



A CONSOLIDATED GRAPHICS COMPANY

Development of a Custom PDF Search (on CD-ROM)

September 27, 2005

Presented by Mark Czajka
Director of New Technology

mczajka@ags.com

Questions to ask...

- How do you capture data, such as abstracts, and apply metadata to PDF document information fields?
- Is the standard full text search acceptable? Is it easy for the users of the CD to use?
- What more can be done to allow the user to drill down through your data?
- What should you watch out for when compiling a large number of PDFs, some coming from different sources?

SPE Case Study

- Abstracts and Papers collected and reviewed via AGS Total Meeting Organization (TMO) website
- 1500 abstracts submitted; 732 papers selected
- TMO System generates text, HTML and PDF files; ColdFusion, SQL Server and activePDF used in development
- SQL database used to build Custom CD Search
- Metadata imported with ISIToolbox; Form Fields and JavaScript used to build Search interface

Full text search picks up all hits...

The screenshot shows the Adobe Reader interface with a search results panel on the right. The main window displays the cover of the ANTEC 2005 Annual Technical Conference. The search panel shows the following information:

- Finished searching for: **Baba**
- Finished searching in: **In the index named ANTEC 2005.pdf**
- Documents found: **6**
- Total instances found: **10**

The search results are listed as follows:

- Effect of Flow Behavior on Internal Structure
 - Satoko **Baba**, Yew Wei Leong and Hiroyuki
 - S. **Baba**, S. Takashima, M. Mizoguchi, T. Ki
 - S. **Baba**, S. Takashima, M. Mizoguchi, T. K
 - S. **Baba**, S. Takashima, H. Hamada, ANTE
- Evaluation of Adjacent Flow Weldline in PC/AE
 - Satoko **Baba**, Umaru Semo Ishiaku, Hiroyu
- Table of Contents
 - S. **Baba**, H. Hamada, U. Ishiaku, Kyoto Ir
 - S. **Baba**, H. Hamada, Y. Leong, Kyoto Ins
- Scratching Behaviors in Injection Molded Prod
 - Baba**, S., Takashima, S., Leong, Y.W., H
- Morphology of PC/ABS Blend Systems
 - Baba**, S., Takashima, S., Leong, Y.W., H

The search panel also includes a 'New Search' button, a 'Sort by: Relevance Ranking' dropdown, a checked 'Collapse file paths' option, and buttons for 'Refine Search Results', 'Done', and 'Find a word in the current PDF document'.

