TeXFolio — a framework to typeset XML documents using TeX

Rishikesan Nair T., Rajagopal C.V.,
Radhakrishnan C.V.

Abstract

TeXFolio is a web-based framework on the cloud to generate standards-compliant, hyperlinked, bookmarked PDF output directly from XML sources with heavy math content, using TeX. TeXFolio is a complete journal production system as well. It can produce strip-ins which are alternate GIF or SVG images of MathML content. In addition, DOI look-up, HTML rendering of XML/MathML, and the whole dataset generation according to the customer's specification are also possible. Customer-specific validation tools can be integrated in this system.

1 Introduction

TeXFolio is a web-based complete journal production system that accepts XML documents as input and generates a variety of outputs per user directives. At the moment, TeXFolio accepts documents tagged per the NLM/JATS or Elsevier Journal Article DTD. The typesetting engine is TeX, which allows hyperlinked, bookmarked and standards-compliant PDF outputs of infinite variants in terms of look and feel. It further permits TeX authors to directly edit their documents at the proof stage and a master copy editor can pass it for publishing with minimal loss of time.

Although the underlying engine of TeXFolio hasn't deviated from the genre of free/libre software, the computing paradigms have markedly shifted to those in vogue to make the system modernized and competitive in terms of usability, technologies and performance. In fact, TeXFolio allows users to undertake any stage of work anywhere in the world, owing to its absolute compatibility with cloud and mobile computing. That is further augmented by the usage of LaTeX3 methodologies and programming to perfect the production system to an efficient, automated and accurate one.

2 The workflow

2.1 XML first

The input can be either XML or LaTeX. XML must be per NLM/JATS DTDs or the Elsevier Journal Article DTD. TeXFolio ingests these two types of XML documents and generates a LaTeX file by applying corresponding an XSLT stylesheet over the document. This LaTeX file is used to edit the content and/or to generate PDF output.

The processing cycle of XML → LaTeX → edit → PDF can be repeated any number of times, as shown in the schematic diagram provided in Figure 1.

Whenever the LaTeX file is edited, the user can either generate PDF or go through another cycle of XML → LaTeX → PDF to ensure high fidelity between XML and PDF, thereby making it a truly XML-first workflow. We call it the "XROUND" process.

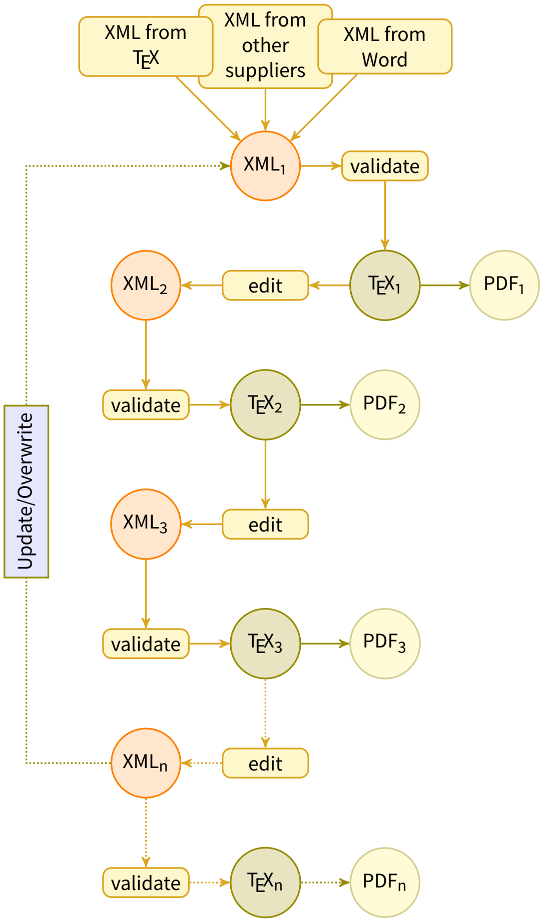


Figure 1: XML-first workflow

2.2 LaTeX input

For LaTeX input, we need to restructure the document to augment XML generation since the front matter is very verbose and structured in a granular way in XML documents. The bibliography must be provided as a BibTeX database. Barring the front and back matters, the main body of the document does not pose any problem during XML generation. The system can digest all the author's macro definitions, `newtheorem`-type declarations, or any other declarations provided by the AMS math packages and other standard LaTeX packages.

TEX4ht is our preferred engine to translate LATEX documents to XML format among all the tools and available software around. Since TEX is being used to process the document, it can easily and effectively handle all the macros and user definitions or declarations seamlessly. See Figure 2 for a schematic diagram of the workflow.

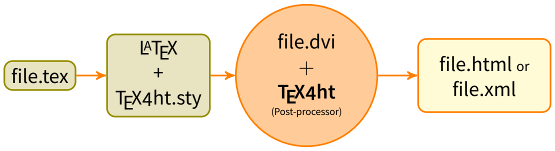


Figure 2: LATEX to XML generation diagram.

