



COMPUTER CREDIT, INC.



**THE
INTEGRATED
COLLECTION
STRATEGY.**

Using Adobe Catalog for PDF Indexing at CCI

Presented by Michael Rosinger

WAVV 2004

Chattanooga, TN

May 2004

mrosinger@cciws.com



Using AdobeTM Catalog for PDF Indexing at CCI



COMPUTER CREDIT, INC.



A little bit about CCI...

- Collection Agency since 1972
- Small company with just over 50 employees
- Serves 490+ facilities in 48 states
- Provides exclusive, low-cost “Third Party” Collection Letter service
- Provides integrated and automated Patient Accounting solutions
- We do *not* send Uncle Vinny out after dark!





The Problem...

- Need an inexpensive report storage solution
- Integrate both mainframe reports with other manuals and documents
- Handle all processing “in house” (immediate results)
- Fast, easy way to search
- Easy to access
- Requires minimal user training





The motivating reasons for considering Adobe...

- PDF is an ad-hoc industry standard
- Relatively inexpensive
- Can quickly implement
- Requires minimal user training
- Easy to configure
- Has plenty of *growth* potential





Could Adobe provide the solution?

- All of our mainframe reports already PDF'ed
 - Generated by VSE2PDF™ via TCP/IP
- Most of our documents and manuals are PDFs
- Most people are familiar with Acrobat Reader
- But what about search capability?





Adobe Catalog

- Catalog is an integrated part of Acrobat Professional
- Provided to create “fast search” indexes
- Fairly flexible...
 - Numbers may be excluded
 - Stop word list may be defined to reduce index size
- Granularity...
 - One index may include multiple PDFs
 - One index may span multiple folders
 - May search multiple indexes concurrently





What it required...

- At least one copy of Acrobat Professional
 - Full version is ~\$400
 - Upgrade is ~\$150
- A dedicated server with plenty of disk space
- Installation of Acrobat Reader on all work stations
- User training
- A “PDF Manager” (a person)





Our PDF Server...


- Recently migrated to new (2nd) server
- Dell PowerEdge 2600
 - Windows 2003 Standard Server
 - 2.4 Ghz Xeon Processor
 - RAID 5 with 4-146 GB SCSI drives, for PDF storage
 - 2-36 GB drives, mirrored for Windows O/S
 - Redundant power supplies
- Cost about \$7,800



Our PDF Server...



Pdf on 'Ntserver07pdf' (P:) Properties [?] [X]

General | Security


 **Label:** df on 'Ntserver07pdf'

Type: Network Drive

File system: NTFS

 Used space:	39,760,932,864 bytes	37.0 GB
 Free space:	400,316,186,624 bytes	372 GB

Capacity: 440,077,119,488 bytes 409 GB



Drive P

Compress drive to save disk space

Allow Indexing Service to index this disk for fast file searching

OK Cancel Apply

Windows Explorer window showing the contents of the folder **P:\Manuals&SystemDocs**.

Address: P:\Manuals&SystemDocs

Name	Size	Type	Modified
BIM CSI		File Folder	4/9/2004 12:32
BSI Products		File Folder	3/5/2004 09:04
CCI Database		File Folder	2/12/2004 07:33
Db2Vse71Pdf		File Folder	2/12/2004 07:33
Db2Vse73Pdf		File Folder	3/11/2004 16:42
Debug Tool		File Folder	2/12/2004 07:33
DF Sort		File Folder	2/12/2004 07:33
Ditto		File Folder	2/12/2004 07:33
Emerald Software		File Folder	2/12/2004 07:33
Group1		File Folder	2/12/2004 07:33
IBM Redbooks		File Folder	2/12/2004 07:32
IBM ServeRAID Manager		File Folder	2/12/2004 07:32
iFrame, Netfinity, Flex, etc		File Folder	2/12/2004 07:32
MacKinney		File Folder	2/12/2004 07:32
MaxBackVse		File Folder	3/12/2004 10:38
MS4 documentation		File Folder	2/12/2004 07:32
Open Software Technologies		File Folder	2/12/2004 07:32
P390 Documentation		File Folder	2/12/2004 07:32
Phoenix 21_0		File Folder	2/12/2004 07:32
Phoenix 21_1		File Folder	2/12/2004 07:32
PowerChute		File Folder	2/12/2004 07:32
QMF for Windows 7		File Folder	2/12/2004 07:32
Series 2000		File Folder	2/12/2004 07:32
SourceSafe		File Folder	2/12/2004 07:32
VSE_Collection_03_2002		File Folder	2/12/2004 07:31
VSE_Collection_09_2002		File Folder	2/12/2004 07:30
WAVV		File Folder	2/12/2004 07:29
Xerox		File Folder	2/12/2004 07:29

