

Building the Table of Contents (TOC) and the Author Index (AI)

Martin Comyn

TRIUMF

Talk Outline

- The database
- Procedures and tools — standards?
- Appearance — standards?
- TOC — standard format?
- AI — rationalized?
- Data required for the TOC and the AI
- Acquiring data for the TOC and the AI
- Manipulating data for the TOC and the AI
- Producing the TOC and the AI

The database

- The veracity of the database is of paramount importance
- The TOC and the AI are almost the last things to be produced:
 - finish producing the final versions of the papers first;
 - confirm the page counts;
 - determine the division of the volumes;
 - determine the ordering of the sections;
 - determine the ordering of the papers;
 - determine any section breaks;
 - allow for sections ending on odd or even pages;
 - determine the page numbers.

Procedures and tools — standards?

- Standards? What standards?
- Platforms and software used:
 - PAC'97
 - * Oracle / $\text{\LaTeX} 2_{\epsilon}$ / Perl
 - Open VMS / Digital UNIX / Linux
 - EPAC'98
 - * FileMakerPro / MS Word / Perl / Excel / Visual Basic
 - Mac / UNIX
 - PAC'99
 - * Oracle / MS Word / Perl
 - UNIX
 - Other conferences?
- The EPAC series benefits from the continuity of the editorial board.

Appearance — standards?

- Hardly any standards at present.
- The paper templates are converging.
- The CD-ROM structure differs.
- The WWW versions differ.
- The book versions differ with regard to the material placed before and after the papers in each volume, and the ordering of the papers.

The EPAC series have section dividers, the PAC series do not.

- PAC'97 has the worst mix of formats for the TOC and the AI:
 - WWW: kluge based on the APS abstract site
 - CD-ROM: produced by Sony Electronic Publishing Services
 - Book: Oracle / $\text{\LaTeX} 2_{\epsilon}$ / Perl

TOC — standard format?

- Very little similarity in structure.
- Overall summary of sections contained in volumes:
 - PAC'97: YES, complete summary in each volume.
 - EPAC'98: NO.
 - PAC'99: YES, in Vol. 2,3,4,5 only.
- Detailed TOC:
 - PAC'97: For each individual volume, 1,2,3.
 - EPAC'98: Complete, Vol. 1–3, in each volume.
 - PAC'99: Complete, Vol. 1–5, in volume 1 only.
- Division of topics:
 - PAC'97: By order presented in oral sessions (invited and contributed mixed) followed by posters.
 - EPAC'98: Invited by topic; contributed by topic; poster by topic.
 - PAC'99: Vol. 1 oral by topic; Vol. 2–5 poster by topic.
- Ordering of topics:
 - PAC'97 and EPAC'98 similar, PAC'99 different.
 - EPAC/PAC could be standardized, and APAC?
 - Editorial board autonomy a problem.

AI — rationalized?

- Rationalized listings:
 - PAC'97: YES.
 - EPAC'98: YES.
 - PAC'99: Partial.
- Ideally all papers for each author should appear as a single list on the AI, even if the format of the name differs on the actual papers.
- Problems:
 - Same person appears on several papers in various formats:
 - * PAC'97 had 1 author in 6 different guises.
 - * Mis-spelt names.
 - * Identical names:
 - ambiguity if only initials;
 - try to solve by affiliation;
 - complications if people move after performing work reported in paper or before/after the conference;
 - cannot always resolve;
 - procedure can only be automated to a limited degree.
 - Name format:
 - * initials + lastname;
 - * first name, initials + lastname;
 - * first name + lastname;
- Define an author name convention in the instructions?
 - Worth trying, but unlikely to succeed. Which format?

Data required for the TOC and the AI

- TOC:
 - volume titles;
 - section titles;
 - section order;
 - section breaks — location, number of pages;
 - paper titles;
 - author and affiliation lists;
 - paper order;
 - paper page counts;
 - paper first page numbers;
 - file names.
- AI:
 - rationalized author names;
 - associated paper first page numbers;
 - associated file names.