3 Features

A list of the most commonly used features built into TEXFolio are given here.

3.1 General

Cloud With the development of TEXFolio, we are able to do text processing in the cloud rather than on a personal computer. This is critically important in the current scenario. Previously, text processing software was installed on each user's desktop and updating the systems with the most recent changes always caused problems. In the text processing world, much software is subject to regular updates and ensuring these updates was a herculean task. With the deployment of TEXFolio, we need to update the software on the server only and users' desktops need not be touched.

TEX For the whole process (i.e., generate XML, PDF, strip-ins, etc.), we are using TEX and friends. TEX being the most sophisticated software for high-quality typesetting, TEXFolio ensures high-quality output especially for mathematics, computer science, economics, engineering, linguistics, physics, statistics.

Low learning curve With the help of user-friendly and an attractive interface, even a novice user can use it without much learning. Since standard LATEX commands are used, a normal LATEX user can quickly start using it.

Cross-platform Since it is browser-based, users with different operating systems can access it without any difficulty.

Browser, sole software A desktop with any current browser and an Internet connection is all that is required to access TEXFolio. Instead of a desktop machine, if you have a Raspberry Pi, that is more than enough.

Self-publishing Supporting self-publishing is another feature of TEXFolio. TEXFolio serves well the "author as publisher" since it accepts TEX documents as input and can easily generate PDF, HTML 5, NLM/JATS or Elsevier Journal Article XML outputs if the sources are marked up per the `elsarticle` or `stm` document classes. These are the deliverables required for a web platform. As mentioned above, the bibliography needs to be provided as a BIBTEX database.

3.2 Inputs

TEXFolio can currently accept the following input formats. The first is for a LATEX-first workflow whereas the second and third are for an XML-first workflow.

LATEX In the case of a LATEX-first workflow, the input source has to be structured according to `neptune.cls`, `elsarticle.cls` or `stm.cls`.

NLM/JATS XML The user loads the XML file in one of these DTDs. TEXFolio generates a TEX file from this loaded JATSXML. From now on, the source will be this machine-generated TEX file. Using this, the user can paginate, make changes and/or generate different deliverables from this source.

Elsevier Journal Article XML The user loads an Elsevier DTD XML file. TEXFolio generates a TEX file from the input. Just as with NLM/JATS, from now on, the source will be the machine-generated TEX file, and the user can generate different deliverables from this source.

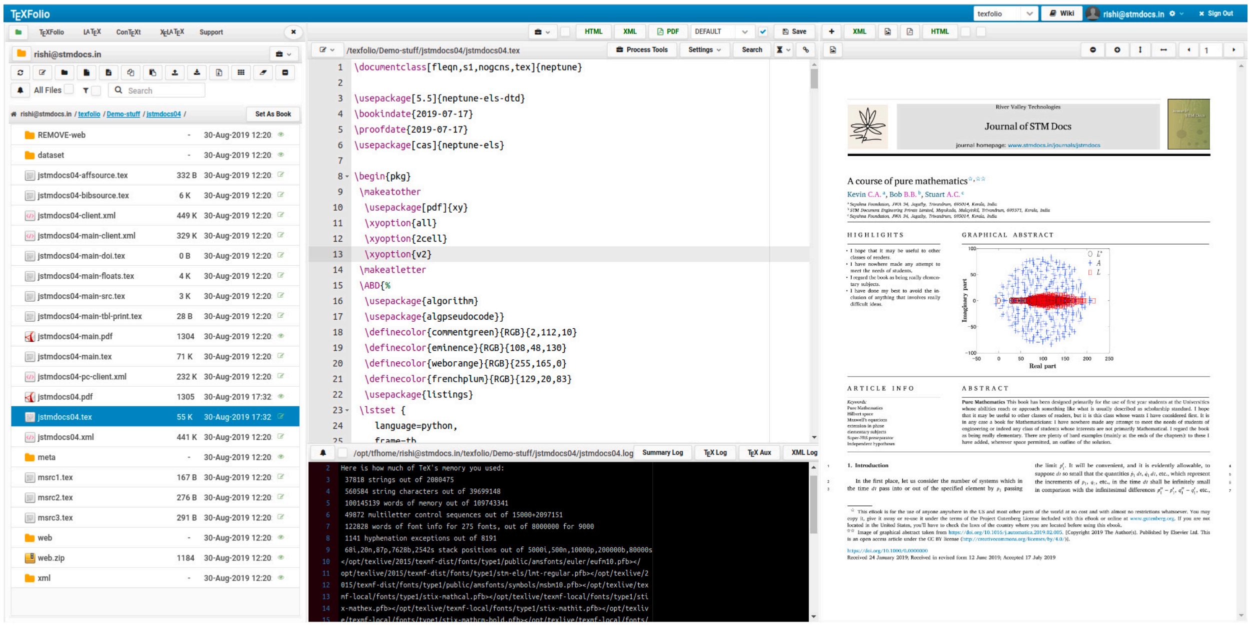
3.3 Outputs

Whatever the input (i.e., LATEX or XML), the user can generate the following output files from the source in a few seconds.

