

GSoD 2021 Project Proposal

Restructuring the NumPy Documentation

Mentors : Melissa Mendonça and Ross Barnowski

Abstract

The aim of good user documentation is to assist end users in using a product or service. Good user documentation is critical because it allows users to learn how to use software, as well as its functionality, tips, and tricks, as well as solve common problems that arise when using it. Without a good user documentation the user might not be able to do the above things efficiently. A good documentation plays a vital role in ensuring good communication to the end user.

Current State

Currently the [NumPy documentation](#) page consists of two main sections , one for new users and beginners and other is for developers and contributors. In its current form the documentation needs more clarity,more examples and restructuring. There are over 100 issues in the GitHub tracker related to documentation , all of them need to be fixed

Why is the proposed documentation an improvement over the current one?

The proposed user documentation will be organized in such a way that it will maximize and maintain performance, accuracy, and peace of mind for any end user.The new documentation will remove as much ambiguity and duplication as possible , which will improve searchability and usability. The newly adopted Diátaxis Framework will greatly enhance the quality.

Diátaxis Framework

The Diátaxis framework aims to solve the problem of structure in technical documentation. It adopts a systematic approach to understanding the needs of documentation users in their cycle of interaction with a product. As per discussion with the community and NumPy enhancement proposal ([NEP44](#)), the updated documentation will follow Diátaxis framework.

Project Plan

As discussed earlier new documentation will be divided into four main sections according to the Diátaxis framework that is

- Tutorials
- How-to guides
- Explanations
- Reference guide

The core project plan will be in close alignment with the [NumPy enhancement proposal](#).

Additional Works : The NumPy documentation could use a little bit of page UI improvement. Current [home page of documentation](#) is not that appealing, hence I prepared a rough design for the same as an example. [Link here](#). Please note that this is just an example. Such UI design improvements can be added as per requirement.

Agile Model

Efficient project management plays a vital role in the success of a project. **Agile software development** methodology is one of the simplest and effective project management methodology. I am planning to use agile

methodology for the entire project. I believe this will help the community and mentors to monitor the project more effectively, easily and efficiently. The whole timeline will be divided into sprints and each sprint will have planned user stories which will contain in depth details of work which are going to be implemented in that week. Review and audit will be conducted at the end of every sprint. Each phase will be divided into two sprints. More on Agile methodology can be found [here](#).

Tentative Timeline

Phase 1 (Till May 17th)

- Become more familiar with the current documentation
- Learn new techniques and professional writing skills that will come in handy while implementing the project
- Participate in more meetings with the community

Phase 2 (May 18th to June 15th)

Sprint 1 (May 18th to May 31st) - Community Bonding and Plan Finalization

- Familiarizing myself with NumPy by working out small coding projects using NumPy.
- Audit the current documentation and refine my objectives
- Setup a communication channel with mentors and community resolving timezone issues.
- Add more details , which are missed in this proposal.
- Prepare the upcoming sprint user stories after discussing with mentors
- Finalize modifications which are to be implemented and get it accepted

Sprint 2 (June 1 to June 15)

- The first task I am planning to start with is the Tutorials
Currently the tutorial for beginners which is available in the NumPy documentation is a comprehensive one and helps beginners a lot. The

first tutorial document I am proposing is *How to write a tutorial document*.

- A separate plan and structure (which I will submit before sprint start) is needed for this task alone as it lays the foundation for upcoming tutorials, hence this one should be written first.
- The document will be prepared , and submitted and will be reviewed by the mentors well before sprint closing.

Phase 3 (June 17th to July 15th)

Sprint 3 (June 17 to June 30)

- As the tutorial guidelines are well organised , I would like to add new tutorials and improve existing ones, currently the NumPy tutorial section contains only two tutorials.
- One example of Improvement is documenting the minimum hardware requirement in the installation section[[Git Tracker Issue](#)]. Detailed sprint plans with all the planned documents will be prepared at sprint start after discussion with mentors.
- Adding more tutorial documentation to the NumPy tutorial section can be done after discussion as per requirement. Improving the existing tutorial with more visualization can be done
- First level restructuring , the tutorials will be done
- Similar to the previous sprint the document will be prepared , and submitted and will be reviewed by the mentors well before sprint closing.

Sprint 4 (July 1 to July 15)

- This sprint will start with preparing How-to guides. Adding How-tos for sections discussed in the NEP44 which are
- Parallelization (controlling BLAS multithreading with threadpoolctl, using multiprocessing, random number generation, etc.)
- Storing and loading data (.npy/.npz format, text formats, Zarr, HDF5, Bloscpack, etc.)

