

Universidad Tecnológica Nacional

Facultad Regional Buenos Aires

Escuela de Posgrado

MAESTRÍA EN INGENIERÍA EN SISTEMAS DE
INFORMACIÓN

Dir: Dra. Ma. Florencia Pollo Cattaneo

Seminario

HERRAMIENTAS PARA EL DESARROLLO DE TESIS 2021

Prof: Florencia Pollo Cattaneo

Trabajo Práctico Final

PRODUCCIÓN ACADÉMICA CON PANDOC

Lisandro Fernández

Abstract

Pandoc como entorno textutal de producción de documentos académicos. Evitar el uso de interfaces captivas 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 Pandoc	1
1.1 Pandoc	1
1.2 Texto Plano	1
1.3 Integración	3
2 Resultados	8
2.1 Metodología	9
2.2 Generar Documentos (PDF)	9
2.3 Ver Apendice	9
3 Conclusion	9
3.1 Futuras líneas de trabajo	10
3.2 Debilidades	10
3.3 Indicaciones	11
Referencias	11

1 Producción gráfica de documentos Pandoc

Este proyecto describe el proceso de confección de escritos académicos, de complejidad considerable, sin la necesidad de interfaces gráficas [1]. Cuenta con capacidades necesarias para la producción de documentación técnica y científica, todo el proceso es controlado mediante línea de comandos evitando así depender de interfaces captivas [2, pp. 88–97]

El objetivo de este trabajo es un entorno el cual *Pandoc* “*a universal document converter*” es la pieza central, e interprete del sistema de composición tipográfica y preparación de documentos de alta calidad LaTeX, estándar de facto para la comunicación y publicación de documentos académicos [3, 4].

Con integraciones sencillas se consigue una infraestructura robusta que incluye funciones diseñadas con gestionar la exposición de extensas referencias, múltiples citas bibliografía, notación matemática, generación gráficos y diagramas.

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,

1.2 Texto Plano

Evitar el uso formatos codificados o de poca legibilidad e interfaces captivas beneficia a todos los usuarios, que deben poder encontrar lo que necesitan,

comprender lo que encuentran y usarlo para realizar tareas [5].

habilita el separar contenido, referencias y estilo.

en un contexto de organizaciones con actividades relacionadas a la publicación donde no solo los documentos son complejos sino que también la tarea involucra a múltiples agentes autores/supervisores/editores

devuelve el control estético y da unidad en estilo en la composición gráfica resultante de diversos productos.

De los 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

incluirla en una cadena de desarrollo propia de las software factories

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.

1.2.1 Markdown

Markdown es una *sintaxis plain text formatting syntax* de formato de texto,

El formato de texto es el marcado que se aplica a un texto simple para añadir datos de estilo más allá de la semántica de los elementos: colores, estilos (negrita, itálica), tamaño, y características especiales (como hipervínculos). Al texto resultante se le conoce como texto formateado, texto con estilos, o texto enriquecido

<https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1348>

This article is categorized under: Software for Computational Statistics > Software/Statistical Software

El con énfasis en la legibilidad

de autoría de texto Lo que distingue a Markdown de muchas otras sintaxis de marcado ligero, 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.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. <https://pandoc.org/MANUAL.html#pandocs-markdown>

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 de software 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

un potente sistema para escribir filtros.

Es el diseño modular: consta de un conjunto de lectores, que analizan el texto en un formato determinado y producen una representación nativa del documento (un árbol sintáctico abstracto o AST), y un conjunto de escritores, que convierten esta representación nativa en un formato de destino. Así, para añadir un formato de entrada o de salida basta con añadir un lector o un escritor. Los usuarios también pueden ejecutar filtros personalizados de pandoc para modificar el AST intermedio. pandoc funciona como la plaza central que integra

De los muchas maneras de personalizar pandoc para que se adapte a sus necesidades, este proyecto se destaca el uso de

incluye un sistema de plantillas y

Pandoc incluye un potente sistema de citas y bibliografías automáticas.