PDF The PDF output generated will be according to the standards. You may generate a web version PDF as well as a print version PDF. The Web version will be hyperlinked, bookmarked and according to the PDF/A-1 standards whereas the print ready or fat PDF will be according to PDF/X-1a standards. It can easily be configured by a developer if the user wants a PDF according to another ISO standard.

HTML5 + MathML This is another deliverable or output which can be generated from the LATEX or XML workflow. MathJax is supported.

XML The user can always generate a client version XML file in either the LATEX-first or XML-first workflow. The XML used for generating PDF can have processing instructions embedded if any vital instructions for TEX was lost in the


TEXFolio
TEXFolio LATEX ConTEXt XELATEX Support
rishi@stmdocs.in
All Files
Search
rishi@stmdocs.in / texfolio / Demo-stuff / jstmdocs04 /
Set As Book
HTML XML PDF DEFAULT Save
/texfolio/Demo-stuff/jstmdocs04/jstmdocs04.tex
Process Tools Settings Search
texfolio Wiki rishi@stmdocs.in Sign Out
XML HTML
1
/opt/tfhome/rishi@stmdocs.in/texfolio/Demo-stuff/jstmdocs04/jstmdocs04.log
Summary Log TEX Log TEX Aux XML Log

REMOVE-web	-	30-Aug-2019 12:20
dataset	-	30-Aug-2019 12:20
jstmdocs04-affsource.tex	332 B	30-Aug-2019 12:20
jstmdocs04-bibsource.tex	6 K	30-Aug-2019 12:20
jstmdocs04-client.xml	449 K	30-Aug-2019 12:20
jstmdocs04-main-client.xml	329 K	30-Aug-2019 12:20
jstmdocs04-main-doi.tex	0 B	30-Aug-2019 12:20
jstmdocs04-main-floats.tex	4 K	30-Aug-2019 12:20
jstmdocs04-main-src.tex	3 K	30-Aug-2019 12:20
jstmdocs04-main-tbl-print.tex	28 B	30-Aug-2019 12:20
jstmdocs04-main.pdf	1304	30-Aug-2019 12:20
jstmdocs04-main.tex	71 K	30-Aug-2019 12:20
jstmdocs04-pc-client.xml	232 K	30-Aug-2019 12:20
jstmdocs04.pdf	1305	30-Aug-2019 17:32
jstmdocs04.tex	55 K	30-Aug-2019 17:32
jstmdocs04.xml	441 K	30-Aug-2019 12:20
meta	-	30-Aug-2019 12:20
msrc1.tex	167 B	30-Aug-2019 12:20
msrc2.tex	276 B	30-Aug-2019 12:20
msrc3.tex	291 B	30-Aug-2019 12:20
web	-	30-Aug-2019 12:20
web.zip	1184	30-Aug-2019 12:20
xml	-	30-Aug-2019 12:20

```
\documentclass[fleqn,s1,nogcns,tex]{neptune}

\usepackage[5.5]{neptune-els-dtd}
\bookindate{2019-07-17}
\proofdate{2019-07-17}
\usepackage[cas]{neptune-els}

\begin{pkg}
 \makeatother
  \usepackage[pdf]{xy}
  \xyoption{all}
  \xyoption{2cell}
  \xyoption{v2}
 \makeatletter
 \ABD{%
  \usepackage{algorithm}
  \usepackage{algpseudocode}}
  \definecolor{commentgreen}{RGB}{2,112,10}
  \definecolor{eminence}{RGB}{108,48,130}
  \definecolor{weborange}{RGB}{255,165,0}
  \definecolor{frenchplum}{RGB}{129,20,83}
  \usepackage{listings}
 \lstset {
    language=python,
    frame=th
```

```
Here is how much of TeX's memory you used:
 37818 strings out of 2080475
 560584 string characters out of 39699148
 100145139 words of memory out of 109743341
 49872 multiletter control sequences out of 15000+2097151
 122828 words of font info for 275 fonts, out of 8000000 for 9000
 1141 hyphenation exceptions out of 8191
 68i,20n,87p,7628b,2542s stack positions out of 5000i,500n,10000p,200000b,80000s
</opt/texlive/2015/texmf-dist/fonts/type1/public/amsfonts/euler/eufm10.pfb></
opt/texlive/2015/texmf-dist/fonts/type1/stm-els/lmt-regular.pfb></opt/texlive/2
015/texmf-dist/fonts/type1/public/amsfonts/symbols/msbm10.pfb></opt/texlive/tex
mf-local/fonts/type1/stix-mathcal.pfb></opt/texlive/texmf-local/fonts/type1/sti
x-mathex.pfb></opt/texlive/texmf-local/fonts/type1/stix-mathit.pfb></opt/texliv
e/texmf-local/fonts/type1/stix-mathrm-bold.pfb></opt/texlive/texmf-local/fonts/
```