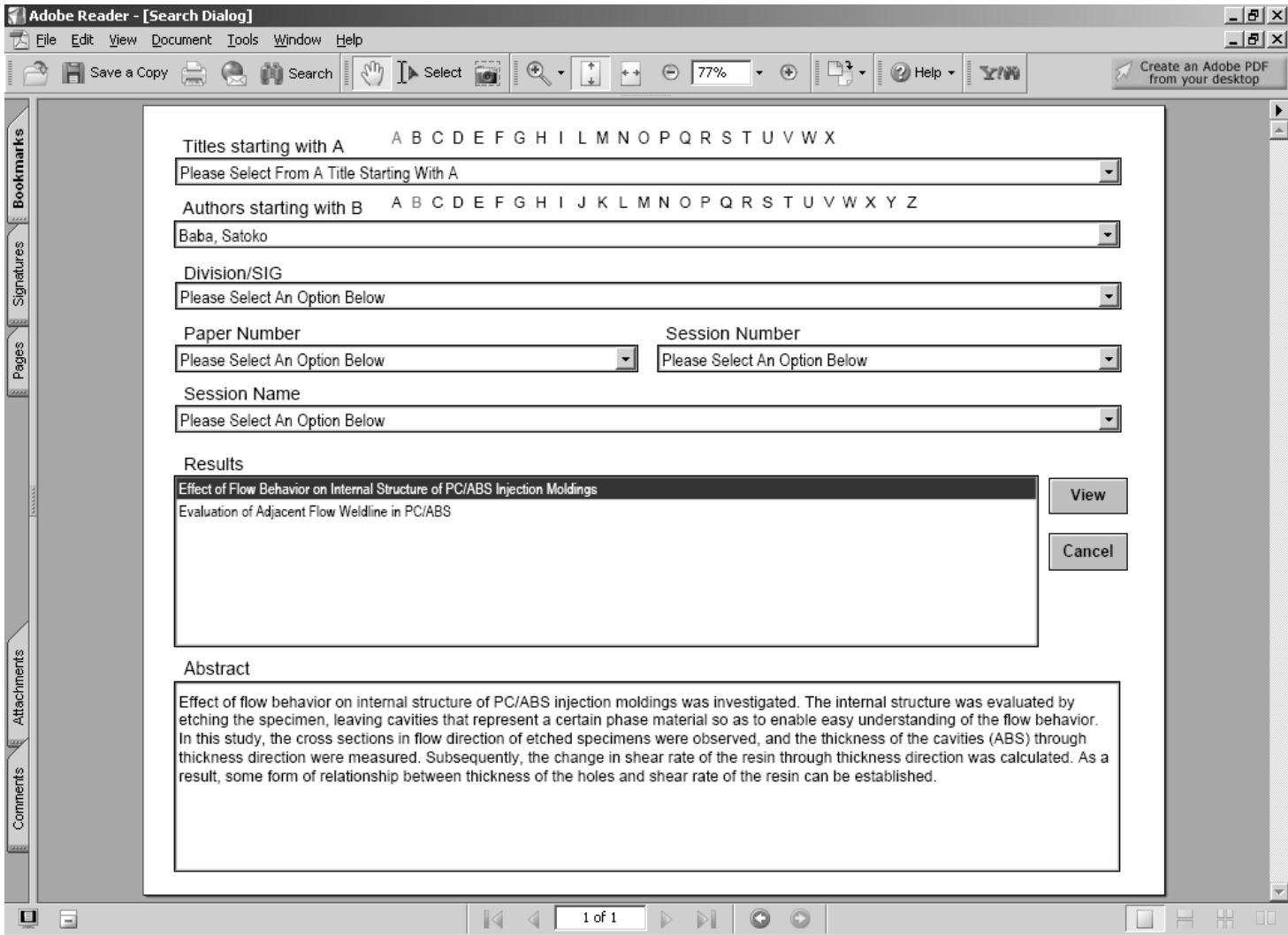
perhaps more than you want.

Advanced Search is better...

The screenshot displays the Adobe Reader interface with a PDF document open. The document content is for the ANTEC 2005 Annual Technical Conference, held from May 1-5 at the Hynes Convention Center in Boston, Massachusetts. The document includes search options such as 'Full Text Search', 'Table of Contents', 'Additional Search Options', and 'Using this CD-ROM'. It also provides contact information for the Society of Plastics Engineers, including the website www.anteec.ws and www.4spe.org. The search panel on the right is open, showing a search query field, a dropdown menu for 'Return results containing:' set to 'Match Exact word or phrase', and a dropdown for 'Look In:' set to 'In the index named ANTEC 2005.pdx'. Under 'Use these additional criteria:', the 'Author' dropdown is set to 'Baba' with the 'Contains' operator. Other criteria include 'Is exactly' and 'Whole words only' (checked). The search panel also includes a 'Search' button and links to 'Search the internet using Yahoo!', 'Use Basic Search Options', and 'Find a word in the current PDF document'.

but it is more difficult to use for the average user.

Custom Search - Advanced



Custom Search - Basic

Adobe Reader - [Menu]

File Edit View Document Tools Window Help

Save a Copy Search Select 123% Help Adobe Reader 7.0

Signatures Pages Attachments Comments

2005 Cable Network Profiles

Search by:

- Demographic
 - Persons
 - Men
 - Women
 - W18-24
 - W25-34
 - W35-49
 - W50+
 - Kids
 - W18-34
 - W18-49
 - W25-54
 - W35-64
 - W35+
- Program
- Keywords

TABLE OF CONTENTS

- How To Use
- Network Profiles
- Local Avails:
 - By Network*
 - By Minute*
- CAB Resources

OneTV World

Helps you identify ad-supported cable networks that target desired audiences

1 of 1

Search Results

The screenshot shows the Adobe Reader interface with a search results panel on the right. The main document is a PDF titled "2005 Cable Network Profiles" with a "TABLE OF CONTENTS" section listing "How To Use", "Network Profiles", "Local Avails" (with sub-options "By Network" and "By Minute"), and "CAB Resources". The document features a large graphic of silhouettes of people standing in front of a screen displaying the "OneTV World" logo. Below the graphic, it says "Helps you identify ad-supported cable networks that target desired audiences".

