

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	2
2	Metodotología	3
2.1	Integración	4
2.2	Gráficos y diagramas	4
2.3	Bibliografías sofisticadas con BibLaTeX	8
2.4	MathJax	9
3	Resultados	10
3.1	Cheat Shet and examples	10
4	Conclusión	10
4.1	Objetivos	11
4.2	Aporte	11
4.3	Limitaciones	11
4.4	Futuras lineas de trabajo	12
	Referencias	12

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 linea de comandos evitando así depender de interfaces captivas [2, pp. 88–97]

El objetivo de este trabaja es un entorno el cual *Pandoc* “*a universal document converter*” es la pieza central, 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].

Mediante integraciones sencillas se consigue una infraestructura robusta con funciones diseñadas para gestionar exposición de extensas referencias, múltiples citas y bibliografía a diferentes fuentes; notación matemática, generación gráficos y diagramas, entre otras capacidades.

1.1 Pandoc

Pandoc es una biblioteca *Haskell* [5] para convertir de un formato de marcado ligero a otro, y una herramienta de línea de comandos que accede a las funciones en esta biblioteca para convertir entre numerosos formatos y de procesamiento de textos.

La biblioteca incluye módulos separados para cada formato de entrada y salida, por lo que para añadir un nuevo formato de entrada o salida sólo hay que añadir un nuevo módulo.

1.2 Texto Plano

Evitar el uso de 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 [6]. Habilita el separar contenido, referencias y estilo.

En un contexto de organizaciones con actividades relacionadas a la publicación donde la complejidad no solo de documentos sino que también lo es la tarea ya involucra a múltiples agentes como autores, correctores y editores,

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

incluirllo en una cadena de desarrollo de propia de las software factories

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

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 doc-

umento 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.

2 Metodología

En este capítulo se describe el método utilizado

Primero se introduce al método de integración de diferentes herramientas

luego gráficos

luego citas y referencias

luego notación matemáticas

El desarrollo de este trabajo fue realizado nota al pie: Como conseguir una instalación funcional de Esta fuera de los alcances de este artículo.

distribuidas con junto con pandoc y otras aportes de la comunidad distribuidos independientemente.

Para la gestión de bibliografía Por ejemplo, esto significa que puede escribir una referencia como @moolenaar2000 o también [@knuth1986texbook p.3-9] y Pandoc la convertirá en una cita con el formato predefinido, utilizando cualquiera de los cientos de Lenguajes de Estilo de Cita (Citation Style Language - CSL) incluyendo estilos de nota al pie, numéricos y autoría, fuente y fechas), 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 diagramación y generación de gráficos

permite crear diagramas y visualizaciones utilizando texto y código. Se trata de una herramienta de diagramación y graficación

que renderiza definiciones de texto inspiradas en Markdown para crear y modificar diagramas de forma dinámica.

2.1 Integración

Un potente sistema para escribir filtros. El diseño modular: consta de un conjunto de lectores, que analizan el texto en un formato determinado y producen una representación nativa del documento (Abstract Syntax Tree - AST), y un conjunto de escritores, que convierten esta representación nativa en un formato de destino.

De las múltiples maneras de personalizar *Pandoc* para que se adapte a los requisitos de cada proyecto, se destaca el uso de un sistema de plantillas y un potente sistema de citas y bibliografías automáticas y la generación de gráficos mediante código.

