

How to contribute to NumPy - Text

Page 1

"At the University of Monty"

Alice: "Whew, I'm so happy we're done with the lab class.",
(Alice is relieved after completing a lab session.)

Bob: "Buuuut we're not done yet."
(Bob reminds Alice that there's more work to do.)

Alice: "We need to look at the cell image data, right? With NumPy?",
(Alice references the need to analyze cell image data using NumPy, indicating they are working on some scientific data processing.)

Alice: "Yes... and I have no idea how to use NumPy. So thanks for meeting after class to help me",
(Alice admits they don't know how to use NumPy and expresses gratitude for the upcoming help.)

Chris: "No problem! We're happy to help. I've played around with NumPy before in a computer science class.",
(Chris is confident and reassures Alice, mentioning prior experience with NumPy in a different context (computer science).)

Page 2

Alice: "You remember what we did in the lab, right?",

Bob: "Yup, we finished staining the cells.",

Alice: "So now all the cell image information is bundled up into a zipped file. So we use NumPy to extract the file.",
(Alice explains that cell images are compressed into a file, and NumPy is used to handle the extraction.)
[Staining, Scanning, Images, Zipping, npz file, Unzipping]

Bob: "Arrays? How are images actually a bunch of numbers?",
(Bob is confused about how images can be represented as arrays of numbers.)

Chris: "Can't we just open the image... as an image? See it with our eyes?",
(Chris asks why they cannot directly view the image instead of dealing with arrays.)

Alice: "Well, in microbiology research we're sifting through lots and lots of images of cells. And with NumPy's arrays (ndarray) we can process and analyze much faster than manually going through images one by one."

(Alice clarifies that in research, many images need to be processed, and using NumPy's ndarrays speeds up the analysis process.)

Bob: "I can work with 3 images.",

Chris: "But 400 images? From just one lab session?",

Bob: "Hmph. That makes sense, I guess.",

Page 3

Chris: "Ok! So how do we get started?",

Alice: "Let's start by making a Python file.",

Alice: "In the file, I'll start with importing NumPy.",

(Alice is explaining the initial step of starting a Python script by importing the NumPy library.)

Alice: "Our professor gave us our cell images as an .npz file named cell_images.npz.",

(Alice explains that they received the cell images in a compressed NumPy archive (.npz format).)

Alice: "Let's combine the function np.load() with cell_images.npz.",

(Alice introduces the use of NumPy's load function to open the .npz file.)

Alice: "And this is our opened .npz file.",

(Alice points out the contents of the file after using np.load().)

Bob: "Great! But... this is a lot of files. How are we supposed to know which one is ours?"

(Bob is confused by the multiple data entries in the loaded file and wonders how to identify the correct one.)

Chris: "You're right. I want my cell images from the 10AM lab."

Bob: "'npzfile A' doesn't tell me if these are from the 10AM lab or 3PM lab.",

(Bob is frustrated because the file naming doesn't provide clarity on which lab session the data belongs to.)

Alice: "Since this looks like a dictionary, let's use the function keys().",

(Alice proposes using the dictionary-like structure of the file to inspect its keys and identify the right data.)

Chris: "Dictionary? Keys?",

(Chris is unsure about how dictionaries work in this context.)

Bob: "I see. Since I'm seeing keysview, this is indeed a dictionary-like object.",

(Bob confirms that the loaded .npz file behaves like a dictionary after inspecting the keys.)

Alice: "The attributes of 09-09-10AM object...."

Bob: "Let's change our approach. Let's use the function list() to turn all this into a proper list.",

(Bob suggests converting the keys into a list to make the data easier to handle.)

Bob: "Now we transformed our npz files into a list of objects! The list of objects is a list of our lab sessions."

Bob: "Each lab session is an object as well. I want to see 09-09-10AM object's images"

Alice: "Let's see all the filenames of the 09-09-10AM objects... There! We got now the cell images captured during our 10AM lab!",

(Alice finds the relevant data after listing the keys and identifies the correct images.)

Page 4

Chris: "WHAT—THE—?!"

Bob: "Why can't you see the filename right there?"

(Bob is frustrated that the filenames aren't visible.)

Chris: "You said NumPy is supposed to be faster! Better than manual!"

(Chris is upset that NumPy is not making the task easier.)

Alice: "I mean, I expected that keysview would tell me the filename. But in reality, it doesn't."

(Alice is explaining that the keysview didn't give the expected result.)

Bob: "So you need to work around its wonkiness!"

(Bob urges Alice to find a way around the issue.)

Alice: "I can work around it since I know how to inspect objects..."

(Alice assures that she knows how to manage despite the problem.)