The search results panel on the right shows the following information:

- Finished searching for: **W35-49**
- Finished searching in: **In the index named index6.pdf**
- Documents found: **13**
- Total instances found: **0**

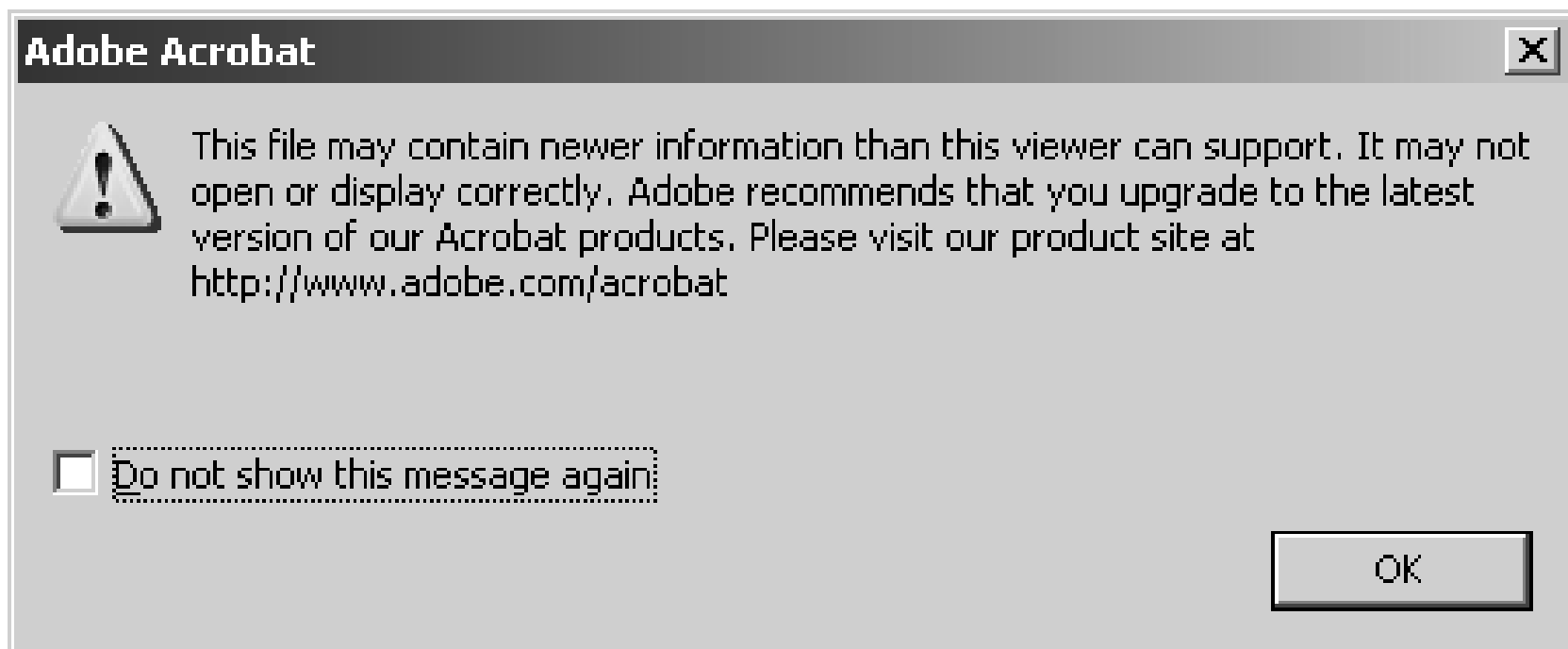
A "New Search" button is visible below the statistics. The "Results:" section lists 13 cable channels:

- Bloomberg Television
- Casa Club TV
- CNBC
- CNN
- CNN en Español
- CNN Headline News
- Court TV
- The Golf Channel
- HGTV
- National Geographic Channel
- News 14 Carolina
- NewsChannel 8
- WE: Women's Entertainment

At the bottom of the results panel, there is a "Sort by:" dropdown menu set to "Relevance Ranking", a checked checkbox for "Collapse file paths", and three buttons: "Refine Search Results", "Done", and "Find a word in the current PDF document".

