GOSCAN

Using GoScan With Free Google Desktop

Table of Contents

Table of Figures

Notice

This document is a proprietary product of GoScan and is protected by copyright laws and international treaty. Information in this manual is subject to change without notice and does not represent a commitment on the part of GoScan. While reasonable efforts have been made to ensure the accuracy of the information contained herein, GoScan assumes no liability for errors or omissions. No liability is assumed for direct, incidental, or consequential damages resulting from the use of the information contained in this document.

The copyrighted software that accompanies this document is licensed to the End User for use only in strict accordance with the End User License Agreement, which the Licensee should read carefully before commencing use of the software. No part of this publication may be reproduced, transmitted, stored in a retrieval system, nor translated into any human or computer language, in any form or by any means, electronic, mechanical, magnetic, optical, chemical, manual or otherwise, without the prior written permission of the copyright owner, GoScan, 26501 Rancho Parkway South, Ste. 103, Lake Forest, CA 92630.

GoScan® is a registered trademark of GoScan, Inc.

This document may use fictitious names for purposes of demonstration; any references to actual persons, companies, or organizations are strictly coincidental.

Contact Information

GoScan Company Headquarters

26501 Rancho Parkway South, Ste. 103
Lake Forest, CA 92630

Phone: 949-829-5822

Email: info@goscan.com

Web Site: www.goscan.com

Abstract

G oScan® provides users a fast, accurate, and easy way to index documents into an electronic document repository. For medium or large companies, an electronic repository usually means a database such as Oracle or an application designed for content management such as Microsoft SharePoint.   Smaller companies may opt to simply store their images in a directory (also called a file system).

There are many differences between storing images in a database or application and a directory.

- Database or Application
    - Requires a software license
    - Requires design, setup and maintenance
    - Provides a search interface
- Directory
    - Requires no software license
    - Requires minimal design, setup or maintenance
    - Does not provide a search interface

The simple solution for providing a search interface for documents stored in a directory is to use free, off the shelf tools from Microsoft or Google. Specifically, Microsoft Search 4.0 ("MS4") or Google Desktop ("GD"). Both of these are free, easily customizable and work with XP or Vista.

In this scenario, GoScan provides the tool to capture and index the documents. MS4 or GD provide a simple search interface. Any existing workstation, server or storage device stores the actual documents. MS4 operates as a program within Windows. GD operates inside a web browser.

The overall solution gives companies a fast, inexpensive way to implement a basic electronic document storage and retrieval system.

Field Name	Value
MemberID	12345
DocumentCategory	Labs

Scanning With GoScan

Use GoScan to scan and index documents to a directory. In this case, the directory is on a server accessible by all computers. The directory could be a workstation, server or a network attached storage device. The picture below shows the results from GoScan scanning a stack of documents, using barcodes to automatically generate filenames and saving one group as .tif files and one group as .pdf files.

Name	Date modified	Type
12345-A98765	12/10/2008 10:09 ...	TIFF Image
54321-A98765	12/10/2008 10:09 ...	TIFF Image
75845-A98765	12/10/2008 10:09 ...	TIFF Image
98547-A98765	12/10/2008 10:09 ...	TIFF Image
IDAUTOMATIONHC39M-RA91-1563-CXHZT_00001	12/10/2008 10:05 ...	Adobe Acrobat D...
IDAUTOMATIONHC39M-RA92-1564-CXHZS_00001	12/10/2008 10:05 ...	Adobe Acrobat D...

Figure 1 - Files scanned by GoScan

Google Desktop and OmniPage Search Indexer

Although the Microsoft Search 4.0 product works well, GoScan is recommending Google Desktop ("GD") at the present time because of the availability of a plug-in called OmniPage Search Indexer ("OSI").   OSI gives added benefit to GD by performing full text OCR on .tif or .pdf images scanned by GoScan. All of the OCR is performed automatically in the background. See http://desktop.google.com/ for details on GD or http://www.nuance.com/omnipage/search/ for details on OSI.

