

PAC Scanning Project

Sara Webber

2002 JACoW Meeting

Berkeley, CA



Scanning Project Goal

- Scan all PAC proceedings that are not currently on the web (1965-1993) and create electronic proceedings for the JACoW website. All proceedings will be fully JACoW compatible—hidden fields will be entered, pages will be cropped to 595 x 792 pt and scans will be fully text searchable.

Short-Term Project Goal

- This is what we are doing with currently available funds. . .
 - Finish the electronic proceedings for PAC 1993-1987 and make them available on JACoW by 12/20/02.

Progress to Date

- PAC 1991 finished and posted on JACoW
- PAC 1991 data sent to SPIRES
- PAC 1993 finished
- PAC 1989 all data entered and TOC verified
- PAC 1987 all data entered and TOC verified
- PAC 1967 all data entered and TOC verified

Labor per Proceedings

- 2 FTE (with project manager guidance) can finish a proceedings in 1 month*. Note that this does not include scanning time.

*We expect the time to finish a proceedings to go down as we move into the older proceedings, which are much shorter.

Current Average Cost per Proceedings

- Total average cost = \$6500 per proceedings
 - Labor = \$2000 per proceedings (includes database and data entry work)*
 - *Project manager salary is donated
 - Scanning = \$4500 per proceedings (includes scanning twice—once for PDF and once for OCR text)*
 - *Again, we expect costs to go down as proceedings get shorter

Funding Sources/Contributions

- US Department of Energy Grant = $\frac{1}{2}$ total funds to date
- PAC 2001 = $\frac{1}{2}$ total funds to date
- Fermilab = Donated manager time, office space, computers and other supplies
- HBar Technologies = Donated data entry time

Steps in Preparing Proceedings

- Scan—done by outside vendor
- Check scans—employees
- Enter data—employees
- Add keywords—database script
- Create preliminary proceedings—database script
- Resize pages and add hidden fields—WinBatch script

Steps in Preparing Proceedings

- Add thumbnails and set open options—Acrobat batch processor
- Add conference stamp to paper—Impress Pro batch processor
- Add copyright stamp to IEEE copyrighted papers—employees (using Impress Pro)
- Proofread database-created table of contents—employees
- Check database-created author index against printed book—employees

Steps in Preparing Proceedings

- Final web proceedings—database scripts
- Create cover page and add front-matter—project manager
- Check all links—employees & automated link check program
- Prepare CD for posting on JACoW—project manager
- Prepare SPIRES catalog file—database scripts

Scans

- All proceedings are scanned by an outside vendor at a current cost of \$1.48 per page
- This includes:
 - Scanning to 400 dpi TIFF and using Acrobat Capture Server to create image + text PDF file
 - Scanning to 400 dpi TIFF and using OCR software to create a text file
 - Providing an Excel index of scanned files
 - Providing TIFF, PDF and OCR text files on CD

Database

- Database design is basically the same as PAC 2001
- Added a conference tag so that multiple conferences could be housed in one database

Add Keyword Script

- Database parses through OCR text files looking for keywords from the keyword table. It adds the five most frequently occurring keywords for each paper to the corresponding abstract record in the database.

Create Web Proceedings Script

- Same as PAC 2001 script
- Creates table of contents, author index, keyword index and session index

WinBatch Script

- Leif Liljeby's PAC 2001 WinBatch script was modified to work with Acrobat 5.0 and the commands for adding the conference stamp and page numbers were removed. This script is driven by a database-generated text file.

SPIRES Catalog File Script

- A PL/SQL script extracts information from the database into a text file, which is formatted as requested by the SPIRES database manager.

Points for Discussion

- What is the best way to inform the community that this service is available?
- How can we track the service's usefulness to the community?
- A future project to add conference stamps to all JACoW papers? (Gerry Jackson's idea)
 - LINAC, EPAC, APAC 1998, PAC 1997, EPAC 1996 and PAC 1995