River Valley Technologies

Journal of STM Docs

journal homepage: www.stmdocs.in/journals/jstmdocs

A course of pure mathematics[☆,☆☆]

Kevin C.A.[a], Bob B.B.[b], Stuart A.C.[c]

[a] Sayahna Foundation, JWA 34, Jagathy, Trivandrum, 695014, Kerala, India
[b] STM Document Engineering Private Limited, Mepukada, Malayinkil, Trivandrum, 695571, Kerala, India
[c] Sayahna Foundation, JWA 34, Jagathy, Trivandrum, 695014, Kerala, India

HIGHLIGHTS

- I hope that it may be useful to other classes of readers.
- I have nowhere made any attempt to meet the needs of students.
- I regard the book as being really elementary subjects.
- I have done my best to avoid the inclusion of anything that involves really difficult ideas.

GRAPHICAL ABSTRACT

ARTICLE INFO

Keywords:
Pure Mathematics
Hilbert space
Maxwell's equations
extension in phase
elementary subjects
Super-FRS preseparator
Independent hypotheses

ABSTRACT

Pure Mathematics This book has been designed primarily for the use of first year students at the Universities whose abilities reach or approach something like what is usually described as scholarship standard. I hope that it may be useful to other classes of readers, but it is this class whose wants I have considered first. It is in any case a book for Mathematicians: I have nowhere made any attempt to meet the needs of students of engineering or indeed any class of students whose interests are not primarily Mathematical. I regard the book as being really elementary. There are plenty of hard examples (mainly at the ends of the chapters): to these I have added, wherever space permitted, an outline of the solution.

1. Introduction

In the first place, let us consider the number of systems which in the time dt pass into or out of the specified element by p_1 passing the limit p_1'. It will be convenient, and it is evidently allowable, to suppose dt so small that the quantities $\dot{p}_1\,dt$, $\dot{q}_1\,dt$, etc., which represent the increments of p_1, q_1, etc., in the time dt shall be infinitely small in comparison with the infinitesimal differences $p_1'' - p_1'$, $q_1'' - q_1'$, etc.,

[☆] This eBook is for the use of anyone anywhere in the US and most other parts of the world at no cost and with almost no restrictions whatsoever. You may copy it, give it away or re-use it under the terms of the Project Gutenberg License included with this eBook or online at www.gutenberg.org. If you are not located in the United States, you'll have to check the laws of the country where you are located before using this ebook.

[☆☆] Image of graphical abstract taken from https://doi.org/10.1016/j.automatica.2019.02.005. [Copyright 2019 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)].

https://doi.org/10.1000/0.0000000

Received 24 January 2019; Received in revised form 12 June 2019; Accepted 17 July 2019

Figure 3: Main page of TEXFolio: File manager (left), text editor (center), output viewer (right), log windows (bottom), tools (top).

translation to XML. Before delivering to the client, these processing instructions can also be removed, and any other necessary changes made. Some publishers even require the XML document to be on a single physical line. Currently, it can ship either NLM/JATS XML + MathML or Elsevier Journal Article XML + MathML.

e-pub It is not one of the standard outputs at present, but is easily configurable.

MathML Math in the above XML documents will be tagged as MathML, SVG, GIF/PNG/JPG, TEX math, or all. Both MathML 2.0 and 3.0 are supported.

3.4 More features

1. Depends on LATEX3 methodologies and paradigms.
2. DOI link fetching and checking
3. Crossmark, ORCID, FundRef linking
4. Linking external objects such as Genbank accession numbers, PDB, CTGOV, OMIM, etc.
5. Source editing with track-changed source, PDF and XML at proof/final stages
6. Author proofing with Neptune
7. Tooltips in PDF
8. Technical support for TEX authors
9. NLM validator to check against NLM Article Publishing, CrossRef Deposit Schema and PMC style checker
10. Supports pdfLATEX, ConTEXt and XELATEX for PDF creation. However, for standard-compliant PDF generation, pdfLATEX is used for the time being.

4 Summary

Figure 3 shows the main page of TEXFolio.

For more details and screenshots, please visit `https://texfolio.org`.

◇ Rishikesan Nair T., Rajagopal C.V., Radhakrishnan C.V.
STM Document Engineering
Pvt. Ltd., River Valley
Campus, Mepukada, Malayinkil,
Trivandrum 695571, India
rishi (at) stmdocs.in
http://stmdocs.com