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 Adobe™ 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!

COMPUTER CREDIT, INC.

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

COMPUTER CREDIT, INC.

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

COMPUTER CREDIT, INC.

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?

COMPUTER CREDIT, INC.

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

COMPUTER CREDIT, INC.

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)

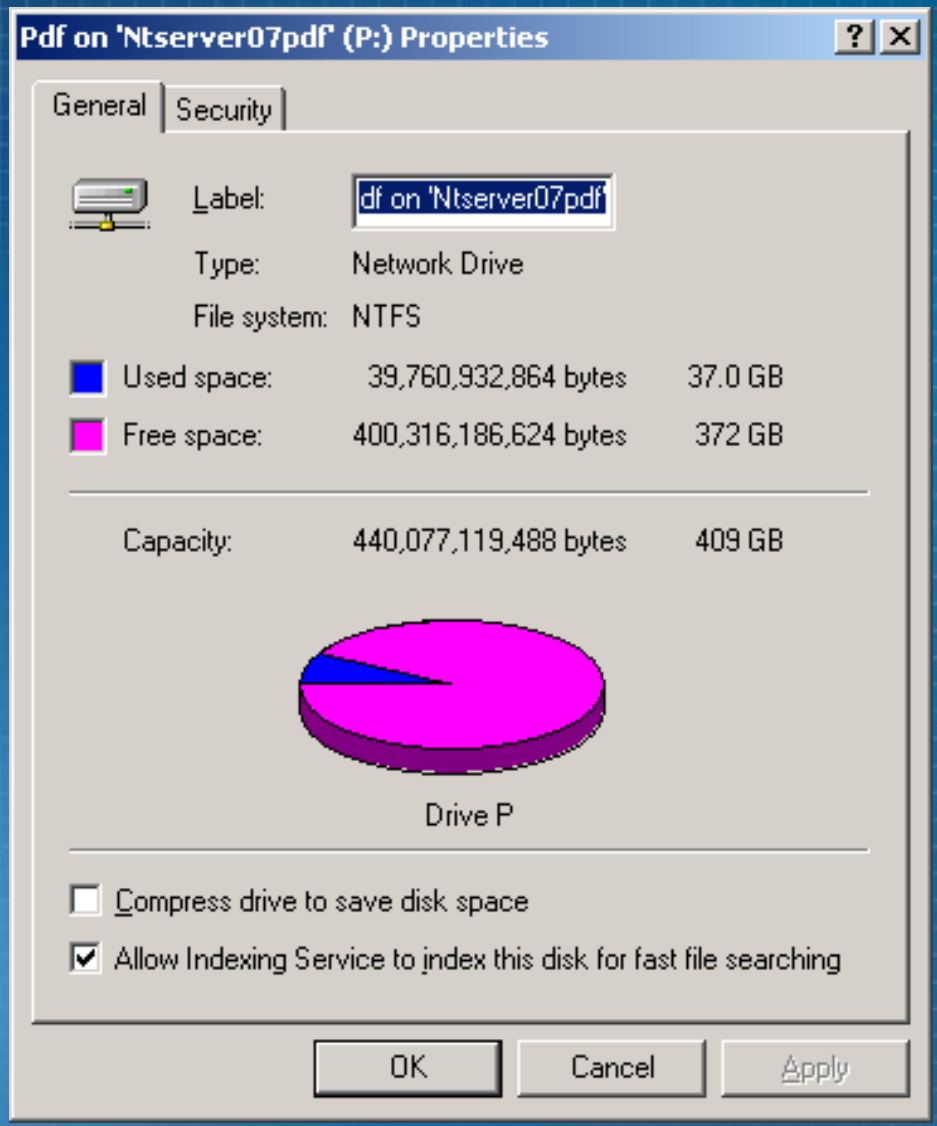
COMPUTER CREDIT, INC.

Our PDF Server...

