pdfTeX and XML in the Workflow for Conference Proceedings

Volker RW Schaa

Gesellschaft für Schwerionenforschung mbH
Darmstadt, Germany

Practical \TeX\ 2004
Holiday Inn Fisherman’s Wharf
San Francisco, California
July 20, 2004
Two years ago appointment as Proceedings Editor for two conferences,

Responsibility for preparation of Abstracts- and Paper-Submissions, and participant data (so-called Affiliations),

web presentation and conference volume (Proceedings)

size of conferences small, but too large for manual production

- **DIPAC2003** 85 papers, 300 pages, 290 authors
- **LINAC2004** 450 papers, 1400 pages, ~1200 authors

existing conference systems in particle physics were examined (DESY/CERN)

- **SRF2003** 185 papers, 620 pages, 520 authors
- **EPAC/PAC** 1200 papers, 4000 pages, 7500 authors

Findings:
Findings

- just the Web presentation is solved in a satisfactory way,
- no platform independent solutions (»Windows only«),
- use of VB scripts in Paper Submissions for
  - »Hidden Field« entries,
  - page numbering, and
  - header/footer information.
- adaption to speed of used equipment (»wait xxx«)
- creation of Proceedings only manually with word processors (Word, Quark, etc.)
- latest decision: no proceedings as printed copies anymore
- no batch oriented processing,
- sole Open Source component is PERL
- but...
leveling out to 7-Bit ASCII
no »special characters« from particle physics
  e^+/e^-
  α, β, λ,…
  big trouble with names like »DAΦNE«

name handles are non honored (van, von, D’, etc.)
sorting of names based by ASCII rules
where to look for your name?
  Bär, before or after Brønsted, and Budyšin?
  Sørby after Szymborska!
  and Čemešič?

mediocre Proceedings (print and structure)
no navigation in Web Proceedings
starting point

1. the idea
   - typesetting: \textsc{pdfTEx}
   - scripting: \textsc{Perl}
   - data: XML

2. the method
   - database export in XML
   - interpretation of XML by \textsc{Perl} scripts
   - transformation to `<html>` and `\pdfTEx`

3. the prerequisites
   - XML structure definitions (as DTD, Schema, RelaxNG, ?)
   - xml database export (native Oracle 9i, 10g)
   - Unicode (input problem)

Now a short excursion to XML and \textsc{Structure}...
What kind of structure(s) do we need for a conference?

```
<conference name="...">
  <session name="...">
    <paper code="...">
      <institute ...>
        <author><name>...</>/<email>...</</author>
        (more »authors«)
      </institute>
      (more »institutes«)
    </paper>
    <paper>...</> (more »papers«)
  </session>
  (more »sessions«)
</conference>
(more conferences??)
```
A *conference* consists of several *sessions*. And for a session you need talks (↦ *paper*) or posters (↦ *paper*).

```xml
<conference name="conference name">
  <session name="session name" abbr="abbreviation of session">
    <paper code="paper code id" abstract="yes|no">
      (xml »paper« sub-structure)
    </paper>
  </session>
  <session name="..." abbr="...">
    (xml »session« sub-structure)
  </session>
  (more »session« structures)
</conference>
```
A paper consists of several sub-structures: You need a title for your paper, and some keywords.

```
<paper code="paper id" abstract="yes|no">
  <title[ note="optional title note"]>
    paper title
  </title>
  <keywords>
    (xml »keyword« structure)
  </keywords>
  <institute>
    (xml »institute« structure)
  </institute>
</paper>
```

And there is always an institute you are working for, so you have to be defined as a part of it.
Now you have to define the *institute*, you are part of. Finally your name appears in an *author* structure.

```xml
<institute name="name of institute"
    abbr="institute (abbreviation)">
    <author [ main="optional: "yes" for main author"]
        [ note="optional author note"]>
        <name>(xml »name« structure)</name>
        <email>email address</email>
    </author>
    (more »author« structures)
</institute>
```

Even your *name* has some structure, and that comes now...
As the *author* preparing the paper, you set the key *main* to "yes".

```xml
<author main="yes"[ note="work funded by ..."]>
  <name>
    <initials>author’s initials</initials>
    <lastname>author’s lastname</lastname>
  </name>
  <email>author’s email address</email>
</author>
```

And your *name* consists of *lastname* and *initials* (and maybe an *email address*).
Finally you have to define the *keywords*. Each entry consists of a single line with a given *keyword*.

\[
\begin{align*}
\text{<keywords>}
\quad &\text{<keyword>keyword</keyword>} \\
\quad &\text{<keyword>another keyword</keyword>} \\
\quad &\text{more xml »keyword« structures} \\
\quad &\text{...} \\
\quad &\text{(a maximum of »5« keyword structures)} \\
\text{</keywords>}
\end{align*}
\]

Now you are done for Proceedings volume and web presentation. . .

or even better, everything is done for you by a script from database entries!
What is the script doing?