Acquiring data for the TOC and the AI

- Procedures:
 - At time of abstract submission:
 - * Enter the title and full author and affiliation list against the unique code for the abstract/paper.
 - Easier to manage if conference is in charge of abstract submission.
 - More problematic if abstracts are submitted to a third party such as the APS.
 - At time of paper submission before/at conference:
 - * check title entry — revise if necessary;
 - * check author list — revise if necessary;
 - * be prepared for surprises — one PAC'97 paper went from 1 to 64 authors;
 - * check affiliations — revise if necessary.
 - Depending upon the platform and software being used, include all mathematical notations, Greek letters and foreign language accents used in the titles, author names and affiliations.
 - Standardize the affiliation names:
 - * PAC'97 developed SQL code to reduce a large number of variants to a standard set with a consistent set of abbreviations.

Manipulating data for the TOC and the AI

- Precise details depend upon the software packages being used.
- The WWW, CD-ROM and book versions all need different variants of the data.
- Normally need to produce parallel databases, or alternate fields in the same database, which contain the same information in an alternate format (modified using SQL code or external Perl scripts) to conform to the final use.
- The main difference concerns the production of the WWW versions of the TOC and the AI which require different strings for the representation of mathematical notations, Greek letters and foreign language accents. Not all characters may be translatable.
- Either SQL code can be developed to produce a complete output file in the required format, or files with one line per paper or author can be generated for external manipulation by Perl scripts and other software packages.

Producing the TOC and the AI

Fully documented in:

POST MORTEM OF THE ELECTRONIC PUBLICATION OF THE
EPAC'98 PROCEEDINGS

L. Liljeby and J. Poole

SL-Note-99-015 DI

<http://www.cern.ch/EPAC/organisers/docs/postmort98.pdf>

Producing the TOC and the AI

TOC — PAC'97

Oracle/L^AT_EX 2_ε (book version)

Write a SQL procedure to DROP, CREATE and UPDATE a table with the data for the:

- chapter number;
- session chair;
- section number;
- section title;
- section first page number.

having already run procedures to create the overall order number and the initial page number of each paper.

Write a SQL procedure to generate the complete L^AT_EX 2_ε code to produce the overall section summary listing for the three volumes.

Write SQL procedures to generate the detailed table of contents for each of the three volumes listing the:

- chapter title;
- section title;
- paper title;
- list of authors (in the same order as on the paper);
- (no affiliations);
- the initial page number.

List of authors field constructed by concatenating all authors for each paper (stored in a database table with an author number and affiliation). Unlimited length field which was also input to the hidden field in the PDF file, even if >256 characters.

Initial page numbers calculated automatically using a series of SQL sequence routines.

Producing the TOC and the AI

TOC — EPAC/Others

FileMaker Pro/HTML (WWW version)

- One record per paper.
- Special characters already converted to HTML format in database fields.
- Author file output from database has paper code and author name (last name, initials) sorted by paper code.
- Input to Perl script which:
 - splits line into 3 components;
 - concatenates authors for each paper;
 - creates array of author lists indexed by paper code.
- First page number file output from database and input to Perl script which:
 - Splits and creates an array of first page numbers indexed by paper code.
- Output file from database containing programme code, abstract code and title.
- Input to Perl script which:
 - Splits fields.
- Run Perl script which outputs HTML code to a file which contains:
 - Title, link to abstract + “Contribution not received”
 - OR
 - Title, link to paper, list of authors, link to abstract, first page #.

FileMaker Pro/MS Word (CD-ROM, book versions)

PDF version for CD-ROM and book generated from a print-merge into MS Word on the Mac → *.ps* → *.pdf* (+ PDFmarks).

AUTHOR INDEX

PAC'97:

Oracle/L^AT_EX 2_ε (book version)

- Write SQL code to generate L^AT_EX 2_ε code with authors ordered alphabetically with concatenated first page numbers for their papers.
- Then manually edit the *.tex* file to rationalize author names.

EPAC:

FileMaker Pro/HTML (WWW version)

- Create duplicate author fields in the database and make global substitutions to convert special characters to HTML format.
- Author names already rationalized on the database.
- Create authors file from the database, ordered alphabetically, each line containing: last name, initials, filename.
- Run Perl script to create HTML file with navigation links to the letters of the alphabet at each letter boundary, and listing by name with links to each author's papers by filename (paper identifier).

FileMaker Pro/MS Word (CD-ROM, book versions)

- Follow a similar procedure to construct a file with one line per author and reimport into the Mac to produce an MS Word *.doc* file → *.ps* → *.pdf* retaining the special characters in the authors' names.