28 object(s) (Disk free space: 372 GB) | 0 bytes | Local intranet

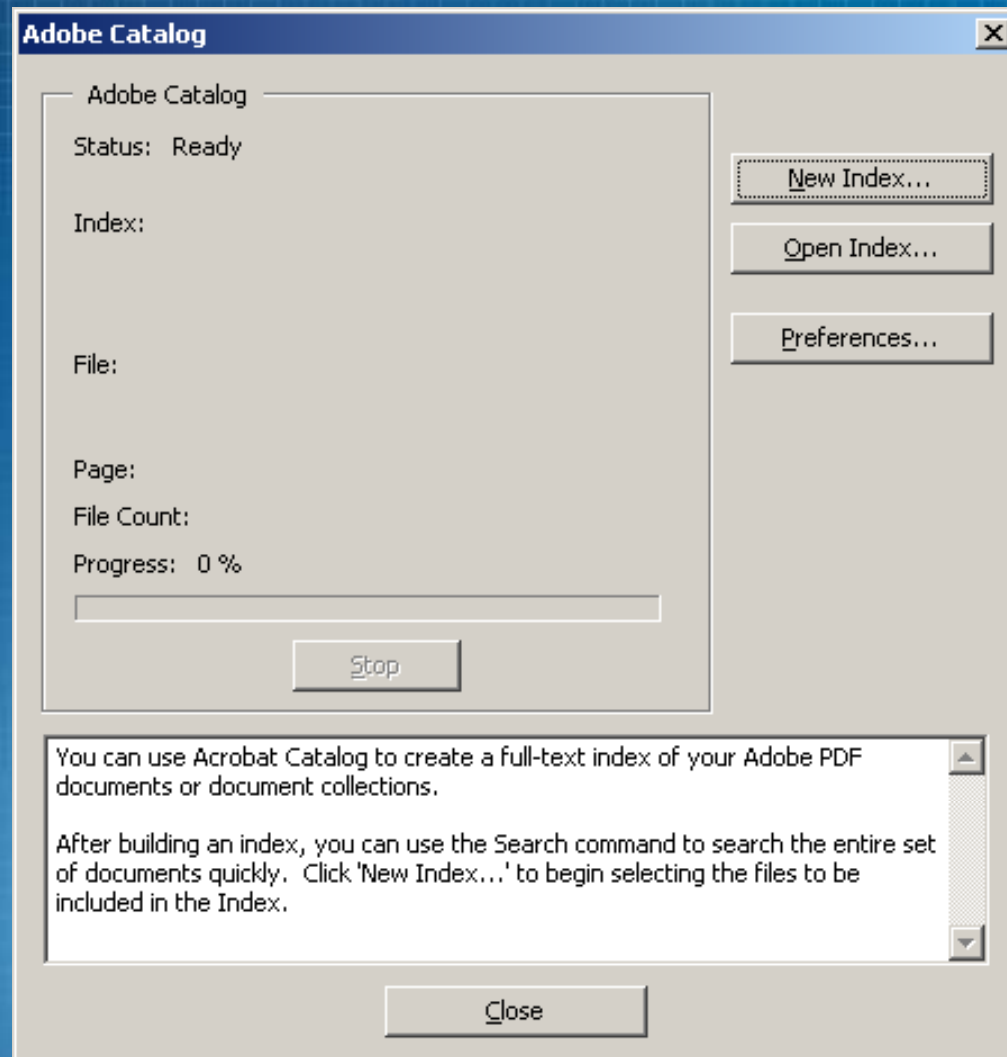


My experience with Acrobat and Catalog...

