



Project Proposal

Converting past R Journal articles to HTML

Project Info

Project Title:

Converting past R Journal articles to HTML

Project Short Title

R Journal article conversion

Project Synopsis:

The Project Converting past R Journal articles to HTML aims to convert **.tex** files to an **HTML** version, based on the LaTeX source files and metadata.

URL of Project Idea Page:

<https://github.com/rstats-gsoc/gsoc2022/wiki/Converting-past-R-Journal-articles-to-HTML>

About Me

I am A Adarsh, a sophomore student pursuing my Bachelor's Degree in Computer Science Engineering from Vidyavardhaka College of Engineering, Mysore, India. I had previously participated in the Google Code-In Contest 2019-20 and was acquainted about the R Project for Statistical Computing.

I'm an Ardent Programmer and Web Developer. I have been following The R Project for the past few years and was astonished by its computing skills in all the fields and would love to continue contributing to the R Programming language. GSoC would be a sublime opportunity for me to go ahead with it. Now I have been enhancing my R Skills with various other languages and Tech stacks such a **C/C++**, **Python**, **HTML**, **CSS**, **JavaScript**, **Go**, **Git**, etc.

GSoC would enable me to level up my journey. Open Source organizations and their tools would help me a lot throughout my journey, it would be my pleasure if I could learn and develop under your precious guidance and mentorship for an esteemed organization like R Project for Statistical Computing.

I went through your project details and found it very insightful and in my skill sets. It would be a great opportunity of learning and developing with you for the project Converting past R Journal articles to HTML.

Contact Information

Name: A Adarsh

Postal Address: 2981, 6th Cross, F Block, Kanakadasanagar, Mysore, Karnataka, India 570022

Email ID: adarshgowda463@gmail.com

Contact Number: +91 8660094048

GitHub: <https://github.com/Adarshgowdaa>

LinkedIn: <https://www.linkedin.com/in/a-adarsh-76695b17a/>

Affiliation

Institution: Vidyavardhaka College of Engineering, Mysore

Program: Bachelor's Degree in Computer Science Engineering

Stage: Sophomore

Contacts to Verify: Dr. Ravikumar V (Professor & Head), Email ID: ravikumarv@vvce.ac.in

Schedule Conflicts:

I have no upcoming plans for this summer except GSoC, Hence there shouldn't be any conflicts in the schedule from my side and I assure my availability for the required amount of hours in the GSoC 2022.

Mentors

Evaluating Mentor: Heather Turner, Email ID: heather.turner@r-project.org

Co-Mentor: Di Cook, Email ID: dicook@monash.edu

Have you been in touch with the mentors? When and how?

Yes, I have been in touch with the mentors from the start of the application process and have been communicating with them through the means of Email for discussion about the project idea as well as clarifying the doubts regarding the application process.

Project Details

Introduction

The R Journal is the open access, refereed journal of the R project for statistical computing. It features short to medium length articles covering topics that should be of interest to users or developers of R. The R Journal intends to reach a wide audience and have a thorough review process. Papers are expected to be reasonably short, clearly written, not too technical, and of course focused on R.

Since it began in 2009, articles have been created in PDF, using LaTeX to convert marked up plain text to the final version with the formatted text, tables, figures, etc. Over the past year, the R Journal has been developing a new website, with the option to publish articles in HTML format. The main goal of the project is to convert published articles to HTML.

Coding Plan and Methods

To convert past R Journal articles to HTML, We will be using Pandoc. It is a free and open-source document converter, widely used as a writing tool and as a basis for publishing workflows. Using this open-source tool the LaTeX source files are converted to markdown in the first instance.

Further, for the next part of the project, we will be focusing on building custom functions to convert the markdown to R markdown using the **rjtools** template.

After thorough scripting of the function, We will be using the rjtools package to convert the generated R markdown file to HTML file.

Later, we will be concluding the project by preparing a thorough vignette of the package including all the details in depth.