Chris: "But hey, it's alright as long as it works."

(Chris tries to downplay the problem, focusing on functionality.)

Bob: "I don't think we should accept 'as long as it works!' It makes our lives harder!"
(Bob pushes back, stating that they shouldn't settle for suboptimal solutions.)

Chris: "You're right. It is hard, but we can change this!"
(Chris agrees with Bob's perspective and expresses hope for improvement.)

Page 5

Alice: "NumPy is open source software. Anyone can contribute and change the codebase. Even as grad students, we can make suggestions, such as making it easier to open .NPZ files. And then other people can build off that suggestion and make that feature a reality!"
(Alice explains that NumPy's open-source nature allows users to improve its functionality.)

Chris: "In that way, we can contribute to open science!"
(Chris highlights the connection between contributing to NumPy and advancing open science.)

Bob: "Buuuuut! What are we actually supposed to do?"
(Bob is unsure about how they can start contributing to NumPy.)

Alice: "Since NumPy is a tool used in microbiology, we are building the tools that make scientific research more collaborative and transparent! Isn't that amazing?"
(Alice emphasizes that improving NumPy can have a positive impact on scientific research, particularly in fields like microbiology.)

Chris: "Umm... I think we're supposed to make an issue on GitHub.com, saying we have a problem, and suggest a solution..."
(Chris recalls that making contributions to NumPy starts with creating issues on GitHub to address problems.)

Bob: "OK, so I never contributed or anything yet, but I learnt a lot at that open science lecture last semester!"
(Bob reflects on their knowledge of open science but admits to not having contributed yet.)

Chris: "I have a GitHub account. We can try it with mine!"
(Chris offers to use their GitHub account to get started.)

Alice: "Great! Let's look at NumPy.org to find their GitHub page."
(Alice suggests they visit NumPy's website to locate their GitHub repository.)

Page 6

Alice: "At NumPy.org."
(Alice is navigating to the NumPy website.)

Chris: "Let's click over here!"

(Chris points out the section on the website to explore further.)

Alice: "Well, I'm looking at NumPy.org and umm..."

(Alice seems unsure of where to go next on the site.)

Chris: "No, I'm pretty sure we can start with GitHub. Let's click here."

(Chris directs Alice to visit NumPy's GitHub repository to begin contributing.)

Bob: "This looks complicated. A mailing list, Slack, community calls? Too much!"

(Bob feels overwhelmed by the community options available on the NumPy website.)

Alice: "There we go! Now we're on NumPy's GitHub page, with a list of issues."

(Alice successfully reaches NumPy's GitHub page where issues are listed.)

Chris: "Let's create a new issue here. What we want is a feature request."

(Chris explains that they need to create a new issue to request a feature.)

Alice: "Let's write an issue."

(Alice prepares to draft their feature request.)

Page 7

Alice: (Writes) "Show filenames in npz archive?"

(Alice creates a feature request regarding the difficulty in identifying filenames within .npz files.)

Chris: "Hi, we're microbiology students using NumPy to look at our cell images."

(Chris explains the context of their use case with .npz files.)

Bob: "I don't know how to use keys."

(Bob expresses confusion regarding the dictionary-style keys in NumPy objects.)

Chris: "Oh yeah! Let me add that part..."

(Chris offers to help by clarifying how keys work in NumPy objects.)

Page 8

Bob: "Now tell them to solve it."

(Bob is eager for a solution and wants immediate action.)

Chris: "I don't think we should be so demanding!"

(Chris suggests a more diplomatic approach to their feature request.)

Alice: (Writes) "Could the filenames be easier to see? Like right when you open an npz?"
(Alice proposes a potential solution for improving .npz file interaction.)

Bob: "Woo! Submitted! Into the internet!"
(Bob celebrates after their feature request is successfully submitted.)

Page 9

Chris: "After all that work, I need a break."

Bob: "B-but we didn't finish our assignment!"

Chris: "We can still use NumPy, it's just... inconvenient."

Chris: "Well, I have GitHub notifications on. So don't worry, we'll hear back."

[Time passes, both are relaxing.]

Bob: "Oh, someone responded!!!"

[They excitedly rush to check the response.]

Chris: "Aghh, fine! Let's take a look."

Page 10

[Looking at the github issue made for showing filenames in npz archive]

Alice : "So this is the issue we made ..."

Alice : "Wow, a maintainer commented, and someone else ?"

Bob : "Who are these people ?"

Bob : "What are they saying ?"

Bob : "Do we need to do something now ?!"

