

Universidad Tecnológica Nacional

Facultad Regional Buenos Aires

Escuela de Posgrado

MAESTRÍA EN INGENIERÍA EN SISTEMAS DE
INFORMACIÓN

Director de Carrera: Dra. Ma. Florencia Pollo Cattaneo

Seminario

HERRAMIENTAS PARA EL DESARROLLO DE TESIS 2021

Profesor: Florencia Pollo Cattaneo

Producción gráfica de documentos académicos con Pandoc

Presentado por: Lisandro Fernández

Abstract

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

Octubre 2022

Buenos Aires, Argentina

Contendios

1 Producción gráfica de documentos académicos Pandoc	1
1.1 Pandoc	1
1.2 Texto Plano	2
1.3 Integración	2
2 Resultados	7
2.1 Generar Portable Document Format (PDF)	8
2.2 Ver Apendice	8
3 Conclusion	8
3.1 Futuras líneas de trabajo	9
3.2 Debilidades	9
4 Referencias	9

1 Producción gráfica de documentos académicos Pandoc

Este proyecto describe el uso de Pandoc como pieza central de un entorno sin interfaz visual 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.

LaTeX es el estándar de facto para la comunicación y publicación de documentos científicos.

El texto plano y legible beneficia a todos los usuarios, deben poder encontrar lo que necesitan, comprender lo que encuentran y usarlo para realizar tareas [2].

1.1 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.

1.2 Texto Plano

De los beneficios que trabajar con texto plano habilita se resalta la accesibilidad y la posibilidad de producir documentos graficos 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 de flujo de trabajo

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

1.2.1 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.

1.3 Integración

pandoc funciona como la plaza central que integra

mermaidjs Mermaid le permite crear diagramas y visualizaciones utilizando texto y código.

Se trata de una herramienta de diagramación y graficación basada en JavaScript que renderiza definiciones de texto inspiradas en Markdown para crear y modificar diagramas de forma dinámica.

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.

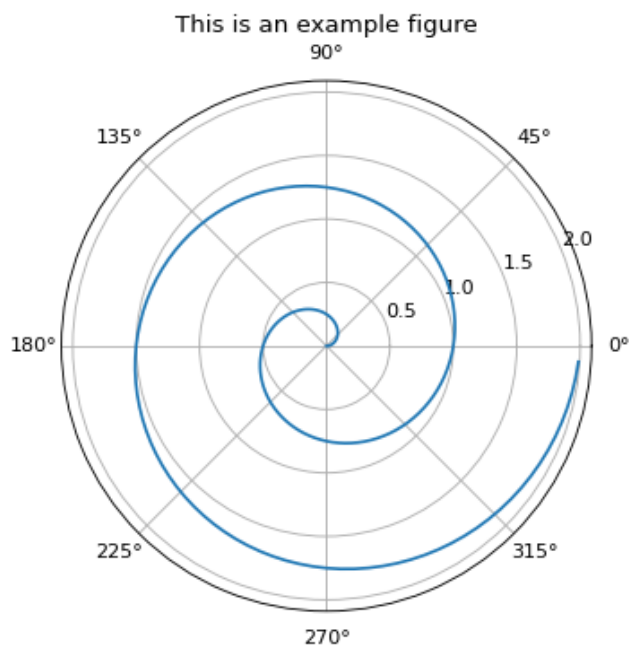
Pandoc incluye un potente sistema de citas y bibliografías automáticas.

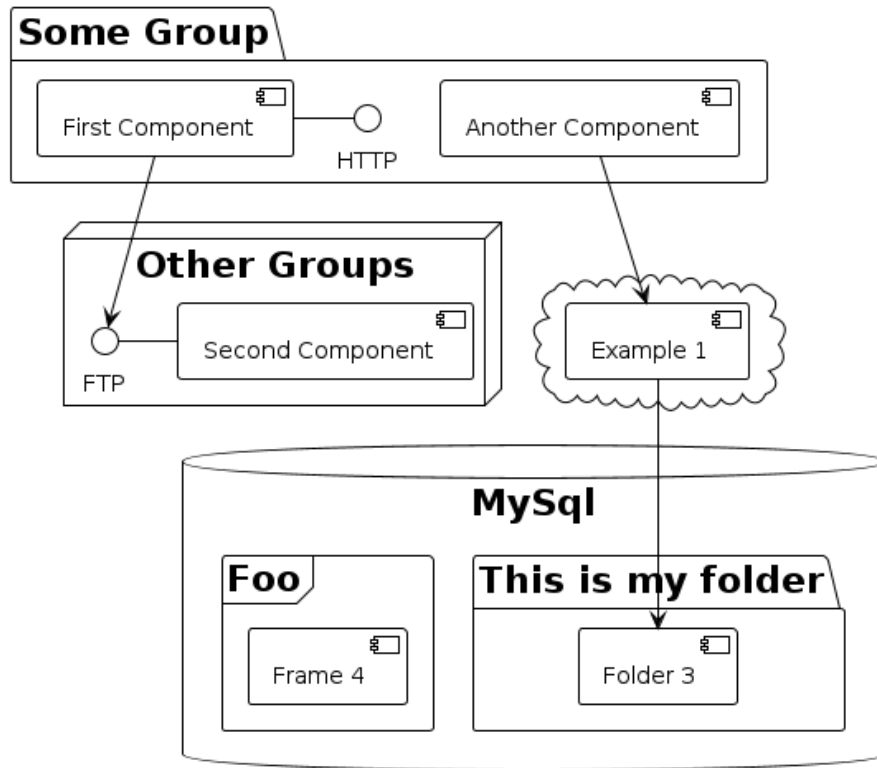
Esto significa que puede escribir una cita como véase [3], también [4, pp. 3–9] y pandoc la convertirá en una cita con el formato adecuado utilizando cualquiera de los cientos de estilos CSL (incluyendo estilos de nota al pie, estilos numéricos y estilos autor-fecha), y añadirá una bibliografía con el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