GD allows you to search your computer as easily as you search the web, using the familiar Google interface. GD requires minimal setup and all the indexing is done automatically in the background. OSI creates index data from document images, without changing the original. OSI includes an OCR engine that is very fast and accurate., and can OCR both .tif and .pdf files. Both GD and OSI are available free.
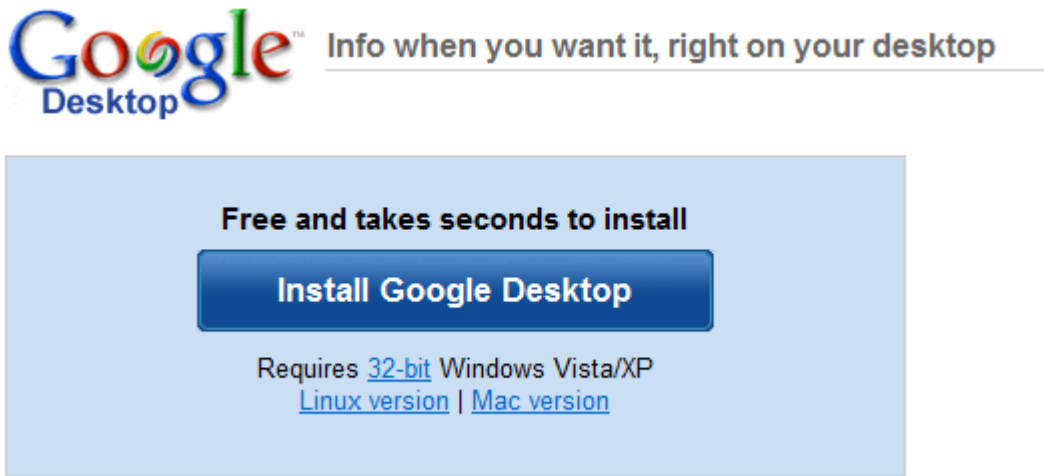

Figure 2 - Free Download For Google Desktop

Figure 3 - Free Download of OmniPage Search Indexer

Searching with Google Desktop

The figure below shows a sample scanned document scanned from GoScan. The name of the file is 12345-A98765.tif. GoScan automatically named the file based on the two barcode fields (only the first barcode is visible in the screenshot). When GoScan scanned the document, it performed barcode recognition on it but not OCR recognition. Therefore it did not know what text was on the document.
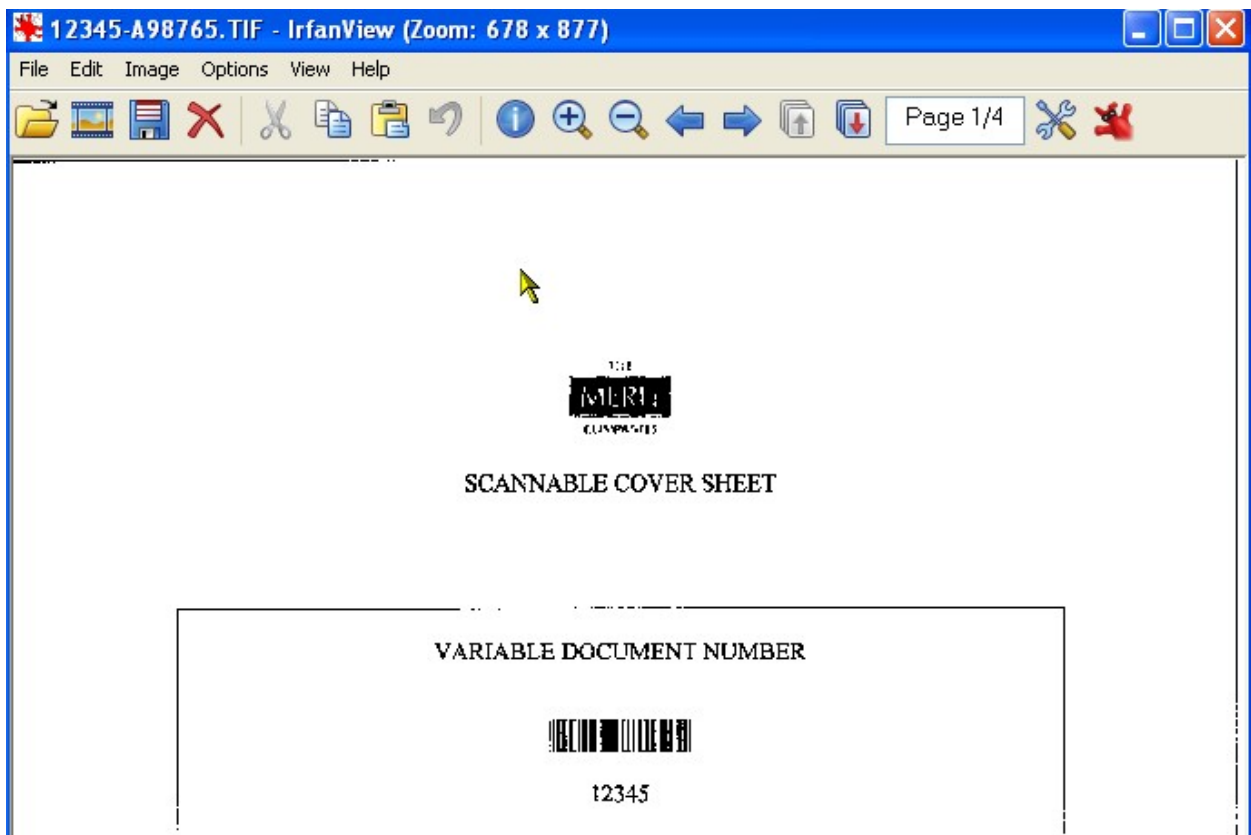

