

Producción gráfica de documentos académicos con Pandoc

Abstract

Pandoc como entorno textutal de producción de documentos académicos. El texto plano beneficia a todos los usuarios, deben poder encontrar lo que necesitan, comprender lo que encuentran y usarlo para realizar tareas.

MISI 2020 - Herramientas para el desarrollo de Tesis 2021 - Lisandro Ferández

Contendios

1	Producción gráfica de documentos académicos con Pandoc	2
2	Justificación	2
2.1	Pandoc	2
2.2	Texto Plano	3
2.3	Integración	3
2.4	MathJax	5
2.5	Gráficos y diagramas	6
3	Resultados	6
4	Conclusion	7
4.1	Ver Apendice	7
4.2	Futuras lineas de trabajo	7
4.3	Debilidades	7
5	Referencias	7

1 Producción gráfica de documentos académicos con Pandoc

Este proyecto describe el uso de Pandoc como pieza central de un entorno puramente textual de producción gráfica de documentos académicos ó de complejidad considerable [1].

Con el sistema de composición tipográfica de alta calidad LaTeX entre bastidores; incluye funciones diseñadas para la producción de documentación técnica y científica.

2 Justificación

Pandoc es a la los documentos gráficos lo que <https://ffmpeg.org/about.html> es para el video. El texto plano beneficia a todos los usuarios, deben poder encontrar lo que necesitan, comprender lo que encuentran y usarlo para realizar tareas [2].

2.1 Pandoc

Pandoc es una biblioteca Haskell para convertir de un formato de markUp a otro, y una herramienta de línea de comandos que utiliza esta biblioteca.

Pandoc puede convertir entre numerosos formatos de marcado y de procesamiento de textos, incluyendo, pero sin limitarse a, varios sabores de Markdown, HTML, LaTeX y Word docx.

Hay muchas maneras de personalizar pandoc para que se adapte a sus necesidades, incluyendo un sistema de plantillas y un potente sistema para escribir filtros.

Pandoc es software libre, publicado bajo la GPL. Copyright 2006-2022 John MacFarlane.

2.2 Texto Plano

De los múltiples beneficios que trabajar con texto plano habilita se resalta la accesibilidad y la posibilidad de producir documentos gráficos de la misma manera que se produce software incluirlo en un pipeline de desarrollo de software y viceversa es un *learn by example* para escritores/editores pueden adquirir capacidad flujo de trabajo

se señala como futuras líneas de trabajo es fácilmente incluíble en operaciones remotas automáticas.

2.2.1 Markdown

Lo que distingue a Markdown de muchas otras sintaxis de marcado ligero, que suelen ser más fáciles de escribir, es su legibilidad.

Como escribe Gruber: El objetivo primordial del diseño de la sintaxis de formato de Markdown es hacerla lo más legible posible. La idea es que un documento con formato Markdown sea publicable tal cual, como texto plano, sin que parezca que ha sido marcado con etiquetas o instrucciones de formato. (<http://daringfireball.net/projects/markdown/>)

<https://github.com/gfm/>

La versión mejorada de Markdown de Pandoc incluye sintaxis para tablas, listas de definiciones, bloques de metadatos, notas a pie de página, citas, matemáticas y mucho más. Véase más abajo en Markdown de Pandoc.

Pandoc comprende una serie de extensiones útiles de la sintaxis de markdown, como los metadatos del documento (título, autor, fecha); las notas al pie; las tablas; las listas de definiciones; los superíndices y subíndices; la tachadura; las listas ordenadas mejoradas (el número de inicio y el estilo de numeración son significativos); las listas de ejemplos en ejecución; los bloques de código delimitados con resaltado de sintaxis; las comillas inteligentes, los guiones y las elipses; el markdown dentro de bloques HTML; y el LaTeX en línea. Si se desea una compatibilidad estricta con markdown, todas estas extensiones pueden desactivarse.

2.3 Integración

LaTeX – A document preparation system

LaTeX is a high-quality typesetting system; it includes features designed for the production of technical and scientific documentation. LaTeX is the de facto standard for the communication and publication of scientific documents. LaTeX is available as free software.

pandoc funciona como la plaza central que integra

Pandoc incluye

un potente sistema de citas y bibliografías automáticas.

Esto significa que puede escribir una cita como véase [3], también [4, pp. 3–9] y pandoc la convertirá en una cita con el formato adecuado utilizando cualquiera de los cientos de estilos CSL (incluyendo estilos de nota al pie, estilos numéricos y estilos autor-fecha), y añadirá una bibliografía con el formato adecuado al final del documento.

Las matemáticas de LaTeX (e incluso las macros) pueden utilizarse en los documentos de markdown.

Las matemáticas de LaTeX se convierten (según lo requiera el formato de salida) en unicode, objetos de ecuación nativos de Word, MathML o roff eqn.

2.3.1 BibLATEX - Bibliografías sofisticadas en LATEX

Los datos bibliográficos pueden estar en formato BibTeX, BibLaTeX, CSL JSON o CSL YAML. Las citas funcionan en todos los formatos de salida.

BibLATEX is a complete reimplementación of the bibliographic facilities provided by LATEX. Formatting of the bibliography is entirely controlled by LATEX macros, and a working knowledge of LATEX should be sufficient to design new bibliography and citation styles. BibLATEX uses its own data backend program called “biber” to read and process the bibliographic data. With biber, BibLATEX has many features rivalling or surpassing other bibliography systems.