Esto significa que puede escribir una cita como véase [6], 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.

sistema de diagramacion y generacion graficos

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

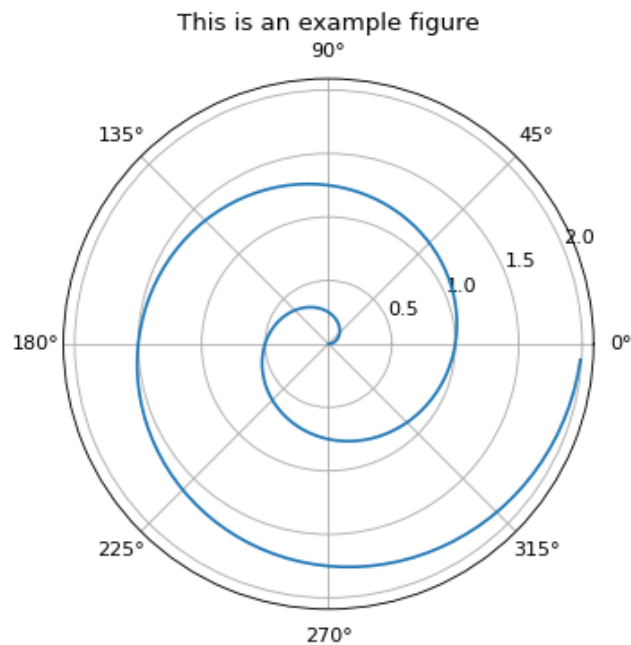
modificar diagramas de forma dinámica.

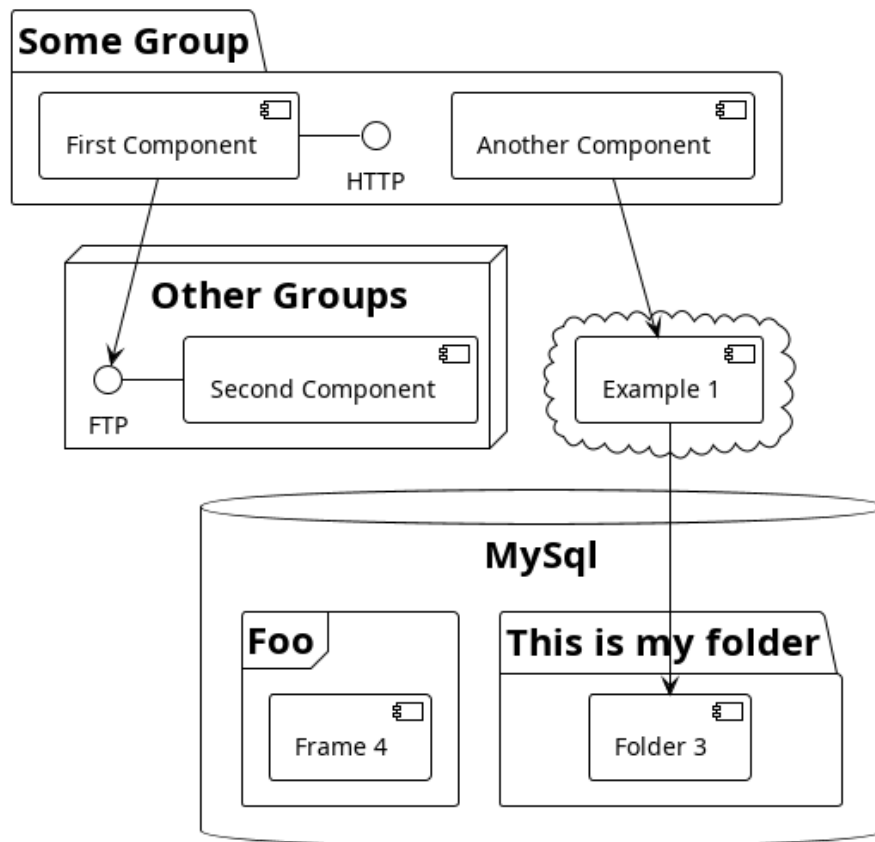
1.3.1 Gráficos y diagramas

Diagramación y la documentación cuestan un tiempo precioso a los desarrolladores y quedan obsoletas rápidamente.

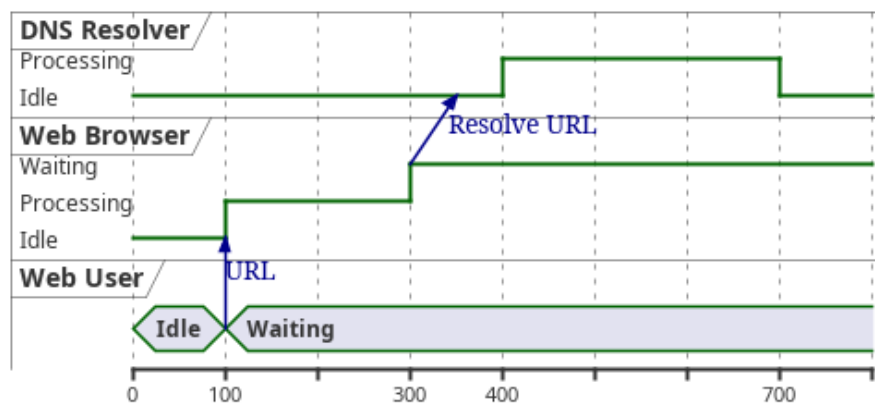
Pero no tener diagramas o documentación arruina la productividad y perjudica el aprendizaje de la organización.