2.2 Gráficos y diagramas

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

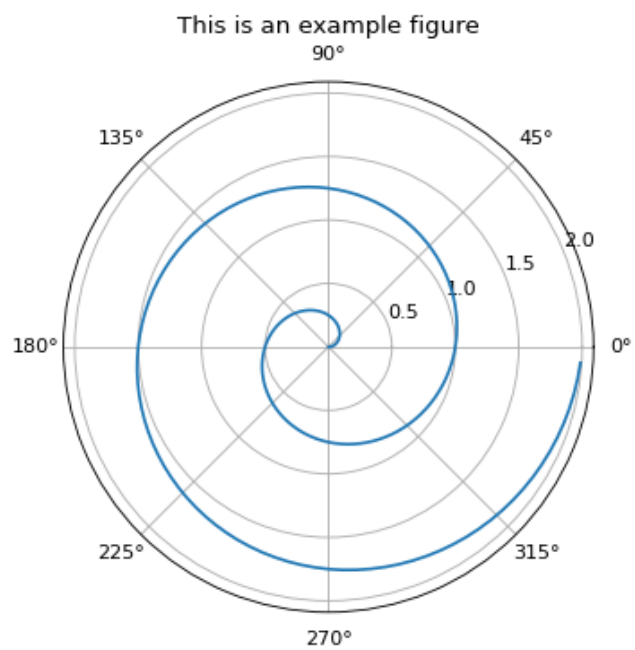
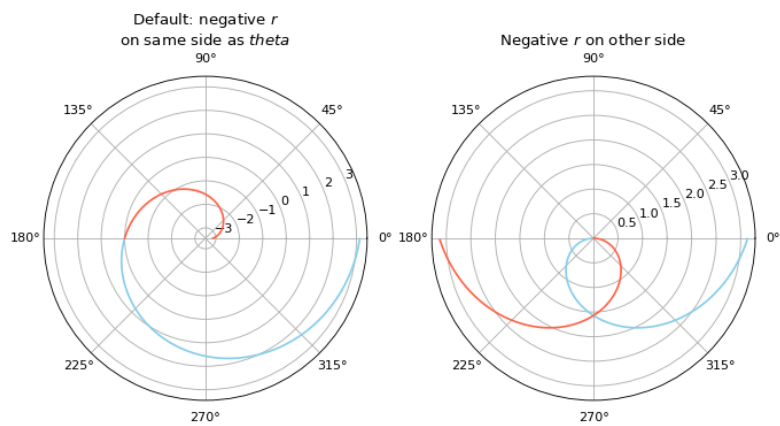
Pero no tener diagramas o documentación arruina la productividad y perjudica el aprendizaje de la organización.

2.2.1 pandoc-plot

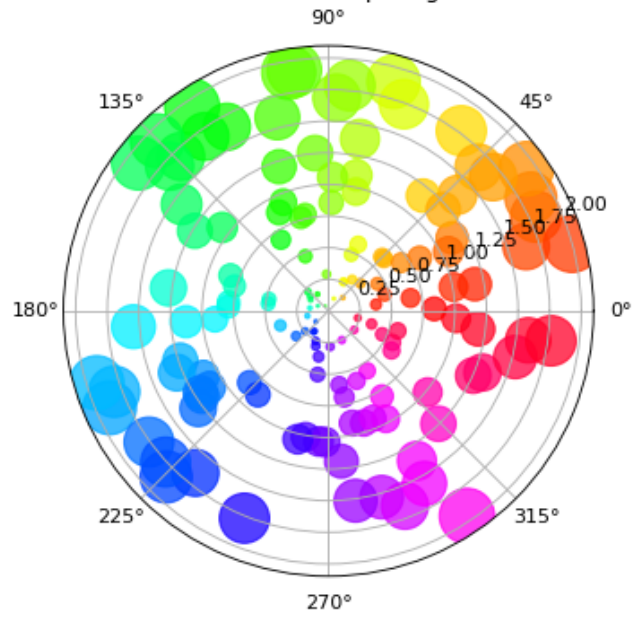
pandoc-plot es un filtro de *Pandoc* para generar figuras a partir de bloques de código en documentos.