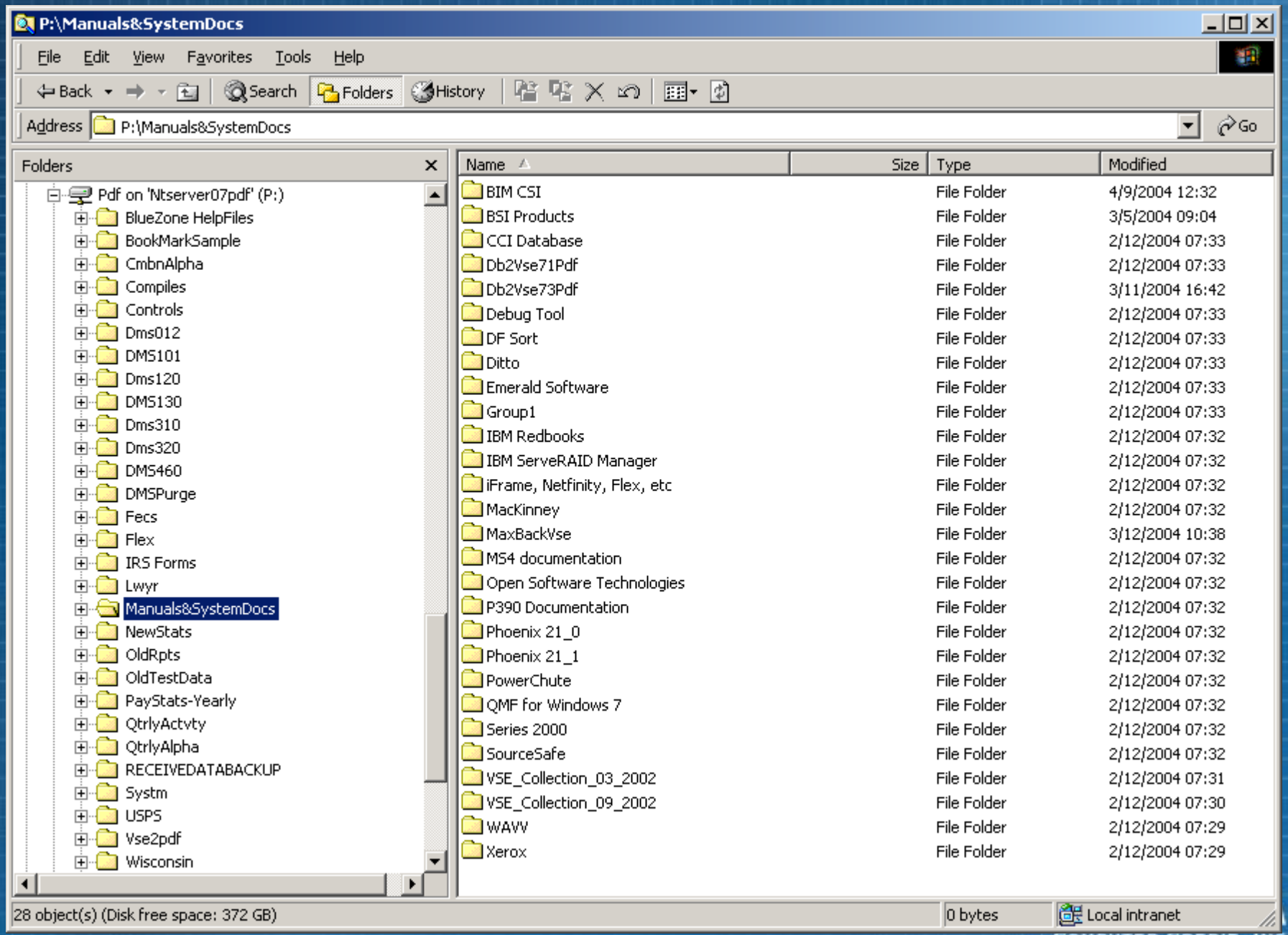
- Recently migrated to new (2nd) server
- Dell PowerEdge 2600
  - Windows 2003 Standard Server
  - 2.4 Ghz Xeon Processor
  - RAID 5 with 4-146 GB SCSI drives, for PDF storage
  - 2-36 GB drives, mirrored for Windows O/S
  - Redundant power supplies
- Cost about $7,800

COMPUTER CREDIT, INC.

Our PDF Server...



COMPUTER CREDIT, INC.



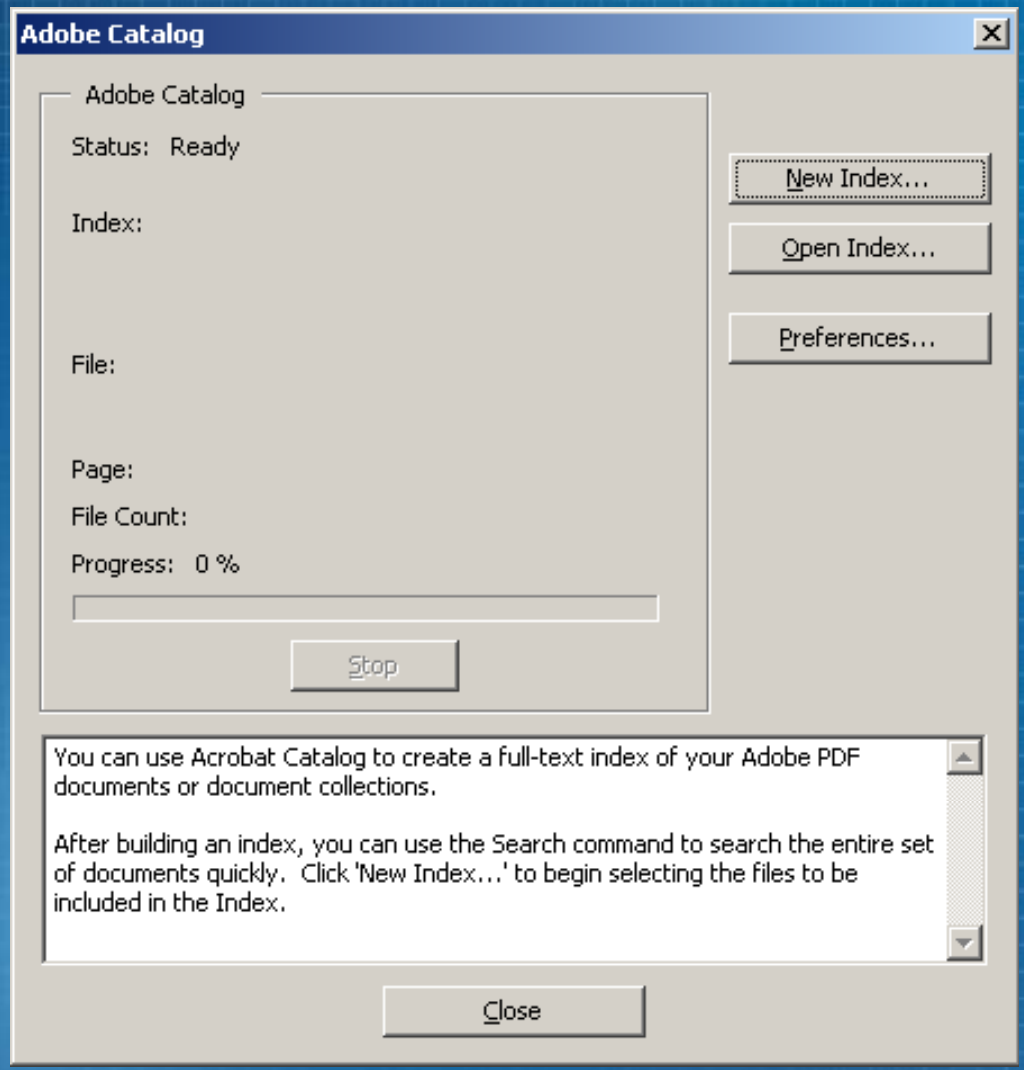
COMPUTER CREDIT, INC.