Figure 4 - Sample Scanned Document

The figure below shows the OCR enhanced search results from GD. Using GD by itself, we would see only the filename. This would be useful but not as useful as having both the filename and some idea of the content. With the OSI plug-in, we see the name of the file and the OCR'd results. This means that you can search either by the filename or by content on the document.
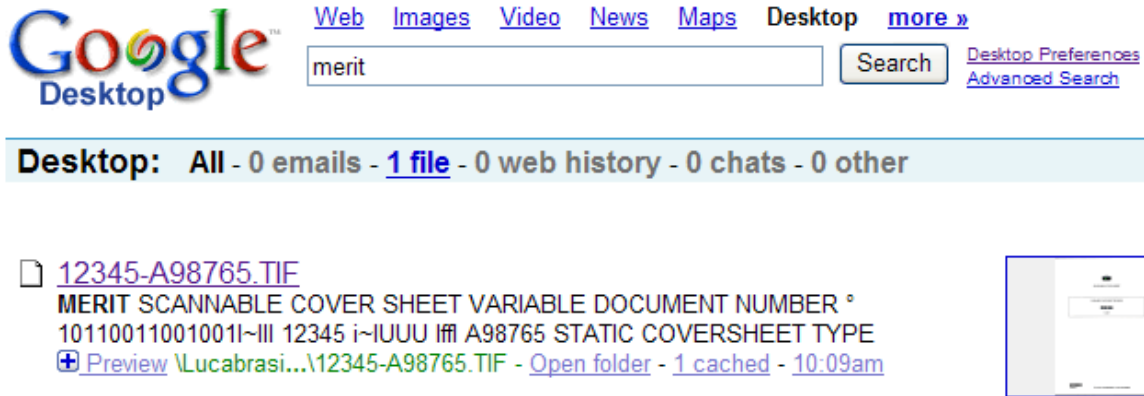