[Alice explain who a maintainer and a contributor is, the picture explains about how the discussion and contribution happens]

Alice : "The first person is a maintainer"

Alice : "The maintainer leads numpy by coordinating with many people towards the big picture"

Alice : "The second person is a contributor"

Alice : "They want to help! To solve our issue by making a pull request !"

Alice : "Together, maintainers and contributors implement new feature, fix bugs and write documentation to make numpy better"

Page 11

Bob : "The contributor made a pull request ! What's that ?"

Alice : "A pull request is asking to make changes to the numpy codebase. This one is changing the npz files."

[There is a pull request comment by a contributor saying they can help, providing a solution on how it can be fixed, raising a pull request]

Alice : "We can see changes the contributor made to the codebase."

Bob : "To be honest, I don't fully understand what's being changed ..."

[Alice points to the summary of the pull request raised by the contributor stating what code changes are being done]

Alice : "That's alright, here's a summary of what has been changed."

[Bob taking a glance over the pull request]

Bob : "I wonder what the contributor is doing ..."

Bob : "They must know everything about numpy to make a pull request like that ..."

Page 12

[Following part shows the view of the contributor thinking]

"Meanwhile, far far away, the creator of that pull request thinks ..."

Contributor : "Ah, I don't know anything about numpy ?"

Contributor : "Wait - Don't beat yourself up. You do know numpy."

[Contributor looking at their pull request getting reviewed successfully]

Contributor : "I'm happy that a maintainer reviewed my work. But how do I respond ?"

Contributor : "The review said I missed out on X. Which makes sense."

Contributor : "Most of my pull request already works. So I'm 80% done."

Contributor : "I just need to rewrite some parts and finish the last 20%!"

[A contributor introspects a little on their contribution and checks the review got from the maintainer. The contributor realises that there is still 20% work yet to be completed.]

Page 13

[Somewhere at the lab, contributor who is thinking about facing npz file issue in the past]

Contributor : "These students... I feel a kinship. It's been years since I was in a lab."

Contributor : "But I remember working on the same npz filetype problem"

Contributor : "Why didn't I submit my own issue ? Saying I needed this help too ?

Contributor : "I ended up making my own work-arounds."

Contributor : "If I had submitted my own workarounds as a pull request all those years ago ..."

Contributor : "Maybe I could have helped others facing the same problem?"

Contributor : "Eh! I was probably too focussed on finishing my thesis to submit a pull request."

Contributor : "I'm glad I'm able to give back now."

Contributor : "I wonder how the maintainer is doing"

Contributor : "He must be knowing everything about NumPy... way more than me."

[The contributor thinks about the past, when there was a problem the contributor was facing with the npz filetype and thinking an issue could have been raised earlier.]

Page 14

"Meanwhile, far far away, the numpy maintainer thinks..."

Maintainer : [Sipping coffee] "I can see the PR maker updated their PR based on my commnts."

Maintainer : "I'll review the code again, and check if it's ready to be merged."

[The maintainer takes a look at the reviewed changes and merges the pull request]

Maintainer : "I'm glad I have these students telling me the problem they're facing."

Maintainer : "I haven't been a professor in lab for a year now, so I'm not always aware of these problems."

Maintainer : "I'm glad that future students will be able to find it easier to navigate with npz file."

["The maintainer reviews the pull request and feels grateful about the students raising the issue about the problem they're facing"]

Page 15

"2 months later"

Alice : "Check out this email!"

[All 3 of them checking the email they have recieved as they signed up for numpy mailing list]

"It's from the numpy mailing list. I subscribed to keep track of our issue."

Alice : "Now our requested feature is here, in this release cycle."

[A glance of the mail showing features in the release]

Chris : "That sounds nice... does it actually work?"

[They now try out the sample code to load npz file to test if the new feature has been updated or not, and it workssssss !!!!!!!!]

[All shockingly excited] : "Whaaa! This is way easier to see if the cell images came from the 10AM or 2PM lab."

Page 16

Chris : "Woooah. I get it now."

Chris : "It's so cool to know that I made a difference to numpy!"

[Chris telling the story about how these changes started off with the single issue]

[Student write an issue, Contributor writes a pull request, maintainer reviews pull request, Pull request merged and code added to numpy codebase, Updated version of the numpy code released across the world]

Chris : "It started with one person and grew from there! The npz file issue and numpy was built by many people, all collaborating and building off each other."

[Excitement continues]

"Now I want to join the next numpy community call"

"I want to understand the numpy codebase too"

"I want to make my own pull request too"

"Meet these contributors who helped us"

"Pull it apart, ask questions"

"Wow, they changed from being the biggest skeptic to biggest fan of numpy!"

"Go for it!"

"The End"