- A detailed report with finalized content will be prepared at the sprint start. The How-to guides will be prepared with minimum ambiguity and will follow the format mentioned in the [How to write a NumPy how-to](#) section.

Phase 4 (July 17th to August 15th)

Sprint 5 (July 17 to July 30)

- This sprint will continue adding other sections mentioned in the NumPy enhancement proposal which are
 - Performance (memory layout, profiling, use with Numba, Cython, or Pythran)
 - Writing generic code that works with NumPy, Dask, CuPy, pydata/sparse, etc.
- Apart from this other How-to guides can be included as per mentor and community suggestions.

Sprint 6 (August 1 to August 15)

- In this sprint I will be adding dedicated Explanations section with explanations on topics
 - Copies vs. Views
 - BLAS and other linear algebra libraries;
 - Fancy indexing.Which are discussed in the NumPy enhancement proposal (NEP44)
- Along with this explanations from the reference guide will be moved to the newly added Explanations section. This will be finalized at the print start after discussion with mentors.

Phase 5 (August 17th to September 15th)

Sprint 7 (August 17 to August 31)

- This sprint will cover Restructuring and reorganizing the documentation page after adding all the new aspects. The newly organised structure will be as per the NumPy enhancement Proposal ([NEP44](#))

Proposed plan for Restructured documentation

- **For users:**
 - Absolute Beginners Tutorial
 - main Tutorials section
 - How Tos for common tasks with NumPy
 - Reference Guide (API Reference)
 - Explanations
 - F2Py Guide
 - Glossary
- For developers/contributors:
 - Contributor's Guide
 - Under-the-hood docs
 - Building and extending the documentation
 - Benchmarking
 - NumPy Enhancement Proposals
- Meta information
 - Reporting bugs
 - Release Notes
 - About NumPy
 - License

Sprint 8 (September 1 to September 15)

- I am planning to dedicate this sprint entirely to fix as many issues as possible from the [NumPy Documentation Issue tracker](#) in Github.

- While some of them will get automatically addressed after the new restructuring , others which remain will be specifically addressed.
- I will try to include questions which are asked in other platforms such as Stack Overflow.
- All the additions will be finalized after further discussions with mentors and community

Phase 6 (September 17th to October 14th)

- This whole phase will be dedicated to proofreading the documentation for typos and errors.
- Adding more UI related designs which could help the documentation
- Ensuring the quality of the documentation
- Completing any tasks which are backlogs of any sprint
- Preparing a final case study and report for final evaluations
- Writing a report of my GSoD experience with NumPy which will be awesome! And updating the same in my personal blog.

About Me

Name : Mahesh S Nair

Email : mahesh6947foss@gmail.com

LinkedIn : <https://www.linkedin.com/in/mahesh-s-nair-8a9aab15>

Time Zone : UTC + 5:30 , India

Location : Kerala, India

Typical Working Hours : 6PM to 11PM (UTC + 5:30) Indian Standard Time

Resume : [My Resume](#)

My Experience

- Won **Google Summer of Code 2018**, with KDE - [Link](#)
- Worked as Mentor for Google CodeIn 2018 with KDE
- Working as Quality Assurance Specialist at Cognizant Technologies since January 2020
- Published documentation article with WordPress - [Link](#)
- Documented code as part of Google Summer of Code
- Delivered Technical talk at International conference , Akademy 2018 at Vienna , Austria which is the annual meet of KDE.
- Was Invited to attend Akademy 2019 at Milan Italy
- Personal blogging

Why am I the right person for the project?

I believe myself to be the right person for this project because

- First and foremost I have past experience with Open-Source since 2017, Which is roughly four years. I have completed the Google Summer of Code program which shares a lot with Google Season of Docs , which makes me more familiar with the program. Also it helps me in understanding the technical aspects better and faster.
- My past experience in writing and publishing documentation articles with Wordpress , and more than one year of work experience at Cognizant , helps me to develop a good tone and style of writing. This also helps me understand project management and I have hands-on experience in working under a project.
- My four year of experience in Open-source shows my enthusiasm towards contributing and learning new things. I have seen , used and learned tutorials and documentations of various applications , which will

definitely help me build the best version of documentation for NumPy. I sincerely wish to be a part of the community even after the program period. I am certain that , my visual design skills will also benefit in presenting the updated documentation in a better way.

References

[NumPy Enhancement Proposal](#)

[Diátaxis Framework](#)

[NumPy GitHub issue tracker](#)

[HIGH-LEVEL RESTRUCTURING AND END-USER FOCUS - NumPy](#)