1. It reads pdf-files and counts pages in each file,
2. reads XML, and generates `<html>` for
   - Session List,
   - Authors’ List,
   - Keyword List,
   - Institute List
3. generates `\pdfTeX` wrappers
   - for each single (raw) pdf-file,
   - for proceedings file,
4. writes command files for
   - generating pdf-files with Authors and Keyword information,
   - building of proceedings file(s).
Actually built-in features:

1. Web pages and proceedings honor special characters,
2. Web pages are in Unicode (UTF8),
3. All names with accented characters and umlauts,
4. Math formulas (in abstracts) on web pages,
5. Sorting author names is rule based (accented letters, umlauts, etc. . . .)
Uršič, R.

**PT05** Experience With Sampling Of 500 MHz Rf Signal For Digital Receiver Applications

*U. Mavrič, S. Bremec, R. Uršič*

I-Tech, Instrumentation Technologies, Solkan, Slovenia

This article will present test results of a prototype system that was built to evaluate feasibility of a direct sampling of a 500 MHz RF signal for use in digital receiver applications. The system consists of a variable gain RF front end, a fast analog to digital converter (ADC) and a field programmable gate array (FPGA) providing glue-logic between the ADC and a PC computer.
Use of Optical Transition Radiation Interferometry for Energy Spread And Divergence Measurements

**R.B. Fiorito, A.G. Shkvarunets**  
IREAP, Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, MD, USA

OTR interferometry (OTRI) has been shown to be an excellent diagnostic for measuring the rms divergence and emittance of relativistic electron beams when the energy spread $\Delta \gamma / \gamma$ is less than the normalized rms divergence $\sigma = \gamma \Theta_{\text{rms}}$. This is the case for most beams previously diagnosed with OTRI. To extend this diagnostic capability to beams with larger energy spreads, we have calculated the effects of all the parameters effecting the visibility of OTR interferences, $V$; i.e. energy spread, angular divergence, the ratio of foil separation to wavelength ratio, $d/\lambda$ and filter bandpass. We have shown that:

1. for a given $\Delta \gamma / \gamma$, the sensitivity of $V$ to $\sigma$ is proportional to the observation angle $\Theta_0$, the fringe order $n$ and the ratio $d/\lambda$;
2. the sensitivity of $V$ to $\Delta \gamma / \gamma$ is independent of $\Theta_0$ and $n$ but is proportional to $d/\lambda$.

Thus, by adjusting $d/\lambda$, and choosing the appropriate fringe order, one can separate out and measure both the energy spread and divergence. However, the filter bandpass must decrease with $\Theta_0$ and $n$. Results of our calculations will be given for various beams of interest.
<html>feature="Sorting Order" (i.e. ö ↔ oe)</html>
Actually built-in features:

1. printing of header and footer information,
2. transfer of all meta-information into pdf-file,
3. (down)scale depending on size of crop/media/object-box,
4. setting of page numbers after count of all pages,
5. author and institute’s index with links to articles,
6. inclusion of paper or "missing" note,
7. config file with settings for directories, sort-rules, and any dependencies etc.
\documentclass[twoside]{book}
\usepackage[papersize={595pt,792pt}, body={ 483pt, 680pt},
   top=54pt, left=56pt, head=18pt, headsep=15pt, footskip=32pt]{geometry}
\usepackage{fancyhdr}\pagestyle{fancy}
\begin{document}
\pdfinfo{%
/Title (Characterisation of Fast Faraday Cups at the ELETTRA Linac)
/Author (M. Ferianis, S. Bassanese, G. D'Auria ELETTRA, Sincrotrone Trieste, Trieste, Italy;
   C. Deibele SNS, Spallation Neutron Source, Oak Ridge, TN, USA;
   M. Poggi INFN-LNL, Legnaro, Italy)
/Subject (Proceedings DIPAC 2003 -- Mainz, Germany)
/Keywords (diagnostics, electron, ELETTRA, instrumentation, linac)
}%
\setcounter{page}{113}
\fancyhead[CE,CO]{\large\sffamily Proceedings DIPAC 2003 -- Mainz, Germany}%
\fancyfoot[RE,LO]{\large\sffamily Posters Monday}%
\fancyfoot[RO,LE]{\large\sffamily\thepage}%
\fancyfoot[CE,CO]{\large\sffamily PM10}%
\IfFileExists{../papers-final/PM10.pdf}{% 
   \includepdf[pages=-, scale=1.0,
      pagecommand={}]{../papers-final/PM10.pdf}%
   {\Huge\textbf{PAPER NOT YET RECEIVED}}}{}
\end{document}
\documentclass[twoside]{book}
\usepackage[papersize={595pt,792pt}, body={483pt, 680pt},
  top=54pt, left=56pt, head=18pt, headsep=15pt, footskip=32pt]{geometry}
\usepackage{fancyhdr}\pagestyle{fancy}
\begin{document}
\pdfinfo{%/Title (Characterisation of Fast Faraday Cups at the ELETTRA Linac)
  /Author (M. Ferianis, S. Bassanese, G. D'Auria ELETTRA, Sincrotrone Trieste, Trieste, Italy;
  C. Deibele SNS, Spalation Neutron Source, Oak Ridge, TN, USA;
  M. Poggi INFN-LNL, Legnaro, Italy)
  /Subject (Proceedings DIPAC 2003 -- Mainz, Germany)
  /Keywords (diagnostics, electron, ELETTRA, instrumentation, linac)}
\setcounter{page}{113}
\fancyhead[CE,CO]{\large\sffamily Proceedings DIPAC 2003 -- Mainz, Germany}%
\fancyfoot[RE,LO]{\large\sffamily Posters Monday}%
\fancyfoot[RO,LE]{\large\sffamily\thepage}%
\fancyfoot[CE,CO]{\large\sffamily PM10}
\IfFileExists{../papers-final/PM10.pdf}{%
  \includepdf[pages=-, scale=1.0,
    pagecommand={}}{../papers-final/PM10.pdf}%
  {\Huge\mbox{\vfill
    \centering\textsf{\textbf{PAPER NOT YET RECEIVED}}}
  \vfill}
\end{document}
```latex
\documentclass[twoside]{book}
\usepackage[papersize={595pt,792pt}, body={ 483pt, 680pt},
       top=54pt, left=56pt, head=18pt, headsep=15pt, footskip=32pt]{geometry}
\usepackage{fancyhdr}\pagestyle{fancy}
\begin{document}
\pdfinfo{\
/Title (Characterisation of Fast Faraday Cups at the ELETTRA Linac)
/Author (M. Ferianis, S. Bassanese, G. D'Auria ELETTRA, Sincrotrone Trieste, Trieste, Italy;
   C. Deibele SNS, Spalation Neutron Source, Oak Ridge, TN, USA;
   M. Poggi INFN-LNL, Legnaro, Italy)
/Subject (Proceedings DIPAC 2003 -- Mainz, Germany)
/Keywords (diagnostics, electron, ELETTRA, instrumentation, linac)
}
\setcounter{page}{113}
\fancyhead[CE,CO]{\large\sffamily Proceedings DIPAC 2003 -- Mainz, Germany}\
\fancyfoot[RE,LO]{\large\sffamily Posters Monday}\
\fancyfoot[RO,LE]{\large\sffamily \thepage}\
\fancyfoot[CE,CO]{\large\sffamily PM10}
\IfFileExists{../papers-final/PM10.pdf}{\%
   \includepdf[pages=-, scale=1.0,
       pagecommand={}){../papers-final/PM10.pdf}]%\
   \Huge\mbox{\vfill\centering\textsf{\textbf{PAPER NOT YET RECEIVED}}}\
   \vfill}{\%
   \end{document}
```
"pdfinfo" transfers all meta info into the pdf file

\documentclass[twoside]{book}
\usepackage[papersize={595pt,792pt}, body={483pt, 680pt},
  top=54pt, left=56pt, head=18pt, headsep=15pt, footskip=32pt]{geometry}
\usepackage{fancyhdr}\pagestyle{fancy}