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

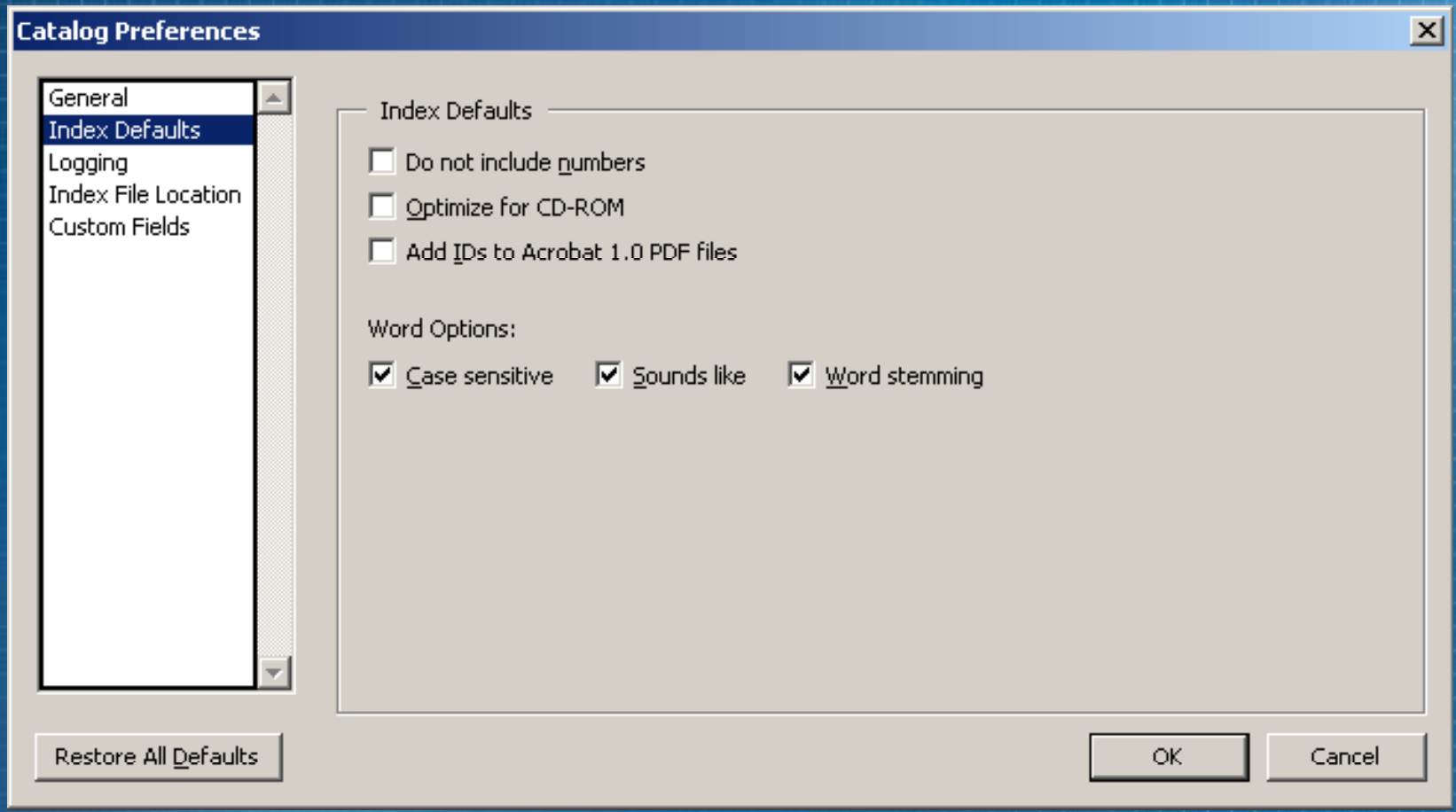
COMPUTER CREDIT, INC.

Let's create an index...