Figure 5 - OCR Enhanced Search Results From Google Desktop

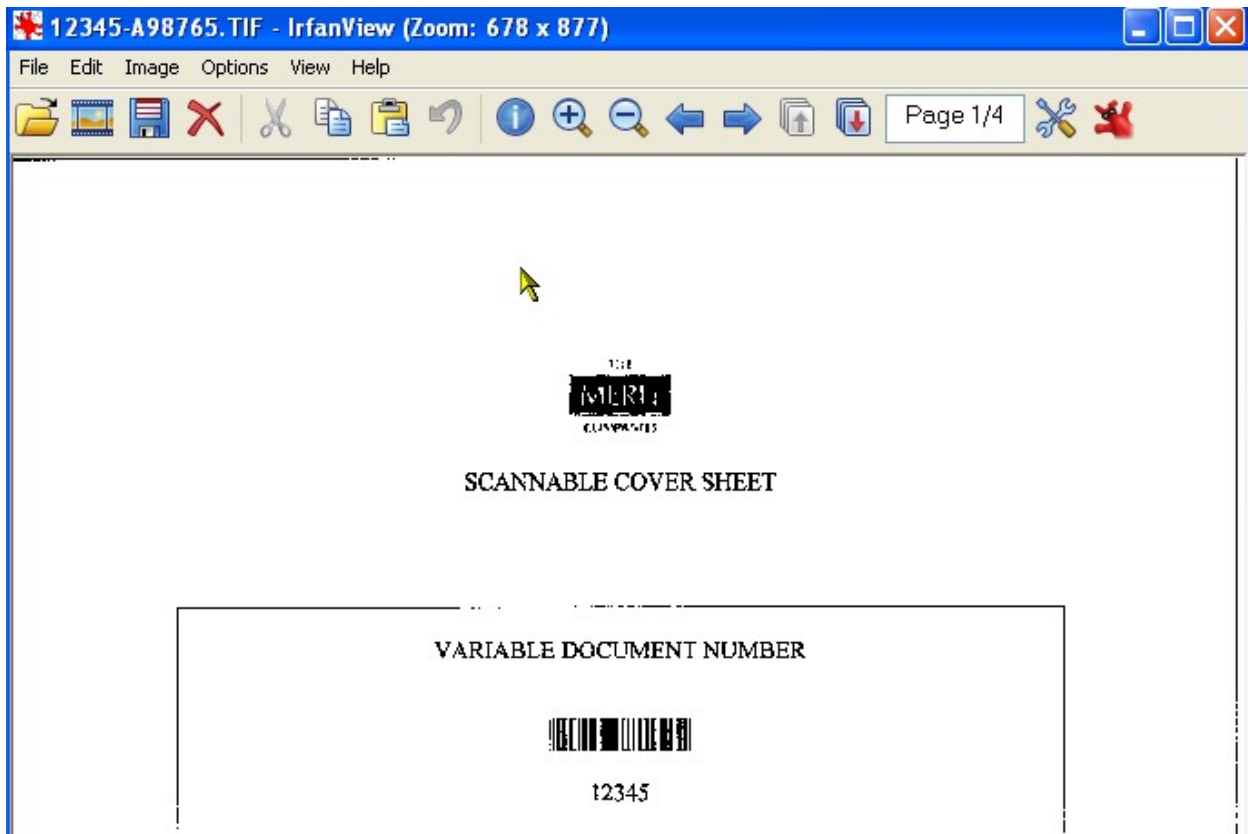Just click on the image to see the results.

Figure 6 - Search Results

Scanning, searching and displaying files works the same for .pdf and .tif image files.

Figure 7 - Sample PDF Image Search

Figure 8 - Sample PDF Image Display

System Architecture

There are a number of variables in setting up the system. The right solution depends on your unique requirements. GD can be installed on a single workstation or on a server accessible by multiple workstations. There is also GD for Enterprise (see http://desktop.google.com/enterprise/about.html for more details). The images can be scanned to a local workstation, a server or a network attached storage device.

Here is an example scenario for implementing a solution using GoScan and GD.

1. Install GoScan on all desktops.
2. Setup a high capacity disk drive on a server accessible by all users. Map this drive letter for all users. As an example, let's say this is i:\images
3. Scan and index from GoScan to the network drive i:\images.
4. Install GD Enterprise and OSI. Configure GD Enterprise so that it is only indexing i:\images.
5. Install GD Enterprise client on local user machines.
6. Search i:\images from local user machines using the GD interface.