

A Short Guide to Destroying the Modern Information Marketplace

Aerith Netzer

Digital Publishing and Repository Librarian

Northwestern University Libraries

aerith.netzer@northwestern.edu

February 1, 2024

Intro

Aerith
Netzer

Background

FOSS

Guerilla Open
Access

Vision

Citations

Possibilities

Solutions

Future
Work

Thanks!

This Presentation's \LaTeX File is Available on GitHub

- https://www.github.com/aerithnetzer/OpenAI_Presentation

The Code Demonstrated in this Presentation is Available on GitHub

- <https://github.com/aerithnetzer/py-bats>
- <https://github.com/nulib-ds/publishing-template>
- <https://github.com/aerithnetzer/djournal>

Intro

Aerith
Netzer

Background

FOSS

Guerilla Open
Access

Vision

Citations

Possibilities

Solutions

Future
Work

Thanks!

Information is power. But like all power, there are those who want to keep it for themselves. The world's entire scientific and cultural heritage, published over centuries in books and journals, is increasingly being digitized and locked up by a handful of private corporations. Want to read the papers featuring the most famous results of the sciences? You'll need to send enormous amounts to publishers like Reed Elsevier. (...) We need to take information, wherever it is stored, make our copies and share them with the world. We need to take stuff that's out of copyright and add it to the archive. We need to buy secret databases and put them on the Web. We need to download scientific journals and upload them to file sharing networks. We need to fight for Guerilla Open Access. With enough of us, around the world, we'll not just send a strong message opposing the privatization of knowledge — we'll make it a thing of the past. Will you join us? Guerilla

Aaron Swartz, 2008
Internet Archive, Guerilla Open Access Manifesto

Intro

Aerith
Netzer

Background

FOSS

Guerilla Open
Access

Vision

Citations

Possibilities

Solutions

Future
Work

Thanks!

Put the big publishers out of business.

Wiley, Elsevier, Springer, Taylor & Francis, Sage, and others will not be able to compete with self-funded operations like us.

Intro

Aerith
Netzer

Background

FOSS

Guerilla Open
Access

Vision

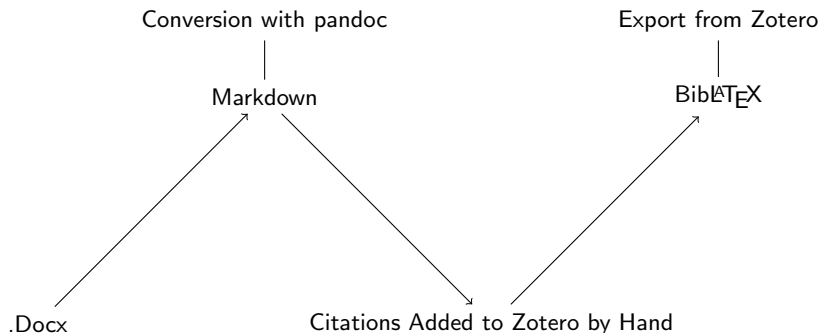
Citations

Possibilities

Solutions

Future
Work

Thanks!



The Current State is Not Scalable

- The current process is time-consuming
- The current process is error-prone
- The current process is not scalable
- Reliant on inaccessible software

Intro

Aerith
Netzer

Background

FOSS

Guerilla Open
Access

Vision

Citations

Possibilities

Solutions

Future
Work

Thanks!

Northwestern Insider to be Published by Northwestern University Libraries!

Website created in two hours using Aerith's abstracted template of Chris Diaz's code.

Intro

Aerith
Netzer

Background

Solutions

O and A

Websites

JMS

Problem &
Solution

py-bats

Future
Work

Thanks!

Operationalize

- Extend and abstract the current template created by Chris Diaz
- Host our own Open Journal Systems instance for future journals
- Create standard operating procedures for journal production and management

Automate

- Facilitate one-click (or few clicks) journal deployments
- Minimize reliance on outside vendors that we have to wait on to get things done
- Automate citation management, format construction, and metadata appending

Intro

Aerith
Netzter

Background

Solutions

O and A

Websites

JMS

Problem &
Solution

py-bats

Future
Work

Thanks!