Benefits to Community

Journal of the R Project for statistical computing consists of articles covering various topics that should be of interest to the R developers. All the articles add great value to the R Project. The option to publish articles in HTML format means that authors can include interactive graphics and animations in their articles. Also, the HTML format is easier to browse online and more accessible for users of assistive tools and technologies such as screen readers. This will help the community remarkably

From the above proposed project, R Project for statistical computing can be used in a more efficient way with better understanding.

Timeline

Having an organized schedule is the most important factor when it comes to handling a project as it increases productivity and allows us be consistent.

I have divided the the project timeline into 5 Phases, Each phase contains **2 to 3 weeks** each and it's respective deadlines.

I understand the valuable time that the mentor's volunteer for the project, hence the weekly discussions and the evaluations that would be required will be done as and when they are most suited and comfortable according to them.

Community Bonding:

May 20, 2022 - June 12, 2022

```
print("Hello World to the R Community")
```

- I would use this time frame to get to know more about the R Community & it's culture.
- Interact with my peers & the mentors and trying to understand different work environments.
- Get ready for the project by bridging up for any required skill gap.

Phase 1:

June 13, 2022 - June 26, 2022

- Discussing base factors with the mentors for the function.
- Creating a base function.
- Implementing **Pandoc** to Convert the file to **markdown**
- Implementing Code in the function to convert **markdown** to **R markdown** using **rjtools** template.

Phase 2:

June 27, 2022 - July 10, 2022

- Checking for special cases to be handled.
- Creating a base package.
- Starting with Documentation and Simple Test run of the Package.

Phase 3:

July 11, 2022 - July 24, 2022

- Checking for special cases (e.g. adding all the required YAML fields to the top of the R markdown file, or handling more complex articles with figures or equations).
- Completing about 50-60% of the modules before the deadline of first evaluations.
- Thorough testing of the created package.

Phase 4:

July 25, 2022 - Aug 13, 2022

- Writing a Vignette containing all the information.
- Starting with Converting the past articles.
- Checking for special cases to be handled in the function.
- Carefully Checking for correctness of the output generated.

Phase 5:

Aug 14, 2022 - Sept 4, 2022

- Continue Converting the articles. (if not completed in Previously)

- Handling Special Cases (e.g. adding alternative text for images).
- Discussing with the mentors and creating pull request to add articles to the R Journal Website.

Wrapping Up:

Sept 5, 2022 - Sept 12, 2022

- Finalizing the Code checks and documentations.
- Getting final reviews and checks from fellow peers.
- Submit to Google team.

Related Work

I started researching unfamiliar document types described in the project and understood the R Journal Template for Writing an article. I tried fiddling with an R Package called **rjtools**. I came across the `create_article()` function from the package.

Just not limiting my learning to R Journal and its related Packages and Modules. I tried brushing my knowledge and understanding the concepts of the R language, package testing and management, creating a vignette for packages.

Test Link - https://github.com/adarshgowdaa/GSoC__2022--R__Journal__articles__to__HTML-tests

Why Me?

I have a good amount of knowledge in the R Markdown and adhere to the good coding standards prescribed by Google's R style guide. Apart from this I am familiar with both Git and Github and have been using them for quite some

time. Also, I have complete working familiarity in both windows and the Linux kernel Operating Systems such as Ubuntu.

Apart from these, I am also the part of IEEE Student Representative Committee. Being part of this committee explains to me the importance of an article.

I have also tried to fulfill all the criteria required for the project and would continue to gain & upskill my knowledge throughout the whole journey.

Post GSoC

GSoC would be an initial part of the journey, but I believe open source is a never ending wonderful & insightful journey.

I will continue maintaining and developing what we would have started and also continue upgrading and contributing to the project and the community. I'll also be looking for more efficient modules and functions for upgrading as the R Project is a rapidly developing community.

This document is created using RMarkdown by A Adarsh