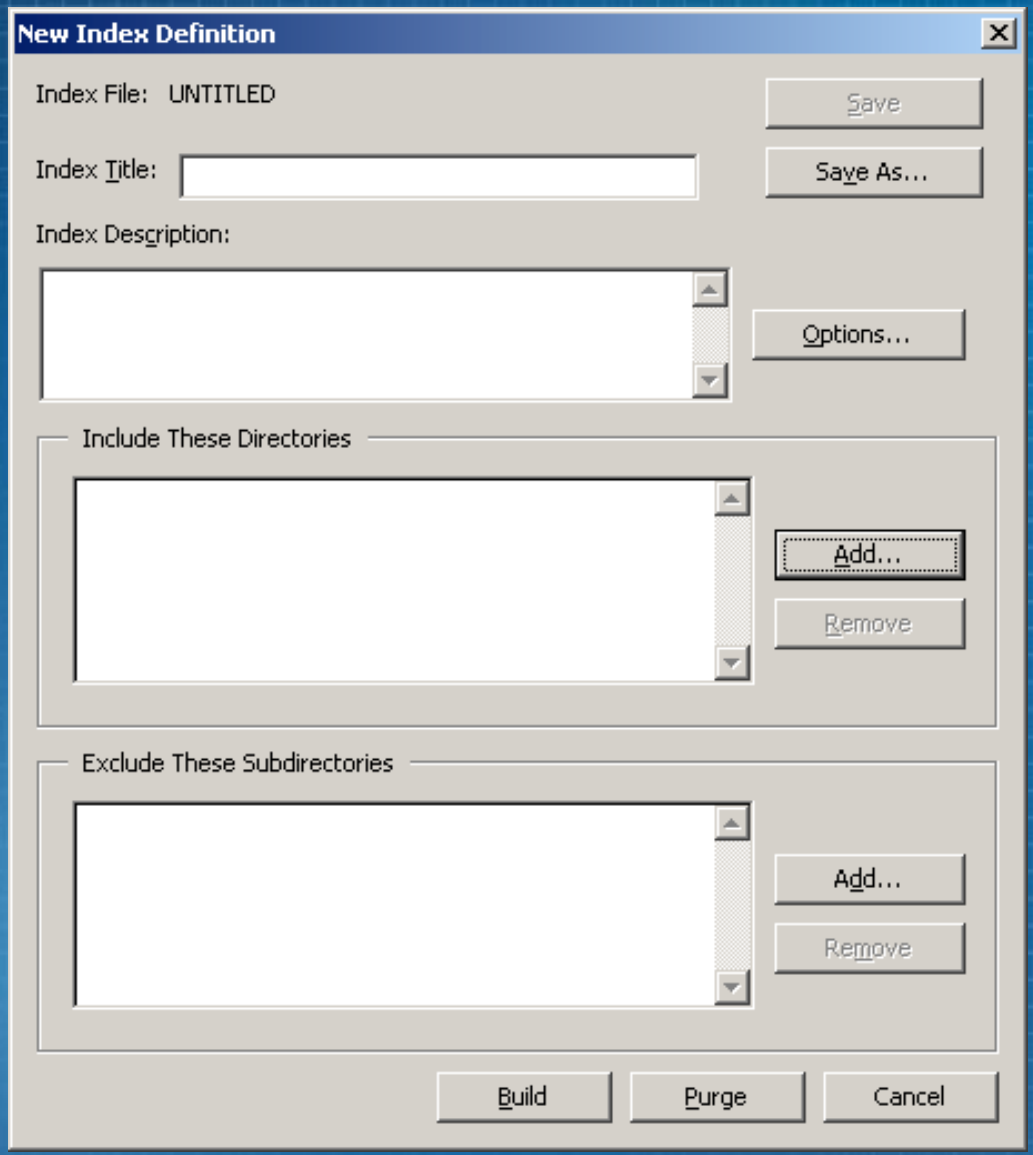
COMPUTER CREDIT, INC.

Determine Preferences...



