GSoD 2021 Project Proposal

NumPy: High-Level Restructuring and End-User Focus

Personal Information

My name is Mukulika Pahari (she/her). I am currently a Computer Engineering undergraduate student at Ramrao Adik Institute of Technology, India.

Email Address: mukulikapahari@gmail.com.
GitHub profile: @Mukulikaa
Personal blog: mukulikapahari.medium.com
LinkedIn profile: linkedin.com/in/mukulikapahari
Location: Navi Mumbai, India
Time Zone: (UTC + 5:30) Indian Standard Time

Please find my resume here.

Professional Information

Writing Samples

Here are a few writing samples:

1. Developer Documentation for Setting up ElasticSearch in CDLI
2. *"Harnessing the Power of Big Data for COVID-19"* in Technophilia 2021 (Pg. 3,4)
3. Starship Mission - What is Spacex up to?
4. CRISPR-Cas9: Road to an Artificial Evolution
5. GeForce RTX 30 Series - Revolutionising Gaming
6. New Tech Club in Town: DSC-RAIT

Roles and Responsibilities

- I have been the Co-Editor-in-Chief of Decrypt, our college's open-source technical blog, since September 2020. We also set up a mirrored publication on Medium. My chief responsibilities include:
  1. Mentoring a team of 10 authors and helping them write blogs every week

2. Editing and sending blogs through a review team consisting of our college professors
3. Publishing approved blogs regularly

● My fellow editors-in-chief and I published our annual newsletter [Technophilia](#) on 24th March 2021. We moulded the newsletter to the theme *Tech in Shining Armour.* This endeavour included:
    1. Planning a relevant and captivating theme
    2. Collaborating with interested writers
    3. Carefully editing the articles and going through a rigorous approval process
    4. Collaborating with the Design Team for designing the newsletter
    5. Officially launching the published newsletter

● I have been volunteering as a content creator at Developer Students Club, RAIT and writing required marketing pitches since October 2020.

Previous Knowledge of NumPy

I completed the course *Zero to Pandas* on [Jovian](#) that included basic concepts of NumPy. It also required a course project which I chose to do on *Solar Power Generation Analysis*. My course notebooks are uploaded to this [repository](#).

Previous Experience with Open Source Software

● Designing pages and fixing minor bugs in [Decrypt](#), the technical blog
    1. [PR #16](#)
    2. [PR #15](#)

● Creating preliminary documentation for setting up ElasticSearch in the CDLI Framework
    1. [MR !282](#)

Project Statement

Project Title: High-Level Restructuring and End-User Focus

Project Description

Overview

The existing documentation of NumPy is quite vast and contains a lot of information. A properly structured and organised documentation will help new and experienced users utilise maximum information from the documentation. The NumPy community already has a clear vision for the future of their documentation in [NEP 44](). Along those lines, I propose to focus on reorganising NumPy Fundamentals and creating content for How-Tos and Tutorials for the next six months.

NumPy Fundamentals

According to the [Diataxis]() framework, this section corresponds to the Explanations. Currently, some explanations are mixed in with the NumPy Reference, leading to issues like [#16924]() and [#15793](). These explanations can have their dedicated pages in NumPy Fundamentals.

NumPy How-Tos

According to the SciPy survey in 2019, How-Tos and Tutorials were the most requested content in the documentation. In issue [#15567](), a need for proper guidance on how to navigate the codebase has been expressed. A How-To can be added based on this insight. It will surely help newcomers to kickstart their code contributions.

NumPy Tutorials

The current tutorials on the [NumPy tutorials]() page follow an engaging story-based approach. To motivate amateur users of this tool, it will be interesting to have tutorials on the applications of NumPy on concepts that they have most definitely encountered in their formative education. Tutorials on topics like *"Identifying elements from their spectra"* and *"Statistical analysis of air pollution before and after lockdowns"* could be added. *"Vegetation cover analysis with satellite imagery data"*, *"Analysing Keeling Curve data with Fourier transforms"*(ref: NEP 44) and *"Random Forest Classifier with NumPy"* also seem like good topics.

I have had some discussions with a few community members and they have indicated that these ideas would indeed benefit the organisation. I would love to contribute to NumPy through this project or even as a regular contributor.

Project Timeline

Phase 0 (Before 17th May 2021)
1. Bonding with the community and setting up preferred communication channels and times
2. Setting up the development environment and becoming familiar with the workflow
3. Researching and learning about NumPy

Development Period - 17th May to 16th November 2021
Phase 1 (4 weeks)
1. Analysing parts of current documentation (User Guide and NumPy Reference) according to the type of content (Explanation, How-To, Tutorial & Reference), uniqueness, crosslinking etc.
2. Preparing a report on my findings

Outcome: A detailed report to aid the implementation of the project. It would also help in triaging current documentation issues.

Phase 2 (9 weeks)
1. Restructuring and reorganising current NumPy Fundamentals and NumPy Reference in the User Guide based on Phase 1 report
2. Adding relevant pages to NumPy Fundamentals
3. Navigation testing to prevent broken links

Outcome: Visibly cleaner and distinct sections in User Guide and additional content in NumPy Fundamentals.

Phase 3 (6 weeks)
1. Writing a How-To based on *"How to navigate the NumPy codebase"*
2. Review and revision iterations with SMEs

Outcome: One How-To will be added to the User Guide.

Phase 4 (7 weeks)
1. Writing a tutorial on one of the topics mentioned above that will be chosen based on community discussions

2. Review and revision iterations with SMEs

Outcome: One Tutorial will be added to the [NumPy tutorials](#) page.

Bonus: If the proposal goals are met before the projected deadline, I would love to continue working on adding more tutorials.