documents, the ability to index a fax document and derive searchable content can be challenging. This is especially true for those organizations who receive significant volumes of inbound faxes from customers, suppliers, clients, or applicants where the fax is initially captured into a centralized mailbox for review or processing. In a typical scenario when a fax is received, a worker must often read it, and then decide where or to whom it gets routed or forwarded based on content or information that was read. Once routed, the fax is often associated with other content related to that specific business function like a claim, a supplier account, a customer account, a healthcare patient, etc. Finally, most organizations today have adopted document retention policies in which all business documents and messages are archived into a variety of storage or content management systems.

## Challenges of Traditional Fax Capture, Routing and Storage

It is a given fact that inbound fax processing (which is often manual) can be slow and subject to errors if not processed accurately. To improve upon this, organizations may opt to utilize some form of automated routing procedure via a network fax server. One popular method is telephony-based routing, in which the fax server leverages the company telephone switch and network to route captured faxes efficiently based on a variety of options. Some common methods include using Direct Inward Dial (DID) plans, identifying the sender's number using CSID or ANI (Call Subscriber ID, Automatic Number ID, etc.) or even assigning a dedicated fax channel for a specialized business function. Telephony methods, although efficient, require additional resources that may challenge company budgets and the phone services are often fee-based and result in incremental monthly or annual expenses. Additionally, these methods do not extract information from the fax document directly, but rather have limited information about the telephony characteristics of the fax like the recipient's DID number, the sender's phone number, etc. Searching and retrieving the fax images results in inefficiencies since no data was extracted about the fax's content and associated with the document.

Better information can be gleaned from the fax, however. A sender-based Optical Character Recognition (OCR) can be used to interpret words and numbers on the fax and use that information to route or forward the document more intelligently - but that too comes with limitations. A sender-based OCR workflow relies on the sender to have the correct information on the page, in the correct location, using an acceptable font and at an acceptable resolution (and in some cases the right language) in order for it to be interpreted correctly. And, information about the fax is limited to the pre-defined OCR words or numbers and this may adversely influence the speed and accuracy of future



search and retrieval operations. This works in business scenarios where there are frequent transactions or where there are pre-defined business forms in use, or where the location of the key information is always known.

A more advanced OCR method involves the use of bar codes. 1 dimensional (1D) or 2 dimensional (2D) codes can be placed on cover pages of incoming faxes which are scanned and recognized. Routing rules can then be applied and the fax sent to the correct recipient for processing. Like sender-based OCR however, future lookups and retrievals may not be as efficient since the code may only contain a certain amount of predefined data. The goal however, is to glean content from the fax and then use that information for indexing which will optimize search and retrieval operations. Below is an overview of traditional methods and their disadvantages:

Method	Description	Drawbacks
Manual	<ul style="list-style-type: none"><li>• Central inbox or network folder is monitored by a user or group of users</li><li>• Faxes are read, then routed or forwarded to a user or into a workflow process</li></ul>	<ul style="list-style-type: none"><li>• Requires user intervention</li><li>• The fax document has no associated meta data, other than a tracking log</li><li>• Faxes are not indexed; search and retrieval of faxes could be inefficient</li></ul>
Telephony	<ul style="list-style-type: none"><li>• Various telephone signaling or line provisioning applications are used to route faxes more efficiently</li></ul>	<ul style="list-style-type: none"><li>• Requires a telephony infrastructure capable of supporting</li><li>• Faxes are not indexed; search and retrieval of faxes could be inefficient</li></ul>
Sender-based OCR (Bar Codes and Characters)	<ul style="list-style-type: none"><li>• OCR is applied to inbound faxes. Looking in a predefined area or for a predefined set of characters or a bar code</li><li>• Data is matched to a routing table and forwarded to correct recipient.</li><li>• A text file accompanies the fax with the OCR information</li></ul>	<ul style="list-style-type: none"><li>• Accuracy of the OCR relies on the sender ensuring the characters and/or bar code is present</li><li>• The fax image file must be associated with a text file of the OCR information.</li><li>• Other content from the fax is not able to be part of the searchable index. Search and retrieval of faxes could be inefficient</li></ul>



## Providing Support for Searchable Faxes

### Solution Overview

OpenText RightFax offers a new fax file format called “Searchable PDF” which as its name implies, creates a searchable portable document file (PDF) after applying an OCR procedure on the captured fax. PDF is the chosen format because of its ubiquitous availability and readability virtually worldwide. Once optically recognized, indexed and stored, the fax is able to be searched more accurately.

