Automation of creation of public material: progress report

Frank Filthaut (Radboud/Nikhef)

Recapitulation

Discussed during January CM: import & adapt ATLAS software (all Python based) to prepare web pages featuring material (figures, tables) in publications

Two steps:

1. Extraction of tables, figures and corresponding captions

2. Integration into a web page that is to be made public

- Possibility for "embargo": restriction of visibility to DUNE

- could conceive of steering this step from docdb instead of CSV (thanks to additional tags!), but would then require arXiv, DOI in addition

Now have 5 papers submitted, and spent a bit of time investigating (some) further details

- most paper sources retrieved from arXiv

Extracting material from sources

Bottom line: essentially working (little change from March)

- Minor tweaking will presumably remain necessary ⇒ it is not prudent to attempt to automate this step (cf. suggestion in August)

  - Home-brewn code to convert LaTeX macros into html code, functional but does not do a perfect job (this is one of the reasons for wanting the "embargo" functionality)

- Issues found thus far:

  - figures: assumes that subfigures are all numbered (a), (b), … not necessarily in agreement with the "left" / "right" used

  - having citations or references to other figures or tables won't work (and would not be very meaningful on a web page in any case)

  - script chokes on e.g. `graphicspath` definition being delegated to separate `preamble.tex`

Technicalities to be sorted out

Hosting the software and having a location on the Fermilab DUNE web pages

- Authorisation & implementation of embargo still to be sorted out

  • have had some interaction with Mike Kirby, but still need to follow up on the details

Customising the output

- have a better understanding now of CSS items, but I believe that we should aim for an appearance like on dunescience.org (fonts, colours)

  • as it seems to me that this is where these pages should end up

So far I've been working mostly on pages for individual papers; need also some work on a top-level page providing access to these individual pages

- don't expect this to be a major issue, but should decide what to put there

Technicalities to be sorted out (cont'd)

Metadata: still to add arXiv & publication (DOI) information + abstract

- not sure (anymore) that automatic abstract retrieval from docdb (or arXiv) is the right way to go, as these are not guaranteed to be updated after iteration with journal referees

  - could reconsider if this is made a hard requirement? Otherwise, abstract will need to be provided separately

    - on ATLAS, use the CERN CDS server for this, along with

- adding arXiv & publication links should not be difficult

  - not sure if naming the publication will be possible as well