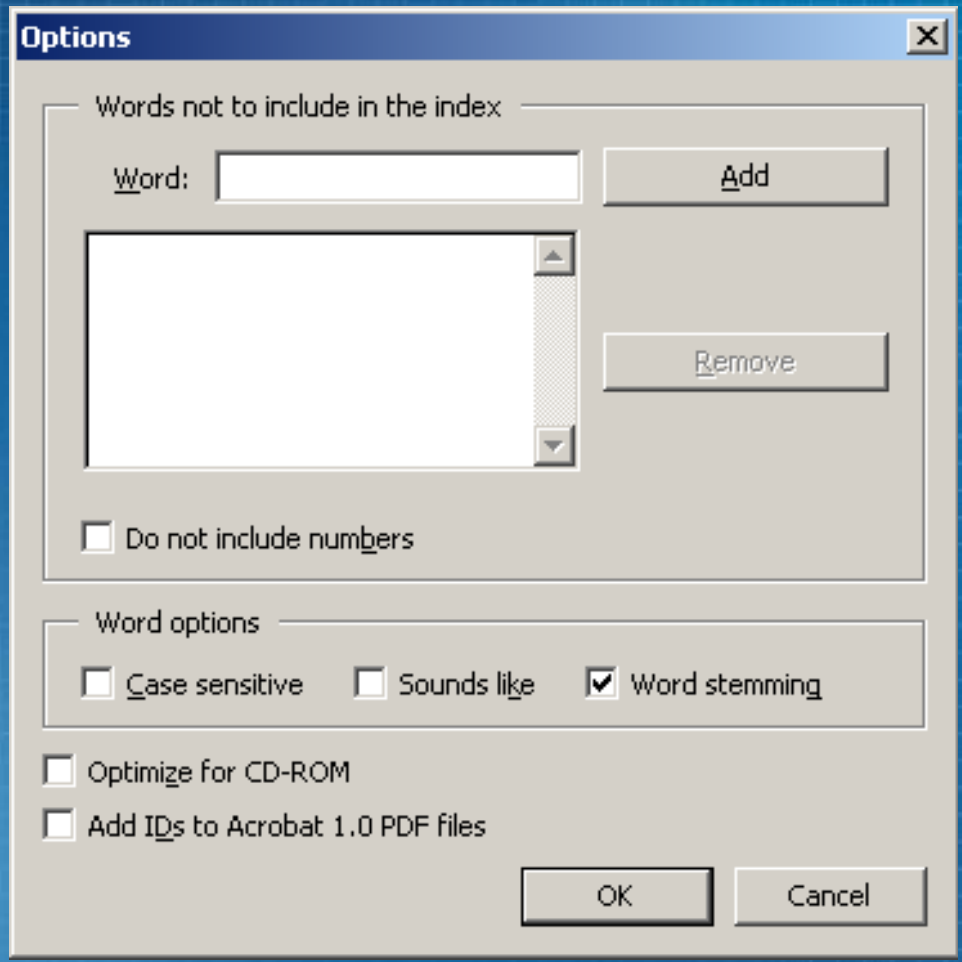
COMPUTER CREDIT, INC.

Define the Index...



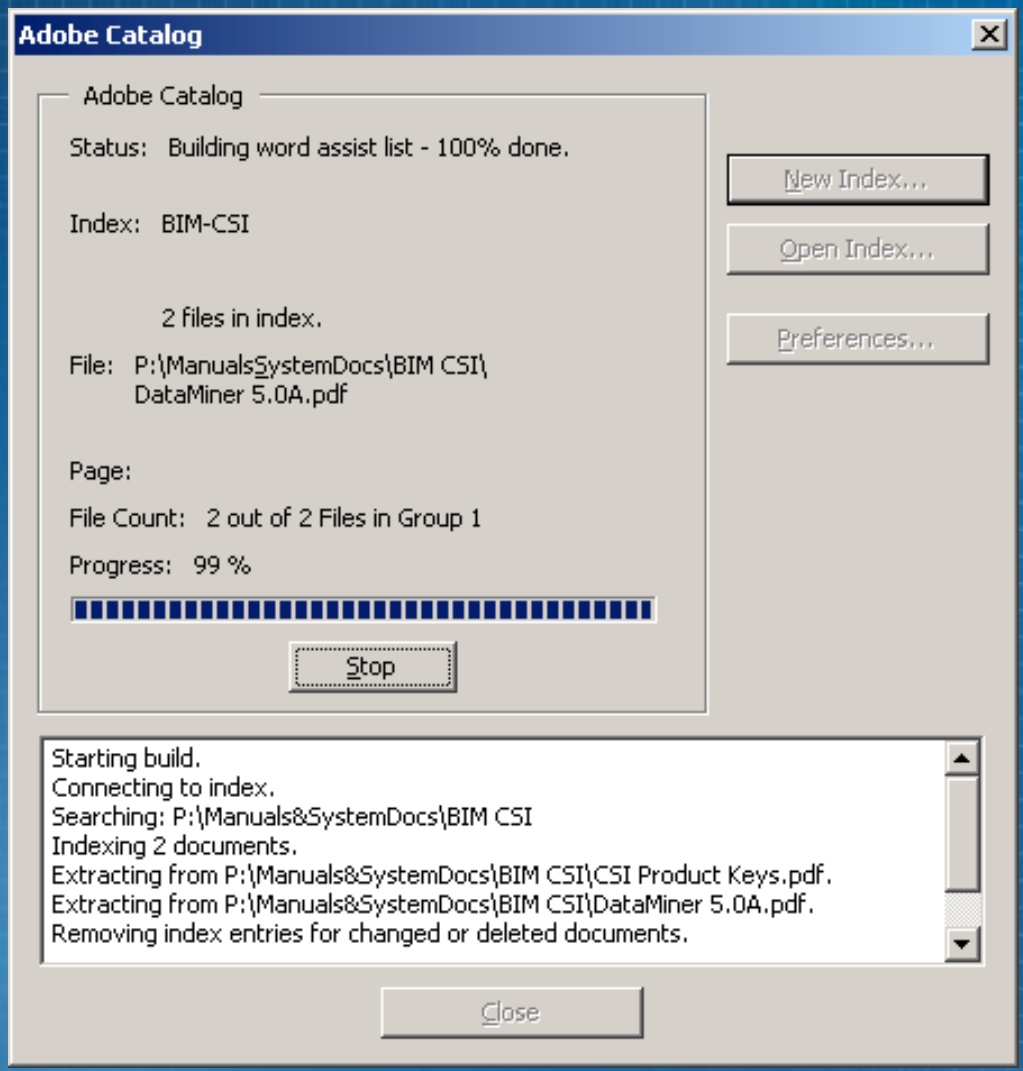
COMPUTER CREDIT, INC.