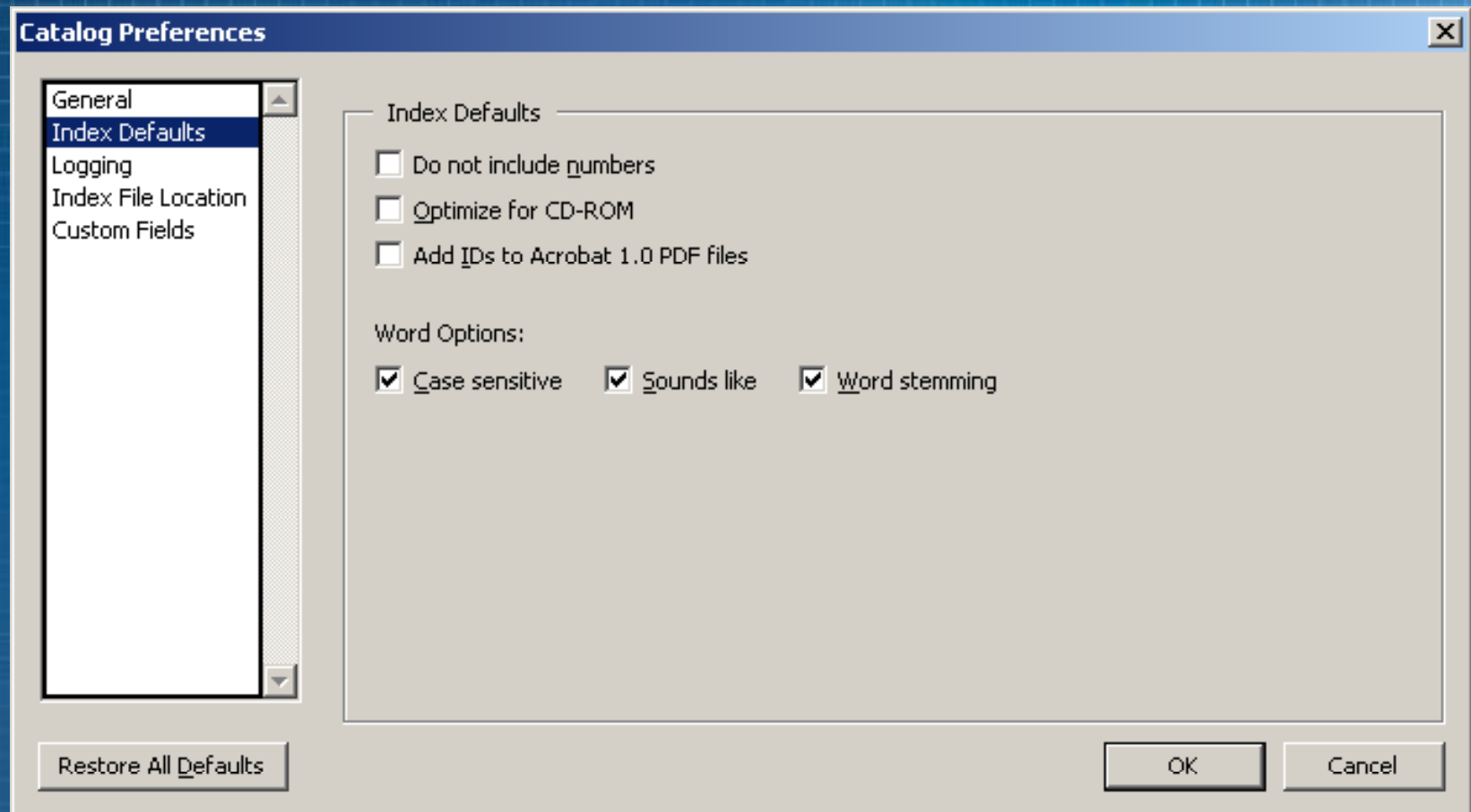
- We started 4 years ago with Acrobat 4.0
- Upgraded to 5.0 about 3 years ago
- Tried to upgrade to 6.0 - but problems...
 - Search dialog was “redesigned” and complicated
 - Search fails on large indexes
 - Boolean search doesn’t work the same
 - Adobe support advised: “stay on 5.0!”
- Staying with 5.0 until?
 - Problem: Adobe supports *only* current and previous releases



Let's create an index...



Determine Preferences...



Define the Index...

New Index Definition [X]

Index File: UNTITLED Save

Index Title: Save As...

Index Description:
 Options...

Include These Directories

Add...
Remove

Exclude These Subdirectories

Add...
Remove

Build Purge Cancel

Define Index Options...

Options [X]

Words not to include in the index

Word:

Do not include numbers

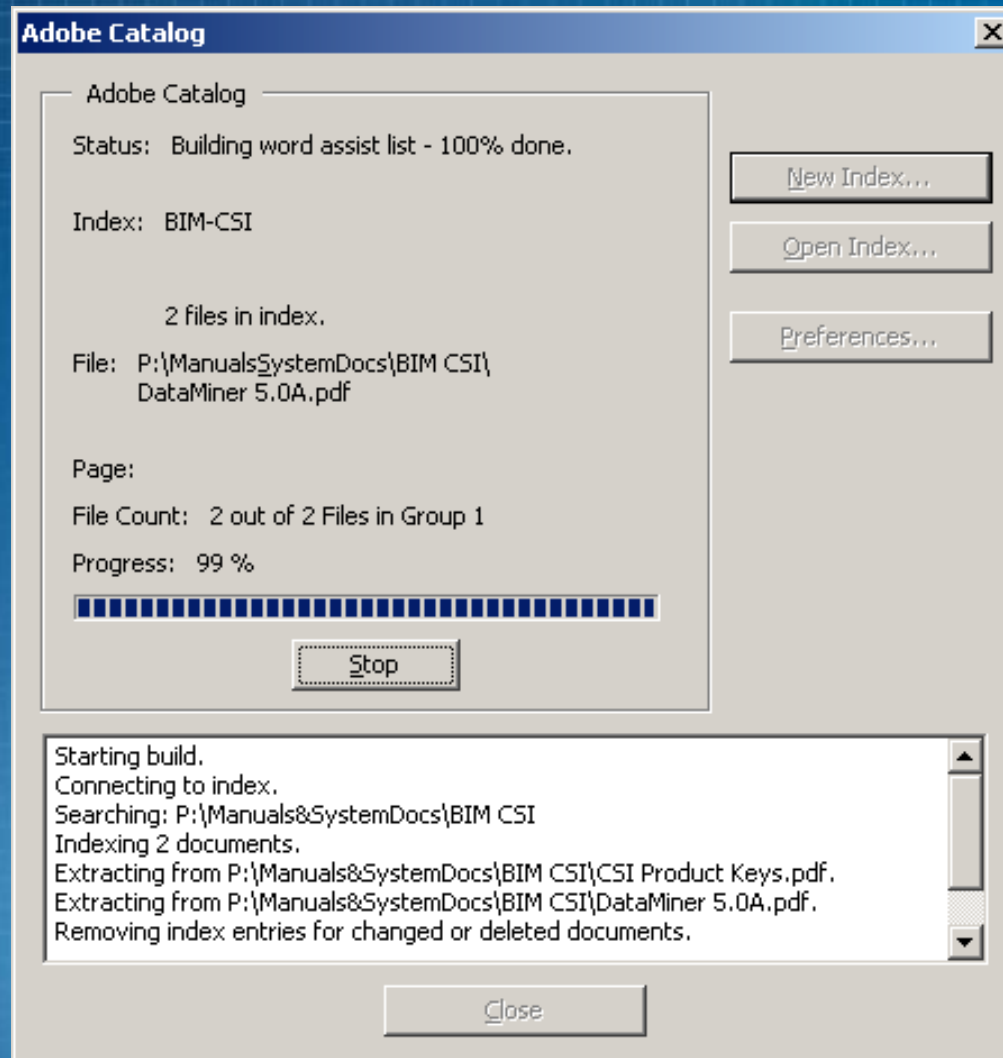
Word options

Case sensitive Sounds like Word stemming

Optimize for CD-ROM

Add IDs to Acrobat 1.0 PDF files

Build the Index...



Execute a Search...

Adobe Acrobat Search [X]

Find Results Containing Text

[Search] [Clear] [Indexes...]

With Document Info

Title []

Subject []

Author []

Keywords []

With Date Info

Created after [/ /] before [/ /]

Modified after [/ /] before [/ /]

Options

Word Stemming Thesaurus Match Case

Sounds Like Proximity

Searching 2 out of 34 indexes.

Index Selection

Available Indexes

- Acrobat 5.0.5 Online Guides
- BIM-CSI
- BSI Manuals
- DB2/VSE 7.1 Manuals
- DB2/VSE 7.3 Manuals
- Debug Tool
- DF Sort 3.4
- DITTO

(currently unavailable indexes are grayed out)

[OK] [Add...] [Remove] [Info...]

Execute a Search...

Adobe Acrobat Search [X]

Find Results Containing Text

file AND sharing

Search

Clear

Indexes...

With Document Info

Title

Subject

Author

Keywords

With Date Info

Created after / / before / /

Modified after / / before / /

Options

Word Stemming Thesaurus

Sounds Like Match Case

Proximity

Searching 4 out of 34 indexes.

Search Results...

Search Results Found 22 out of 30 documents.