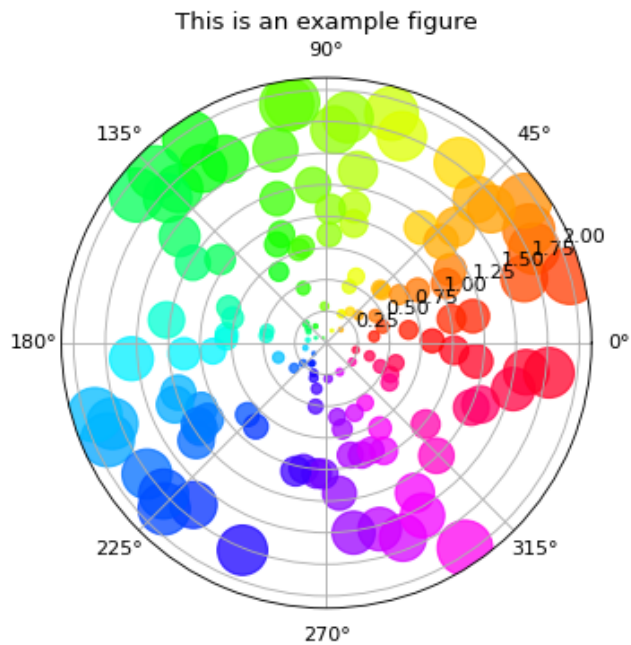
1.3.1.1 **pandoc-plot** <https://laurenttrdc.github.io/pandoc-plot/>





<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.

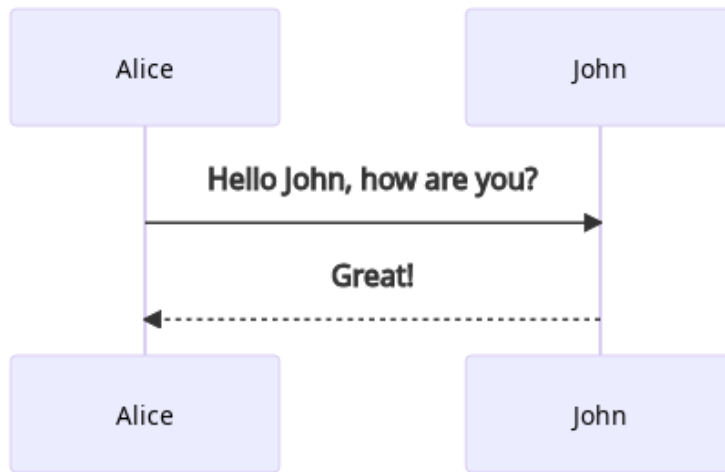
Mermaid aborda este problema permite crear diagramas fácilmente modificables, del código de piezas de software.

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

```

~~~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 <https://www.mendeley.com/guides/csl-editor/>

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 proyecto CSL, citationstyles.org.

Si alguna vez has escrito un trabajo de investigación, habrás trabajado de investigación,

probablemente has incluido referencias a otros trabajos. Las referencias son importantes en la comunicación académica, ya que proporcionan la atribución, enlazan referentes.

Sin embargo, formatear manualmente las referencias puede llevar mucho tiempo, especialmente cuando se trata de múltiples publicaciones con diferentes estilos de citación.

El software de gestión de referencias puede ayudar. no sólo ayudan a gestionar bibliotecas 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 (CSL) 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 la integración de diferentes piezas de software andamiaje necesario para reproducir este proyecto esto incluye:

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

El documento y su contraparte

Para recrear el proceso que se utilizo para confeccionar este documento

2.1 Metodologia

El desarrollo de este trabajo fue realizado

nota al pie: Como conseguir una instalacion funcional de Esta fuera de los alcances de este articulo.

2.2 Generar Documentos (PDF)

Con una version funcionando en ejecutar en la terminal del mismo el siguiente sentencia:

```
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
```

La salida de este comadno produce un documento grafico en formato PDF a partir del LINK de este mismo texto cuya fuente es el markdown

2.3 Ver Apendice

apendice.pdf

3 Conclusion

contar lo que se hizo

Se consigue un cadena de produccion

Características generales

- Un formato libre y abierto
- Componentes aislados, compactos y robustos
- Amplia compatibilidad con requisitos de estilo
- Miles de estilos disponibles predifindos por la comunida
- Etiquetas de bibliografía altamente personalizables
- Fuentes de datos remotas para publicaciones frecuentes con informacion dinámico.