Once recognized, the faxes can be sent to an email gateway (e.g., SMTP, Exchange or Lotus), a network folder (where it can be searched), or directly to the RightFax XML Generator for integration into other Content Management Systems, or any application that can ingest XML data, including numerous multi-function products (MFPs). Highlights of the Searchable PDF (SPDF) feature are below:

Feature	Description	Benefits
<ul style="list-style-type: none"><li>• Applies OCR to incoming faxes, converts to a searchable PDF</li></ul>	<ul style="list-style-type: none"><li>• Inbound faxes are captured and converted to a searchable PDF file.</li><li>• Conversion process can be run remotely for optimal load balancing and availability.</li></ul>	<ul style="list-style-type: none"><li>• Reduces manual processing</li><li>• Increases productivity</li></ul>
<ul style="list-style-type: none"><li>• Delivers faxes with searchable content as Portable Document Format (PDF) files</li><li>• Sends faxes as searchable PDFs from MFPs and XML Generator</li></ul>	<ul style="list-style-type: none"><li>• Delivers OCR results to either a Network Folder, Email, or a Network MFP</li><li>• Users can view the PDF file embedded with the original fax image file (TIFF).</li><li>• Optionally integrates with RightFax XML Generator for integration to other applications or MFPs</li></ul>	<ul style="list-style-type: none"><li>• Streamlines workflow efficiency</li><li>• Reduces error</li><li>• Leverages existing IT investments</li></ul>
<ul style="list-style-type: none"><li>• Faxes are stored with searchable content</li></ul>	<ul style="list-style-type: none"><li>• iFilter indexing plugin is used to create the PDF index</li><li>• Desktop searches can be performed by anyone with a typical desktop search engine</li><li>• Search engines look for a match within an index.</li></ul>	<ul style="list-style-type: none"><li>• Improves worker productivity</li><li>• Reduce risk associated with document handling requirements</li></ul>



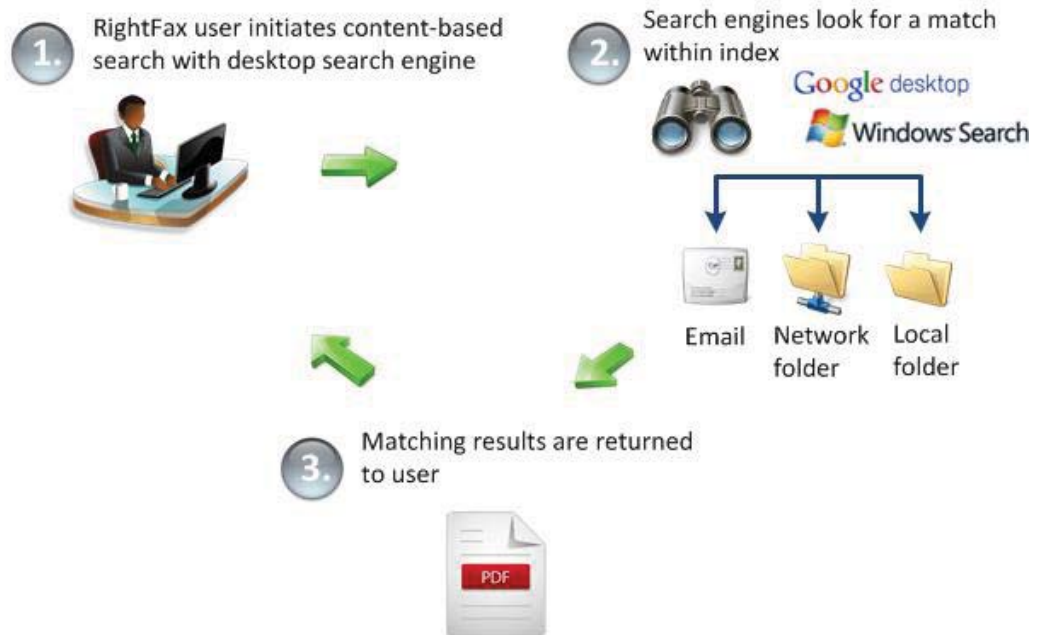
## Technology Overview: Searchable Faxes

With Searchable PDF, RightFax improves fax document routing, retrieval, and document management.



*Incoming faxes are converted to searchable portable document files (PDF) which can be delivered to any application and stored using the captured data*

Search and retrieval operations are now easier for faxes that have been stored since relevant data captured from the fax are now within the search index.



*Using typical Desktop Search Engines, a RightFax user can search and retrieve fax content*



## Specifications

- RightFax 10.0 or greater
- The RightFax Searchable PDS (SPDF) module
- A desktop search engine
- Optional RightFax XML Generator (for integration to other applications or MFP's)
- iFilter (for PDF indexing)

## Summary

The OpenText RightFax Searchable PDF option can significantly streamline the ability to store and retrieve fax content. Integration into a variety of content management systems or multi-function products is possible with the use of the RightFax XML Generator. This allows organizations to leverage existing investments, bolster productivity, and improve their document handling policies.

---

**[www.ATechnologies.com](http://www.ATechnologies.com)**

Phone (866) 730-1700 | [www.ATechnologies.com](http://www.ATechnologies.com) | [sales@ATechnologies.com](mailto:sales@ATechnologies.com)