Score Title

- Master Index and Glossary
- Performance Tuning Handbook
- CYGNET Reference Guide
- Installation and Configuration Guide
- Database Services Utility
- Data Restore Guide
- System Administration
- Database Administration
- Operation
- Program Directory DB2/VSE 7.3
- Opti-Cache
- Control Center Operations Guide for VSE
- C3381
- Messages and Codes
- Interactive SQL Guide and Reference
- ECHO User Reference Guide
- Quick Reference
- SQL Reference

P:\Manuals\SystemDocs\Db2Vse73Pdf\Master Index and Glossary.pdf

Adobe Acrobat - [Performance Tuning Handbook.pdf]

File Edit Document Tools View Window Help

151%

Bookmarks
Thumbnails
Comments
Signatures

1. Make sure you have read access to the VMSES/E code (MAINT 5E5 disk) and read/write access to the Software Inventory disk (MAINT 51D) or SFS directory.
2. Establish the access order.

```
vmfsetup 5697F42X {DB2VM|DB2VMSFS}
```

5697F42X is the PPF that was shipped with the product. If you have your own PPF override, substitute that name for 5697F42X shown in this command. You also need to substitute your PPF name in the VMSES/E commands in any subsequent steps.

Use DB2VM for installing on minidisks or DB2VMSFS for installing in **Shared File System** directories.
3. Rebuild DB2 Server for VM DBSS member or component in the ARISQLLD LOADLIB. You must do both steps 3a and 3b.
 - a. Rebuild the ARISQLLD LOADLIB.

```
vmfbld ppf 5697F42X {DB2VM | DB2VMSFS} ARIBLLD ARISQLDS (a11  
vmfview build
```

Use DB2VM for installing on minidisks or DB2VMSFS for installing in Shared File System directories.

ARIBLLD is the name of the VMSES/E build list used to build the ARISQLLD LOADLIB.

Review the build message log (\$VMFBLD \$MSGLOG). If necessary, correct any problems before you continue. Use the F2 key, ALL, to review all of the messages.

200 of 263 8.5 x 11 in

COMPUTER CREDIT, INC.

Adobe Acrobat - [Adobe Acrobat Help]

File Edit Document Tools View Window Help


150%

Searching and Indexing Document Collections

Searching indexes

The Adobe Acrobat Search command allows you to perform full text searches of PDF documents and collections that have been indexed with the Acrobat Catalog feature. The Search command is more powerful and flexible than the Find command. It lets you search multiple documents, and define advanced query criteria. Search is faster than Find because it reads the index rather than the entire document.

To search an index created using Adobe Acrobat Catalog, you first select the indexes to search, define a search query, and view the occurrences of the search term within the documents you selected to review. You can also select the documents to review from those returned by the search. A *search query* is an expression made up of text and other items to define the information you want to find.

 Opening a PDF document associated with an index automatically makes the index searchable.

Selecting indexes

You can search any or all indexes displayed in the Index Selection dialog box. Dimmed indexes are not available for searching.

To add or remove indexes to search:

- 1 Choose Edit > Search > Select Indexes to list the currently available indexes and to add or delete indexes, and then do one of the following in the Index Selection dialog box:

215 of 287 8.25 x 11 in



Limitations, Problems or Gotcha's with Catalog...

- Logical size of index
 - Number of documents in index
 - Number of “words” in each document
 - Include numbers?
 - Other search options at index creation
- Possible circumvention...
 - Keep it granular!
 - Group documents in folders
 - One index per folder
 - Make use of stop word list
 - Exclude numbers where possible





Limitations, Problems or Gotcha's with Catalog...

- Numbers in the text
 - Option to include or exclude
 - No distinction between dollar amounts and numeric strings
 - Can greatly inflate index size unnecessarily
- Possible circumvention...
 - Exclude numbers where possible
 - Or... modify report to append an alpha character to number





Limitations, Problems or Gotcha's with Catalog...

- Stop word list
 - Cannot be re-used
 - Cannot be copied from index to index
- Possible circumvention...
 - None - must re-key the words for each index
 - Enhancement for this submitted years ago but no response - *yet*





Limitations, Problems or Gotcha's with Catalog...

- Adding/deleting documents in existing index
 - Catalog is “smart”
 - Identifies new documents, only indexes them and merges into existing index
 - If done a lot over a period of time, will grossly inflate size of index
 - May slow index access due to “dead space”
- Resolution...
 - Purge and rebuild the index on a scheduled basis





Limitations, Problems or Gotcha's with Catalog...

- Internal and External names don't match
 - Catalog uses the internal name
 - Other dialogs reference external (physical) name
 - Can be confusing
- Resolution...
 - Update internal name in Document Properties so that both names match





In Summary...

- Acrobat Professional 5.0/Reader 5.0 has worked very well - *even* with the limitations
- Adobe support has been a little disappointing
- No response to enhancement requests
- Searching with the indexes is very fast
- A good investment and solution for our needs
- Concerned about new releases!





For More Information...

- Adobe's web site...
 - <http://www.adobe.com/products/acrobat/main.html>
- Or feel free to contact me:
 - mrosinger@cciws.com
 - 1-336-761-1524



COMPUTER CREDIT, INC.