\begin{document}
\pdfinfo{%
  /Title (Characterisation of Fast Faraday Cups at the ELETTRA Linac)
  /Author (M. Ferianis, S. Bassanese, G. D'Auria ELETTRA, Sincrotrone Trieste, Trieste, Italy;
          C. Deibele SNS, Spalation Neutron Source, Oak Ridge, TN, USA;
          M. Poggi INFN-LNL, Legnaro, Italy)
  /Subject (Proceedings DIPAC 2003 -- Mainz, Germany)
  /Keywords (diagnostics, electron, ELETTRA, instrumentation, linac)
}
\setcounter{page}{113}
\fancyhead[CE,CO]{\large\sffamily Proceedings DIPAC 2003 -- Mainz, Germany}%
\fancyfoot[RE,LO]{\large\sffamily Posters Monday}%
\fancyfoot[RO,LE]{\large\sffamily\thepage}%
\fancyfoot[CE,CO]{\large\sffamily PM10}
\IfFileExists{../papers-final/PM10.pdf}{%
  \includepdf[pages=-, scale=1.0, pagecommand={}]{../papers-final/PM10.pdf}}%
  {\Huge\mbox{\vfill}
   \centering\textsf{\textbf{PAPER NOT YET RECEIVED}}
  \vfill}
\end{document}

Volker RW Schaa

pdfTeX and XML in the Workflow for Conference Proceedings
\documentclass[twoside]{book}
\usepackage[papersize={595pt,792pt}, body={483pt, 680pt},
  top=54pt, left=56pt, head=18pt, headsep=15pt, footskip=32pt]{geometry}
\usepackage[fancyhdr]{fancyhdr}
\pagestyle{fancy}
\begin{document}
\pdfinfo{
  /Title (Characterisation of Fast Faraday Cups at the ELETTRA Linac)
  /Author (M. Ferianis, S. Bassanese, G. D'Auria ELETTRA, Sincrotrone Trieste, Trieste, Italy;
     C. Deibele SNS, Spallation Neutron Source, Oak Ridge, TN, USA;
     M. Poggi INFN-LNL, Legnaro, Italy)
  /Subject (Proceedings DIPAC 2003 -- Mainz, Germany)
  /Keywords (diagnostics, electron, ELETTRA, instrumentation, linac)
}
\setcounter{page}{113}
\fancyhead[CE,CO]{\large\sffamily Proceedings DIPAC 2003 -- Mainz, Germany}\
\fancyfoot[RE,LO]{\large\sffamily Posters Monday}\
\fancyfoot[RO,LE]{\large\sffamily\thepage}\
\fancyfoot[CE,CO]{\large\sffamily PM10}
\IfFileExists{../papers-final/PM10.pdf}{\includepdf[pages=-, scale=1.0,\
  pagecommand=]{../papers-final/PM10.pdf}}{\Huge\mbox{}\vfill\
  \centering\textsf{\textbf{PAPER NOT YET RECEIVED}}}\
\vfill\end{document}
\documentclass[twoside]{book}
\usepackage[papersize={595pt,792pt}, body={483pt, 680pt},
  top=54pt, left=56pt, head=18pt, headsep=15pt, footskip=32pt]{geometry}
\usepackage{fancyhdr}
\pagestyle{fancy}
\begin{document}
\pdfinfo{
  /Title (Characterisation of Fast Faraday Cups at the ELETTRA Linac)
  /Author (M. Ferianis, S. Bassanese, G. D'Auria ELETTRA, Sincrotrone Trieste, Trieste, Italy;
          C. Deibele SNS, Spalation Neutron Source, Oak Ridge, TN, USA;
          M. Poggi INFN-LNL, Legnaro, Italy)
  /Subject (Proceedings DIPAC 2003 -- Mainz, Germany)
  /Keywords (diagnostics, electron, ELETTRA, instrumentation, linac)
}
\setcounter{page}{113}
\fancyhead[CE,CO]{\large\sffamily Proceedings DIPAC 2003 -- Mainz, Germany}\
\fancyfoot[RE,LO]{\large\sffamily Posters Monday}\
\fancyfoot[RO,LE]{\large\sffamily \thepage}\
\fancyfoot[CE,CO]{\large\sffamily PM10}\
\IfFileExists{../papers-final/PM10.pdf}{%
  \includepdf[pages=-, scale=1.0,\n    pagecommand={\Huge\mbox{\vfill\n        \centering\textsf{\textbf{\small{\textbf{PAPER NOT YET RECEIVED}}}}}\n        \vfill}\n  }{\Huge\mbox{\vfill}}\
\end{document}