To mention a few:

- Full Unicode support
- Highly customisable sorting using the Unicode Collation Algorithm + CLDR tailoring
- Highly customisable bibliography labels
- Complex macro-based on-the-fly data modification without changing your data sources
- A tool mode for transforming bibliographic data sources
- Multiple bibliographies and lists of bibliographic information in the same document with different sorting
- Highly customisable data source inheritance rules
- Polyglossia and babel support for automatic language switching for bibliographic entries and citations

- Automatic bibliography data recoding (UTF-8 -> latin1, LATEX macros -> UTF-8 etc)
- Remote data sources
- Highly sophisticated automatic name and name list disambiguation system
- Highly customisable data model so users can define their own bibliographic data types
- Validation of bibliographic data against a data model
- Subdivided and/or filtered bibliographies, bibliographies per chapter, section etc.

2.3.2 CLS

The Citation Style Language (CLS) is an XML-based format to describe the formatting of citations, notes and bibliographies, offering:

An open format Compact and robust styles Extensive support for style requirements Automatic style localization Infrastructure for style distribution and updating Thousands of freely available styles (Creative Commons BY-SA licensed)

For additional documentation, the CLS schema, styles, and locales, visit the CLS project home, citationstyles.org.

<https://docs.citationstyles.org/en/stable/specification.html>

2.4 MathJax

Services

Se proporcionan varios métodos diferentes para representar las matemáticas en HTML, incluyendo MathJax y la traducción a MathML.

When $a \neq 0$, there are two solutions to $(ax^2 + bx + c = 0)$ and they are

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Content Transformation: Stylized Epub logo MathJax provides tools to transform your content from traditional print sources into modern, accessible web content and ePubs.

High-quality typography: MathJax uses CSS with web fonts or SVG, instead of bitmap images or Flash, so equations scale with surrounding text at all zoom levels.

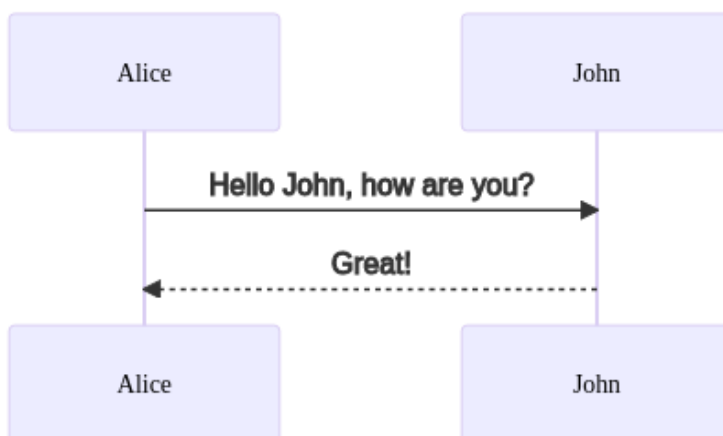
Modular Input & Output: MathJax is highly modular on input and output. Use MathML, TeX, and ASCIImath as input and produce HTML+CSS, SVG, or MathML as output.

Accessible & reusable: MathJax works with screenreaders & provides expression zoom and interactive exploration. You also can copy equations into Office, LaTeX, wikis, and other software.

2.5 Gráficos y diagramas

```
@startuml
Alice -> Bob: Authentication Request
Bob --> Alice: Authentication Response
Alice -> Bob: Another authentication Request
Alice <-- Bob: another Response
@enduml

~~~mermaid
sequenceDiagram
    Alice->>John: Hello John, how are you?
    John-->>Alice: Great!
    ~~~
```



3 Resultados

Un ejemplo de como utilizar estas herramientas es el mismo repositorio con el cual es

Ejecutando el siguiente comando se obtiene un documento grafico en formato PDF a partir del LINK de este mismo texto cuya fuente es el markdown

```
pandoc README.md --mathjax \
-F pandoc-crossref -F mermaid-filter --citeproc \
--template=plantilla --pdf-engine-opt=--shell-escape \
-s --toc --toc-depth=3 --number-sections --columns=80 \
-o README.pdf
```

4 Conclusion

4.1 Ver Apendice

contar lo que se hizo

4.2 Futuras líneas de trabajo

4.2.1 Build remoto



4.3 Debilidades

Dado que la representación intermedia de un documento por parte de pandoc es menos expresiva que muchos de los formatos entre los que convierte, no hay que esperar conversiones perfectas entre todos los formatos. Pandoc intenta conservar los elementos estructurales de un documento, pero no los detalles de formato, como el tamaño de los márgenes. Y algunos elementos del documento, como las tablas complejas, pueden no encajar en el modelo de documento simple de pandoc. Mientras que las conversiones del Markdown de pandoc a todos los formatos aspiran a ser perfectas, las conversiones de formatos más expresivos que el Markdown de pandoc pueden tener pérdidas.

HTML vs PDF

5 Referencias

- [1] W. Caleb McDaniel, “Why (and how) i wrote my academic book in plain text – w. Caleb McDaniel,” *W. Caleb McDaniel*. Disponible: <http://wcaleb.org/blog/my-academic-book-in-plain-text>
- [2] D. A. S. U. Harvard, “Use plain language,” *Digital Accessibility*. Digital Accessibility Services. Disponible: accessibility.huit.harvard.edu/use-plain-language
- [3] B. Moolenaar, “Seven habits of effective text editing,” 2000. mooleenaar.net/habits.html
- [4] D. E. Knuth, D. Knuth, and D. Bibby, *The TeXbook*. Addison-Wesley, 1986. Disponible: www-cs-faculty.stanford.edu/~knuth/abcde.html