Define Index Options...



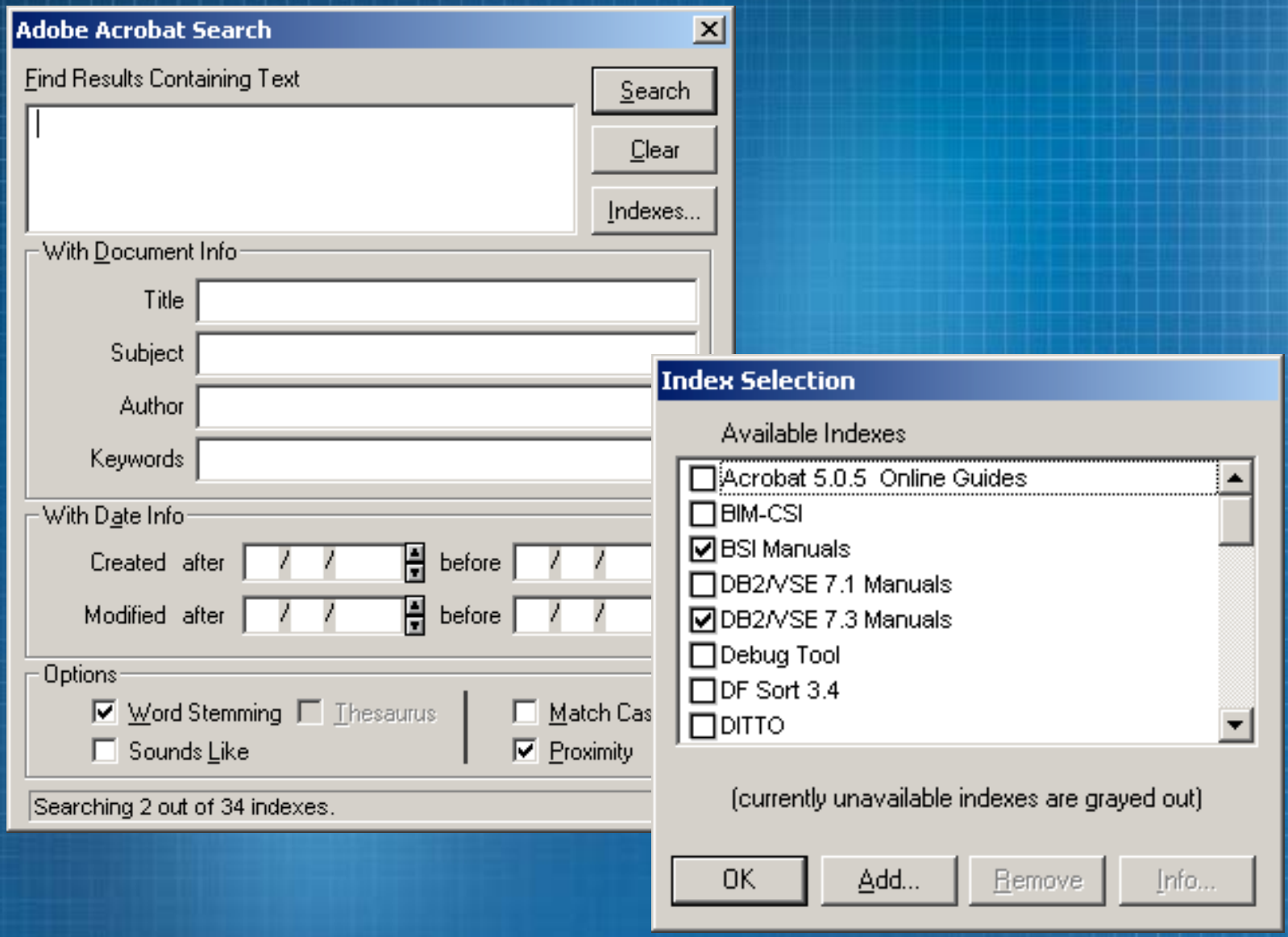
COMPUTER CREDIT, INC.