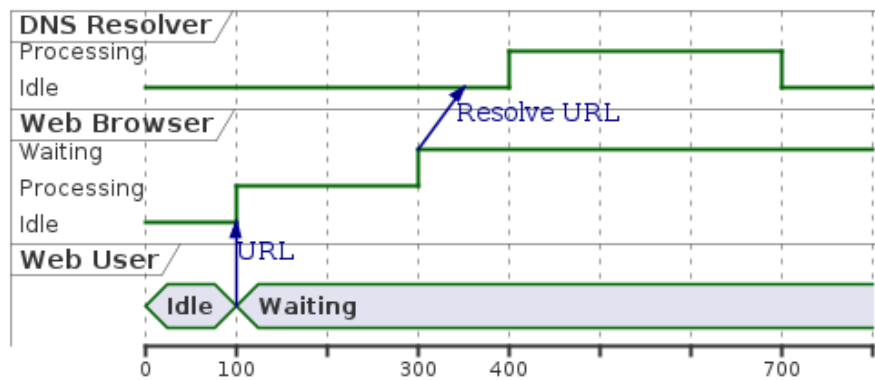
1.3.1 Gráficos y diagramas

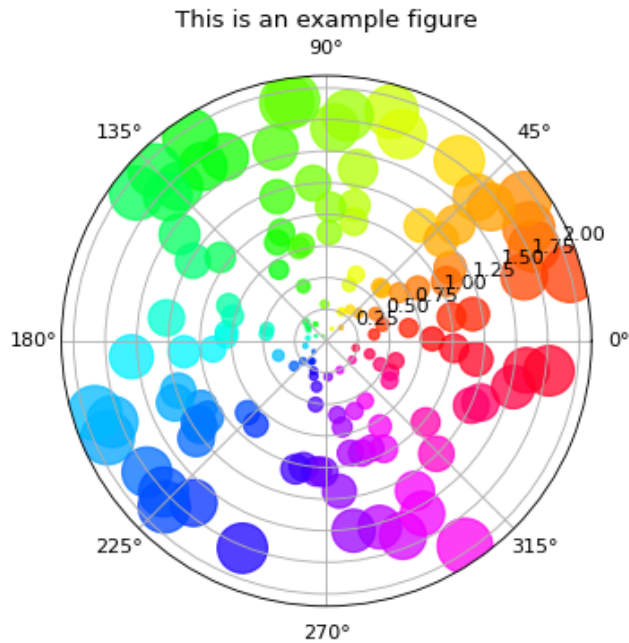
1.3.1.1 **pandoc-plot** <https://laurenttrdc.github.io/pandoc-plot/>





<https://the-lum.github.io/puml-themes-gallery/>





1.3.1.2 MermaidJS Mermaid es una herramienta de diagramación y gráficos basada en JavaScript que utiliza definiciones de texto inspiradas en Markdown y un renderizador para crear y modificar diagramas complejos. El objetivo principal de Mermaid es ayudar a que la documentación se ponga al día con el desarrollo.

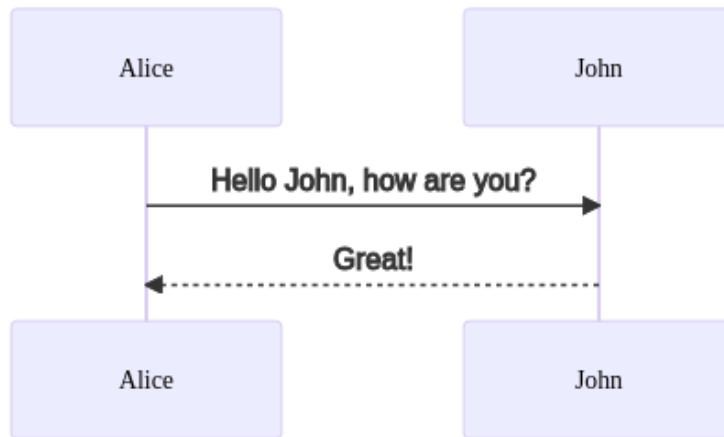
La diagramación y la documentación cuestan un tiempo precioso a los desarrolladores y quedan obsoletas rápidamente. Pero no tener diagramas o documentación arruina la productividad y perjudica el aprendizaje de la organización.