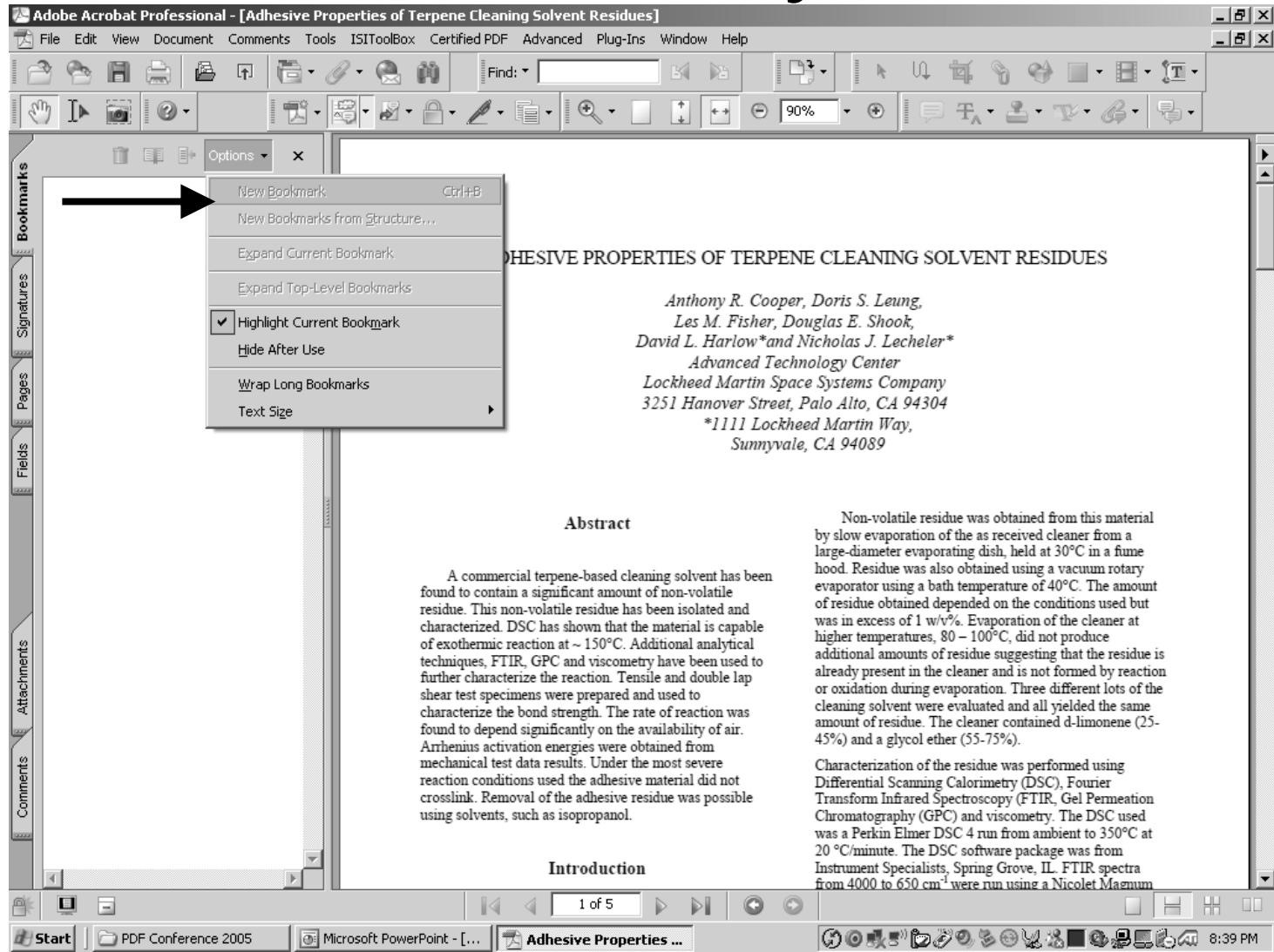
PDF Version

PDFs coming from different sources may be created in different versions of Acrobat.



Use the PDF Optimizer to make your PDFs compatible with Acrobat 5 or higher.

Security



How do you add bookmarks to a PDF that is secure?

Fonts

- Supplied PDFs may not have embedded fonts (an issue if you are printing a book)
- Supplied PDFs may not be searchable (an issue if you are developing a CD)
- Using Batch with PDF Optimization could unembed certain fonts – Be Careful!

Tools mentioned in this Case Study

- activePDF (www.activePDF.com)
- ISIToolbox (www.imagesolutions.com)
- TMO (www.tmo-system.com)

Adobe Reader Distribution Agreement

- Visit Adobe's website to fill out the online request to distribute Adobe Reader.
- <http://www.adobe.com/products/acrobat/acrrdistribute.html>