Build the Index…



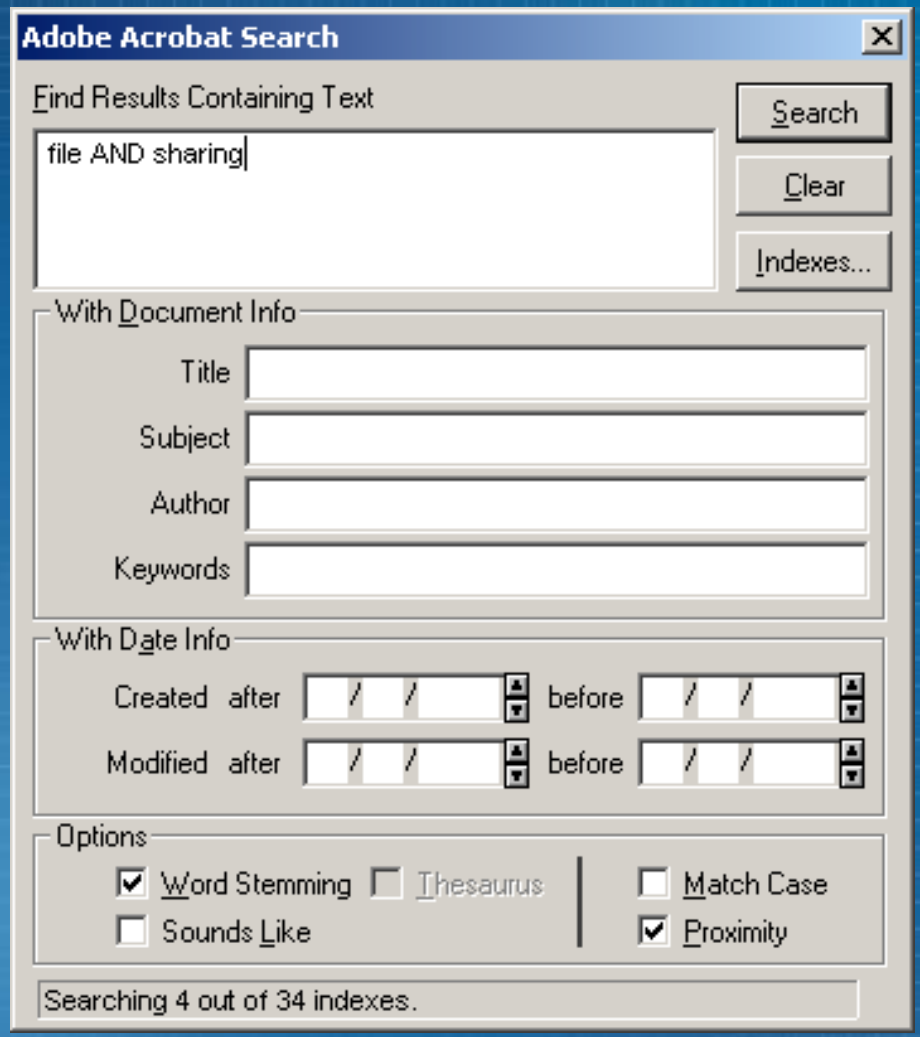
COMPUTER CREDIT, INC.

Execute a Search...



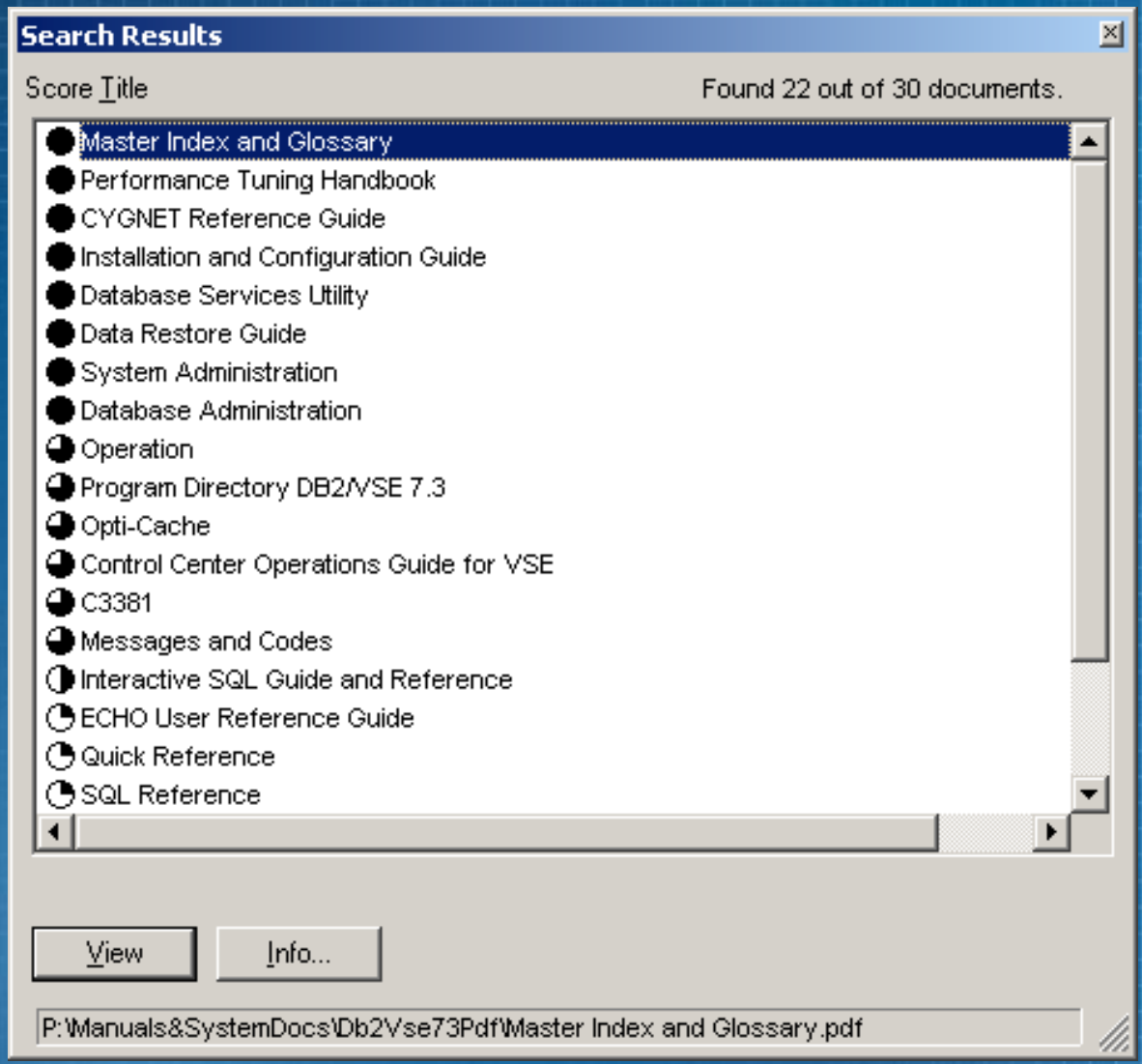
COMPUTER CREDIT, INC.