Mermaid aborda este problema permitiendo a los usuarios crear diagramas fácilmente modificables, que también pueden formar parte de los scripts de producción (y otras piezas de código).

Mermaid permite incluso a los no programadores crear fácilmente diagramas detallados y a través del Mermaid Live Editor.

Tiene tutoriales en vídeo. Utilice Mermaid con sus aplicaciones favoritas, consulte la lista de integraciones y usos de Mermaid.

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



1.3.2 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 es una reimplementación completa de las facilidades bibliográficas proporcionadas por LATEX. El formato de la bibliografía está totalmente controlado por las macros de LATEX, y un conocimiento práctico de LATEX debería ser suficiente para diseñar nuevos estilos de bibliografía y citación.

Con biber BibLATEX tiene muchas características que rivalizan o superan a otros sistemas bibliográficos.

1.3.2.1 CLS El Lenguaje de Estilo de Citación (CSL) es un formato basado en XML para describir el formato de citas, notas y bibliografías, ofreciendo:

<https://docs.citationstyles.org/en/stable/primer.html#what-is-csl>

Para obtener más documentación, el esquema CSL, los estilos y las localizaciones, visite la página web del proyecto CSL, citationstyles.org.

Si alguna vez has escrito un trabajo de investigación, habrás trabajado de investigación, probablemente has incluido referencias a otros trabajos. La página web de referencias son importantes en la comunicación académica, ya que proporcionan la atribución y enlazan la atribución y enlazan la investigación publicada. Sin embargo, formatear manualmente las referencias puede llevar mucho tiempo, especialmente cuando se trata de múltiples revistas que tienen su propio estilo de citación.

Por suerte, el software de gestión de referencias puede ayudar. Programas como Zotero, Mendeley y Papers no sólo te ayudan a gestionar tu biblioteca de investigación, sino que también pueden generar automáticamente citas y bibliografías. Pero para formatear las referencias en el estilo deseado, estos

programas necesitan descripciones de cada estilo de citación en un lenguaje que el ordenador pueda entender. Como habrás adivinado, el Lenguaje de Estilos de Citación (LEC) es ese lenguaje

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

1.3.3 MathJax

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

Cuando $a \neq 0$, hay dos soluciones a $(ax^2 + bx + c = 0)$ las cuales son

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

Transformación de contenidos: Logotipo estilizado de EpubMathJax proporciona herramientas para transformar sus contenidos de fuentes impresas tradicionales en contenidos web y ePubs modernos y accesibles.

Tipografía de alta calidad: MathJax utiliza CSS con fuentes web o SVG, en lugar de imágenes de mapa de bits o Flash, por lo que las ecuaciones se escalan con el texto circundante en todos los niveles de zoom.

Modular la entrada y la salida: MathJax es altamente modular en la entrada y la salida. Utiliza MathML, TeX, y ASCIImath como entrada y produce HTML+CSS, SVG, o MathML como salida.

Accesible y reutilizable: MathJax funciona con lectores de pantalla y proporciona zoom de expresión y exploración interactiva. También puede copiar ecuaciones en Office, LaTeX, wikis y otro software.

2 Resultados

El producto de este proyecto son una conjunto de diferentes piezas de software integradas el repositorio del código y andamiaje necesario para reproducir estem proyecto esto incluye:

- configuraciones
- estructura
- filtros
- cls
- plantilla latex
- resaltado de código
- configuraciones para operaciones remotas

2.1 Generar Portable Document Format (PDF)

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
```

2.2 Ver Apendice

apendice.pdf

3 Conclusion

cls

- Un formato abierto
- Estilos compactos y robustos
- Amplia compatibilidad con los requisitos de estilo
- Localización automática de estilos
- Infraestructura para la distribución y actualización de estilos
- Miles de estilos disponibles gratuitamente (con licencia Creative Commons BY-SA)

Por bibtex algunos:

- Soporte completo de Unicode
- Ordenación altamente personalizable mediante el Algoritmo de Cotejo Unicode + adaptación de CLDR
- Etiquetas de bibliografía altamente personalizables
- Compleja modificación de datos sobre la marcha basada en macros sin cambiar las fuentes de datos
- Un modo de herramienta para transformar las fuentes de datos bibliográficos
- Múltiples bibliografías y listas de información bibliográfica en el mismo documento con diferente ordenación
- Reglas de herencia de fuentes de datos altamente personalizables
- Soporte de poliglosia y babel para el cambio automático de idioma de las entradas y citas bibliográficas
- Recodificación automática de datos bibliográficos (UTF-9 -> latin1, macros LATEX -> UTF-8, etc.)
- Fuentes de datos remotas

- Sistema de desambiguación automática de nombres y listas de nombres altamente sofisticado
- Modelo de datos altamente personalizable para que los usuarios puedan definir sus propios tipos de datos bibliográficos
- Validación de los datos bibliográficos con respecto a un modelo de datos
- Bibliografías subdivididas y/o filtradas, bibliografías por capítulo, sección, etc.

contar lo que se hizo

3.1 Futuras líneas de trabajo

3.1.1 Build remoto



3.2 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

4 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