<https://github.com/nulib-ds/publishing-template>

```
47
48 # enable hero section's is-bold effect
49 heroBold = true
50
51 # enable PWA, prepare your icons and DON'T forget to modify manifest.json
52 pwa = true
53
54 # disable jdeliver cdn, default false
55 noCDN = false
56
57 # international standard serial number (issn) for the journal
58 issn = "0000-0000"
59
60 # relative path to the favicon image file, name the file favicon.png and put it in the static/img folder
61 favicon = "/img/favicon.ico"
62
63 # Wordmark image file, name the file wordmark.png and put it in the static/img folder
64 wordmark = "/img/wordmark-invert.png"
65
66 # Lockup image file, name the file lockup.png and put it in the static/img folder
67 lockup = "/img/lockup-invert.png"
68
69
70 # to enable different highlight themes in light/dark mode
71 [markup]
72 [markup.highlight]
73   noClasses = false
74
75 [menu]
76 [[menu.main]]
77   identifier = "about"
78   name = "About"
79   url = "/about/"
80   weight = 1
81 [[menu.main]]
82   identifier = "browse"
83   name = "Browse"
84   url = "/issues"
85   weight = 2
86 [[menu.main]]
87   identifier = "authors"
88   name = "Guidelines"
89   url = "/guidelines/"
90   weight = 3
91 [[menu.main]]
92   identifier = "styleguide"
93   name = "Style Guide"
94   url = "/style-guide/"
95   weight = 4
96 [[menu.main]]
97   identifier = "submit"
98   name = "Submit"
99   url = "/submit/"
100   weight = 5
101
102 [taxonomies]
103 category = "categories"
```


[Intro](#)[Aerith
Netzer](#)[Background](#)[Solutions](#)[O and A](#)[Websites](#)[JMS](#)[Problem &
Solution](#)[py-bats](#)[Future
Work](#)[Thanks!](#)

Djournal

- A django application for journal management systems
- Homegrown
- Lightweight, elastic, and no more icky, icky PHP

This is nowhere close to ready for deployment.

Intro

Aerith
Netzer

Background

Solutions

O and A

Websites

JMS

Problem &
Solution

py-bats

Future
Work

Thanks!

The Problem of Plaintext Citations

- Plaintext citations are not machine-readable
- Many different citation styles cause confusion
- Citations are often not linked to the cited work

Workaround: OpenAI GPT-3.5

- Parse citations from article submission
- Create a .txt file with the citations in plaintext
- Call the OpenAI API to generate a Bib \LaTeX file

The Optimal Solution: Bib \LaTeX Evangelising

- Create resources that help create .bib files
- Communicate the importance of Bib \LaTeX to researchers
- Create a .tex file with the citations in Bib \LaTeX

Intro

Aerith
Netzer

Background

Solutions

O and A

Websites

JMS

Problem &
Solution

py-bats

Future
Work

Thanks!

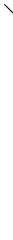
Clean and Isolate
Plaintext Citations

Create .tex File
with Bib^LA_TE_X

Docx Conversion
to Markdown

Call OpenAI API to
Generate Bib^LA_TE_X

Automate citation additions
to markdown file



Intro

Aerith
Netzer

Background

Solutions

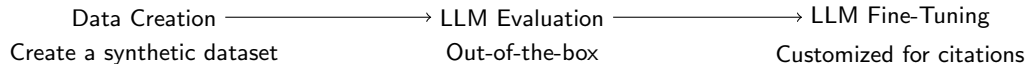
Future
Work

The Roadmap

Journal
Articles

Dev Work

Thanks!



Intro

Aerith
Netzer

Background

Solutions

Future
Work

The Roadmap

Journal
Articles

Dev Work

Thanks!

Evaluating Large Language Models for Transforming Plaintext Citations to Machine-Readable Formats

- Highlight the importance of Bib \LaTeX for citations and parsing by Google Scholar
- Create a synthetic dataset of plaintext citations and their Bib \LaTeX equivalents across many citation styles
- Evaluate the effectiveness and economic efficiency of large language models out-of-the-box
- Fine-tune a large language model to create a citation parser and measure its effectiveness and economic efficiency

A Modern, Maintainable, and Scalable Diamond Open Access Journal Publishing Workflow

- Present hugo templating for journal websites
- Provide code, resources, and documentation for other libraries to spin up their own operations

[Intro](#)[Aerith
Netzer](#)[Background](#)[Solutions](#)[Future
Work](#)[The Roadmap](#)[Journal
Articles](#)[Dev Work](#)[Thanks!](#)

Implementing Error Correction in Crossref API

- Relatively often, HTML code is returned when Bib \LaTeX is requested
- Reduces effectiveness of the citation parser
- Error correction could be implemented to decrease confusion

Intro

Aerith
Netzer

Background

Solutions

Future
Work

Thanks!

Chris Diaz

Never met the guy but I figure he should get some credit for doing all the hard work before me

David Schober

For trusting me with access to GPT-3.5

Intro

Aerith
Netzer

Background

Solutions

Future
Work

Thanks!

Fin.