Execute a Search...



COMPUTER CREDIT, INC.

Search Results...



COMPUTER CREDIT, INC.

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.

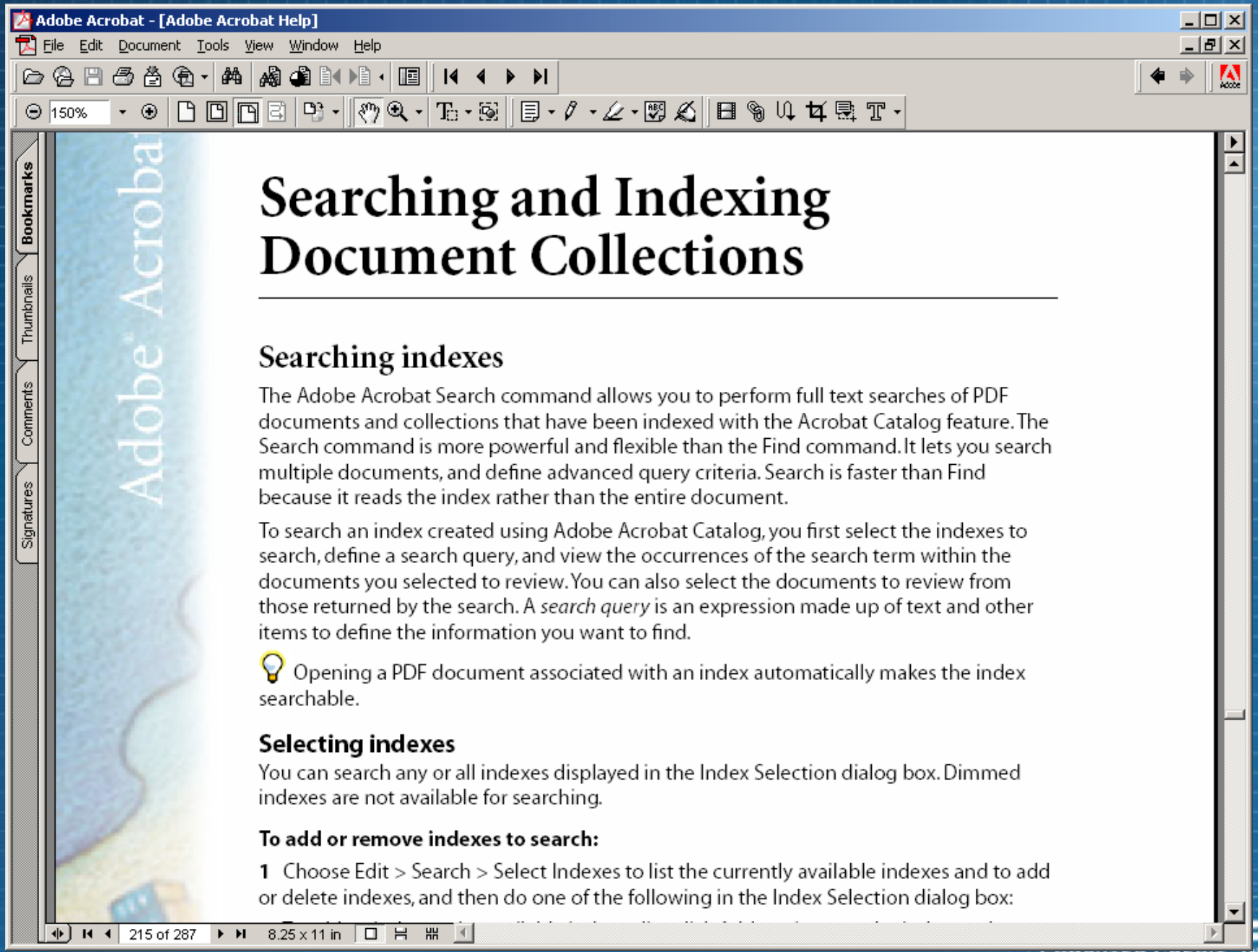
      ```
      vmfbld ppf 5697F42X {DB2VM | DB2VMSFS} ARIBLLLD ARISQLDS (all
      vmfview build
      ```

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

      ARIBLLLD 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%
Bookmarks
Thumbnails
Comments
Signatures
Adobe Acrobat
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

COMPUTER CREDIT, INC.

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

COMPUTER CREDIT, INC.

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*

COMPUTER CREDIT, INC.

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

COMPUTER CREDIT, INC.

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

COMPUTER CREDIT, INC.

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!

COMPUTER CREDIT, INC.

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.