Estilos de citas

- Localización automática de estilos
- Infraestructura para la distribución y actualización de estilos

- Compleja modificación de datos sobre la en produccion, basada en macros sin cambiar las fuentes de datos

3.1 Futuras líneas de trabajo

3.1.1 Operaciones remotas automáticas

<https://pandoc.org/installing.html#github-actions>

disponible el repositorio con



Integración continua <https://about.gitlab.com/features/continuous-integration/>

3.1.2 Revisión sistemática de literatura

Para revisiones de literatura que comprometan múltiples bibliografías y listas de información bibliográfica en el mismo documento con diferente ordenación

- Ordenación altamente personalizable mediante el Algoritmo de Cotejo Unicode (Unicode collation algorithm - UCA) (<https://www.unicode.org/reports/tr10/>)
- Adaptación de CLDR (<https://cldr.unicode.org/>)
- Bibliografías jeraquizadas por capítulo, sección, etc.
- Soporte de poliglosia para el cambio automático de idioma de las entradas y citas bibliográficas
- 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

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 exactas entre todos los formatos. Mientras que las conversiones de Markdown de Pandoc a todos los formatos aspiran a ser perfectas, las conversiones de formatos más expresivos que Markdown de Pandoc pueden tener diferencias.

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 por ejemplo tablas complejas, pueden no encajar en el modelo de documento simple de Pandoc.

Si bien los módulos empleados maduraron y se dirtribuyen dentro con empaquetatos la distribución principal de Pandoc cambiar configuraciones agrerar funcionalidades y rutinas puede variar y depender de versiones de las herramientas en las diferente distribuciones y sistemas operativos puede inferir en la integración y operaciones.

Pero en una implementación organizacional esto puede ser solucionado ejecutando en servidor remoto donde solamente se carguen como insumos los ficheros de contenido y las configuraciones y plantillas estén

3.3 Indicaciones

es un proyecto que sirva como de guía para patrones de diseño y buenas practicas para proyectos similares

si bien este proyecto está enfocado a la producción de literatura académica esta misma cadena puede ser considerado en en el desarrollo de cualquier sistemas de gestión documental, registros médicos, documentos legales, certificados, etc

donde productos gráficos imprimibles se generan mediante rutinas directamente de bases de datos, una capa codificada extra que opaca la relación con el interprete el y el contenido, se recomienda un proceso similar al descripto de respaldo del la información en contenedores de formato simple y legible, sin codificar u opacar habilitando ser manipulados con herramientas obícuas, con acceso directo mediante sistemas rudimentales.

Referencias

- [1] W. Caleb McDaniel, “Why (and how) i wrote my academic book in plain text – w. Caleb McDaniel,” *W. Caleb McDaniel*. Available: <http://wcaleb.org/blog/my-academic-book-in-plain-text>
- [2] M. Gancarz, *Linux and the unix philosophy*. Elsevier Science, 2003. Available: <https://books.google.com.ar/books?id=qqstCSlk5MIC>
- [3] john MacFarlane, “Pandoc - a universal document converter,” *Pandoc - a universal document converter*. 2022. Accessed: Sep. 14, 2022. [Online]. Available: <https://pandoc.org/>
- [4] D. E. Knuth, D. Knuth, and D. Bibby, *The TeXbook*. Addison-Wesley, 1986. Available: www-cs-faculty.stanford.edu/~knuth/abcde.html
- [5] D. A. S. U. Harvard, “Use plain language,” *Digital Accessibility*. Digital Accessibility Services. Available: accessibility.huit.harvard.edu/use-plain-language
- [6] B. Moolenaar, “Seven habits of effective text editing,” 2000. mooleenaar.net/habits.html
- [7] B. A. Kitchenham, D. Budgen, and O. P. Brereton, “Using mapping studies as the basis for further research – a participant-observer case study,” *Information and Software Technology*, vol. 53, pp. 638–651, Jun. 2011, doi: [10.1016/j.infsof.2010.12.011](https://doi.org/10.1016/j.infsof.2010.12.011).
- [8] N. Guarino, G. Guizzardi, and J. Mylopoulos, “On the philosophical foundations of conceptual models,” *Frontiers in Artificial Intelligence and Applications*, vol. 321, pp. 1–15, Dec. 2020, doi: [10.3233/FAIA200002](https://doi.org/10.3233/FAIA200002).