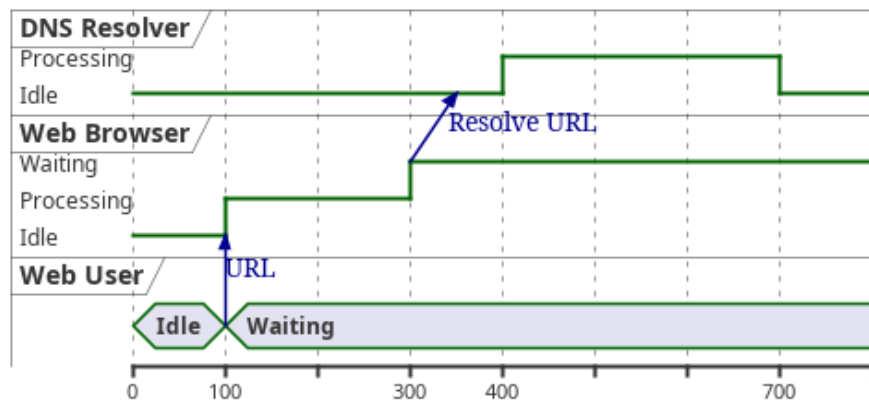
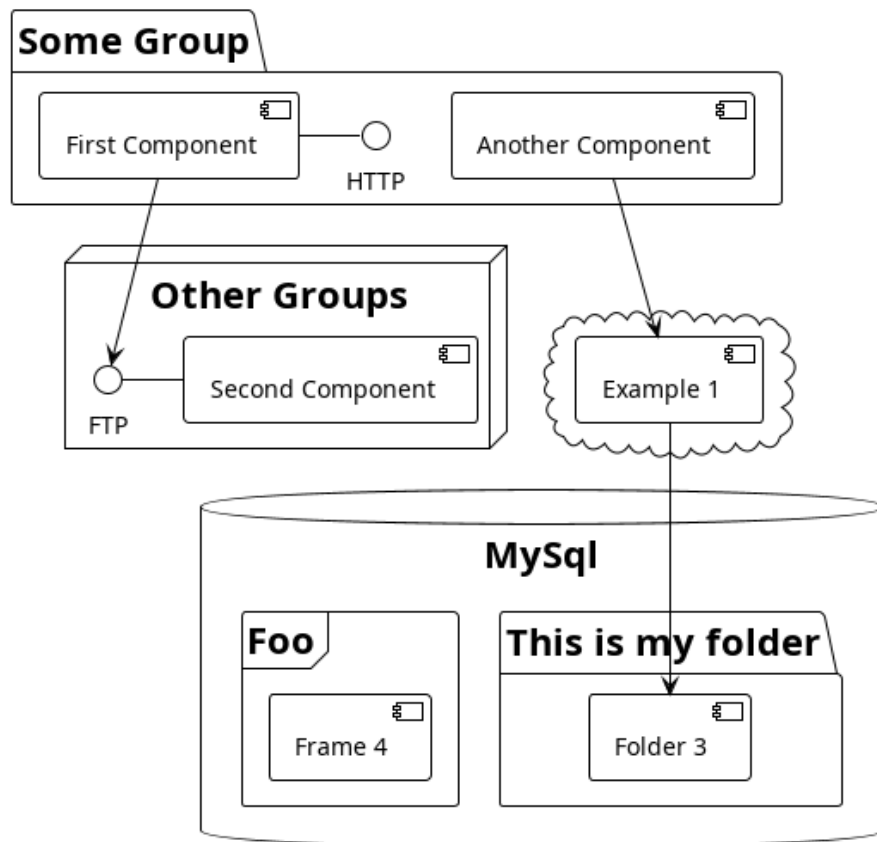
Esta página está generada a partir de *Markdown* utilizando *pandoc-plot*, para que puedas hacerte una idea de lo que es posible.

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



This is an example figure





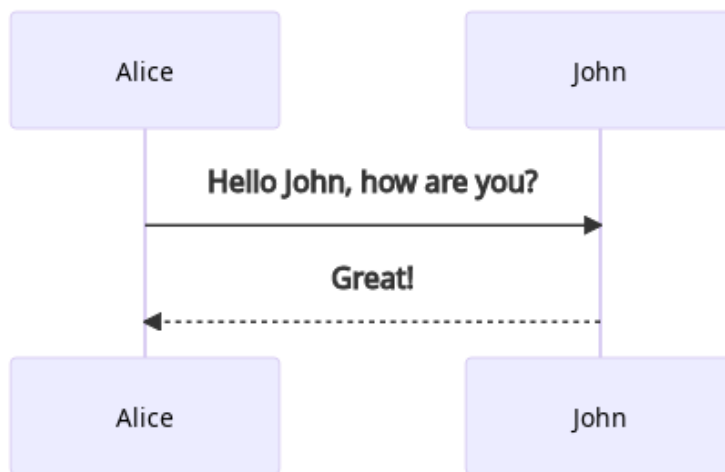
2.2.2 MermaidJS

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!
~~~
```



2.3 Bibliografías sofisticadas con BibLaTeX

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.

2.3.1 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 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 (Citation Style Language -CSL) es ese lenguaje

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

2.4 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.