documentclass[twoside]{book}
\usepackage[papersize={595pt,792pt}, body={ 483pt, 680pt},
   top=54pt, left=56pt, head=18pt, headsep=15pt, footskip=32pt]{geometry}
\usepackage{fancyhdr}\pagestyle{fancy}
\begin{document}
\pdfinfo{
   /Title (Characterisation of Fast Faraday Cups at the ELETTRA Linac)
   /Author (M. Ferianis, S. Bassanese, G. D'Auria ELETTRA, Sincrotrone Trieste, Trieste, Italy;
      C. Deibele SNS, Spallation Neutron Source, Oak Ridge, TN, USA;
      M. Poggi INFN-LNL, Legnaro, Italy)
   /Subject (Proceedings DIPAC 2003 -- Mainz, Germany)
   /Keywords (diagnostics, electron, ELETTRA, instrumentation, linac)
}
\setcounter{page}{113}
\fancyhead[CE,CO]{\large\sffamily Proceedings DIPAC 2003 -- Mainz, Germany}\
\fancyfoot[RE,LO]{\large\sffamily thepage}\
\fancyfoot[RO,LE]{\large\sffamily \thepage}\
\fancyfoot[CE,CO]{\large\sffamily PM10}\
\IfFileExists{../papers-final/PM10.pdf}{%\
   \includepdf[pages=, scale=1.0, pagecommand={}]{../papers-final/PM10.pdf}\
   \Huge\mbox{\vfill}\
   \centering\textsf{\textbf{PAPER NOT YET RECEIVED}}}\
\vfill\end{document}
Information are set in the config file
»scaling« is determined by maximum of crop/media-box sizes

\documentclass[twoside]{book}
\usepackage[papersize={595pt,792pt}, body={ 483pt, 680pt},
    top=54pt, left=56pt, head=18pt, headsep=15pt, footskip=32pt]{geometry}
\usepackage{fancyhdr}\pagestyle{fancy}
\begin{document}
\pdfinfo{%
    /Title (Characterisation of Fast Faraday Cups at the ELETTRA Linac)
    /Author (M. Ferianis, S. Bassanese, G. D'Auria ELETTRA, Sincrotrone Trieste, Trieste, Italy;
    C. Deibele SNS, Spalation Neutron Source, Oak Ridge, TN, USA;
    M. Poggi INFN-LNL, Legnaro, Italy)
    /Subject (Proceedings DIPAC 2003 -- Mainz, Germany)
    /Keywords (diagnostics, electron, ELETTRA, instrumentation, linac)
}
\setcounter{page}{113}
\fancyhead[CE,CO]{\large\sffamily Proceedings DIPAC 2003 -- Mainz, Germany}%
\fancyfoot[RE,LO]{\large\sffamily Posters Monday}%
\fancyfoot[RO,LE]{\large\sffamily \thepage}%
\fancyfoot[CE,CO]{\large\sffamily PM10}
\IfFileExists{../papers-final/PM10.pdf}{%
    \includepdf[pages=-, scale=1.0,
    pagecommand={}]{../papers-final/PM10.pdf}}%
\Huge\mbox{\vfill\centering\textsf{\textbf{PAPER NOT YET RECEIVED}}\vfill}
\end{document}
What happens, if you compile this \TeX script without the necessary pdf-file?

You only get the "missing" note.
What happens, if you compile this \TeX script without the necessary pdf-file?

You only get the "missing" note.

The text is configurable in the config file.

And if it's there, you get...
the scripts are usable, and tested on DIPAC/LINAC, and EPAC related conference papers and settings,

- translation of special character to Unicode has to be extended
  - actually 65 accented letters,
  - 17 special characters,
  - 113 math symbols, and
  - 39 Greek letters.

- generating wrappers for pdf-files and Web pages is one script at the moment (if you want one, you get the other free)
- I’m going to maintain and extend the scripts, having them checked at a bigger conference (EPAC with 1239 papers)
- the conference database system (SPMS: Scientific Programme Management System) and all scripts will be available under GPL license in September 2004

- <advertisement>
  everything runs with the \TeX\ Collection Set
  that’s unnecessary here of course!
</advertisement>
Some New Scripts for the Wrapper

Volker RW Schaa

Gesellschaft für Schwerionenforschung mbH
Darmstadt, Germany

JACoW Team Meeting,
Trieste, Italy,
Dec 2003

Thank you!

Volker RW Schaa
Problems, what's missing, . . .

- document internal links are lost when documents are included by `pdfpages`,
- fast Web Display: not yet,
- (raw) documents must be O.K.,
  - color space (CMYK, RGB, etc.),
  - imbedded fonts,
- printing house has no access to embedded XObjects,
- changed behavior of Acrobat
  (version $\leq 5$: first, version 6: last document info),
- . . .
Thank you!
pdfTeX and XML in the Workflow for Conference Proceedings

Volker RW Schaa

Gesellschaft für Schwerionenforschung mbH
Darmstadt, Germany

Practical \TeX\ 2004
Holiday Inn Fisherman’s Wharf
San Francisco, California
July 20, 2004