3 Resultados

El producto de este proyecto la integración de diferentes piezas de software y andamiaje necesario para reproducir este proyecto esto incluye: Configuraciones, estructura ejemplo filtros, cls, plantilla LaTeX, resaltado de código, configuraciones para operaciones remotas

Un ejemplo de como usar esta herramienta es el mismo comando que sirve para conseguir este documento desde su fuente en *Markdown* [markdown?].

```
$ pandoc README.md \
-F pandoc-plot --metadata-file=metadata.yaml --mathjax \
-F mermaid-filter -F pandoc-crossref --citeproc \
--highlight-style pygments.theme \
--template=plantilla --pdf-engine-opt=-shell-escape \
-s --toc --toc-depth=2 --number-sections --columns=80 \
-o README.pdf
```

Ejecutar el comando precedente en una sistema con una instancia de *Pandoc* funcionando produce las este documento y declara: plantillas: General LaTeX, lenguaje de estilo de citas, resaltado de código. Módulos extra como filtros: Notación matemática, gestión de Referencias, citas y generación de gráficos. Entre otras configuraciones generales involucradas en el proceso, la profundidad la tabla de contenidos. archivo de entrada y de salida.

3.1 Cheat Shet and examples

apendice.pdf

El documento y su contraparte

Para recrear el proceso que se utilizo para confeccionar este documento

4 Conclusión

Este capítulo concluye el estudio. El objetivo de esta tesis es investigar el proceso de producción de documentos con herramientas rudimentales. En primer lugar, se responde a las preguntas de investigación. El segundo subcapítulo presenta la contribución de esta tesis, y los dos últimos subcapítulos presentan las limitaciones del estudio y las sugerencias para desarrollos futuros, respectivamente.

4.1 Objetivos

Se consigue un cadena de producción

Características generales: un formato libre y abierto, componentes aislados, compactos y robustos; amplia compatibilidad con requisitos de estilo, miles de estilos disponibles predefinidos por la comunidad, etiquetas de bibliografía altamente personalizable y fuentes de datos remotas para publicaciones frecuentes con información dinámico.

Citas y citas: localización automática de estilos, infraestructura para la distribución y actualización de estilos, compleja modificación de datos sobre la producción y basada en macros sin cambiar las fuentes de datos.

4.2 Aporte

Es intención que este proyecto que sirva como plantilla para proyectos similares, patrones de diseño y guía de buenas practicas para la producción de documentos gráficos de alta complejidad.

Si bien que este proyecto está enfocado a la producción de literatura académica, esta misma cadena puede ser considerada 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 de la información en contenedores de formato simple y legible, sin codificar u opacar habilitando ser manipulados con herramientas obicuas, con

Pero en una implementación organizacional esto puede ser solucionado ejecutando en servidor públicos donde con acceso web como un servicio de preparación de documentos gráficos, tanto preparando para impresión o para ser consumidos mediante exploradores de internet solamente se carguen como insumos los ficheros de contenido o

Aunque los escuadrones son autónomos, es importante que los especialistas (por ejemplo, los desarrolladores de) se alineen en las mejores prácticas.

4.3 Limitaciones

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.

Los módulos empleados maduraron y se incluyen en la distribución principal de *Pandoc* cambiar configuraciones agregar funcionalidades y rutinas puede variar y depender de versiones de las herramientas en las diferentes distribuciones y sistemas operativos puede inferir en la integración y operaciones.

4.4 Futuras líneas de trabajo

Se señala como áreas de desarrollo

4.4.1 Operaciones remotas automáticas

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

disponible el repositorio con



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

4.4.2 Revisión sistemática de literatura

Para revisiones de literatura que comprometen múltiples bibliografías y [7]

Que deben 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

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] S. Marlow *et al.*, “Haskell 2010 language report,” *Available online http://www.haskell.org/(May 2011)*, 2010.
- [6] D. A. S. U. Harvard, “Use plain language,” *Digital Accessibility*. Digital Accessibility Services. Available: accessibility.huit.harvard.edu/use-plain-language
- [7] B. Kitchenham, “Evidence-based software engineering and systematic literature reviews,” in *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 2006, vol. 4034 LNCS, p. 3. doi: [10.1007/11767718_3](https://doi.org/10.1007/11767718_3).