The SciPy Documentation Project

Joseph Harrington

University of Central Florida

...with lots and LOTS of help

from all of you!

Examples

User Docs

API

Outline

- Motivation
- Requirements and Design
- Timeline and Plans
- People
- Results
- Challenge
- Future

Once upon a time...

- Taught data analysis in IDL at Cornell
- Challenging class but doable
- Programming all the time, BUT not focus
- About half of students could program at start
- Course highly successful:
  - Best students went on to teach their own advisors
  - Hired worst student, wrote spectrum extraction algorithm and test suite to prove best in world

Brought Class to UCF

- Changed course to NumPy
- Better students were similar! All were capable.
- Given *Guide to NumPy*, web site, mailing lists, me, sources, astro tutorial, encouraged to work together
- BUT:
  - *Often* spent 2 hours to find and learn a simple routine
  - Students (and I) put heroic effort into class
  - All bombed, even experienced programmers
  - Docs not even poor; mostly nonexistent
- I teach it again this fall...what to do?
- Must either have docs or go back to IDL (Nooooooo...)

If I can't find a reindeer, I'll *make* one instead.
-T. Grinch, 1957

- Except I'm a pre-tenure prof
- Research very active, time very constrained
- Saw what happened to Travis, expected same
- Resources:
  - Community
  - Some surplus funding
  - Experience and skill as a communicator
  - About 8 months

Requirements - General

- Reference pages and manual
- Tutorial user guide
- Quick "Getting Started" guide
- Topic-specific user guides
- Examples that work
- "Live" collection of worked examples (cookbook)
- Tools to view and search docs interactively
- Variety of access methods (help(), web, PDF, etc.)
- Mechanisms for community input and management
- NumPy, then SciPy, then others

Requirements - Reference

- Reference page per function/class ("docstrings")
  - Thorough, complete, professional text
  - Accessible one level below expected user
  - Examples, crossrefs, math, images (limited)
- Pages for general topics
  - Glossary
  - Slicing, broadcasting, indexing
- Reference manual
  - Organized by topic areas
  - Index

General Timeline

- Infrastructure
  - Doc wiki (and SVN link, stats, reviews, etc.)
  - Templates and standards
  - Indexing, math, image mechanisms in ReST
  - Produce text, PDF, HTML automatically
- NumPy
  - Reference (functions/objects, topic pages)
  - Getting Started Guide
  - Tutorial
- Reading tools beyond help(), browsers, PDF readers
- SciPy (yikes)

First Steps and Later Questions

- NumPy is big – page triage, UCF course funcs first!
- 2008: by...
  - Fall semester: NumPyI 50%+ to "Needs review"
- 2009: by...
  - Spring semester: NumPyI 100% to "Needs review"
  - Summer: NumPyI Proofed
  - Fall release: SciPy 25%+ to "Needs review", etc.
- When/if/who/how to do "uninteresting" NumPy?
- Who/how to do user guides
  - Community or individual?
  - Free guide + encourage many commercial books!

Organization – Core Team

- Leerless Feeder: jh – organization, plans, standards and processes, funding, whip, person who will die if this fails...
- Reliable and available expert (on UCF contract): Stéfan van der Walt – interface with SVN, PDF and HTML, standards and processes, wiki, chief writer, the one person I could dump on mercilessly...
- Emmanuelle Guillart – wiki, wiki hosting
- Pauli Virtanen – wiki, writing
- Perry Greenfield – numpy.doc pages

Shirts

- Write 1000 words or do equivalent work
  - Wiki
  - Reviewing
  - Proofing
- Now or later (we'll mail)
- Offer good until we withdraw it
- Don't be an "expert"
  - Tell us the size you want
  - Not the list of all possible sizes you could wear!

SCIPY DOCUMENTATION MARATHON 2008
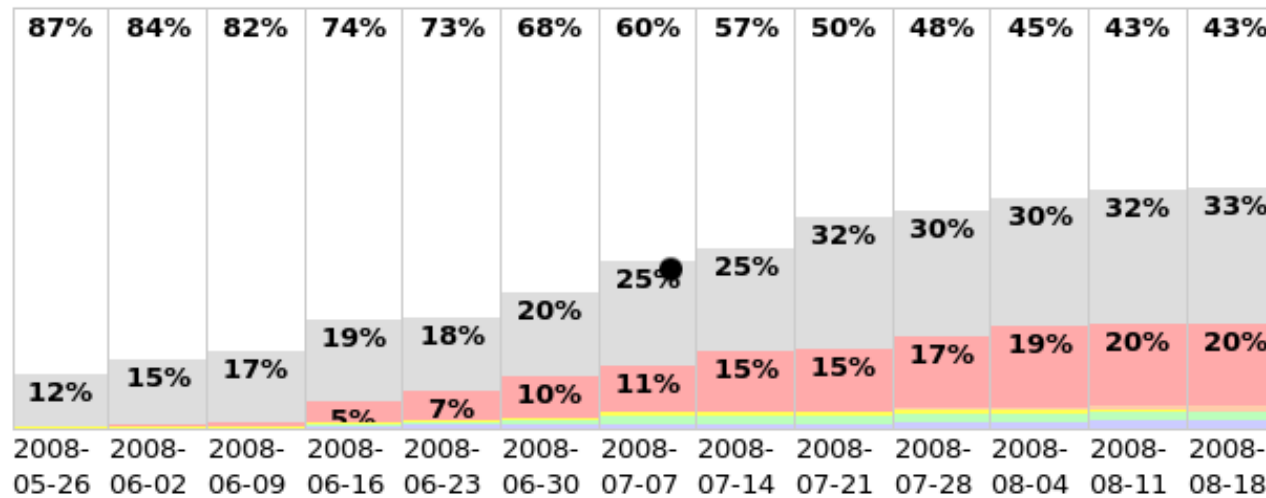
Examples
User Docs
API

I WTFM
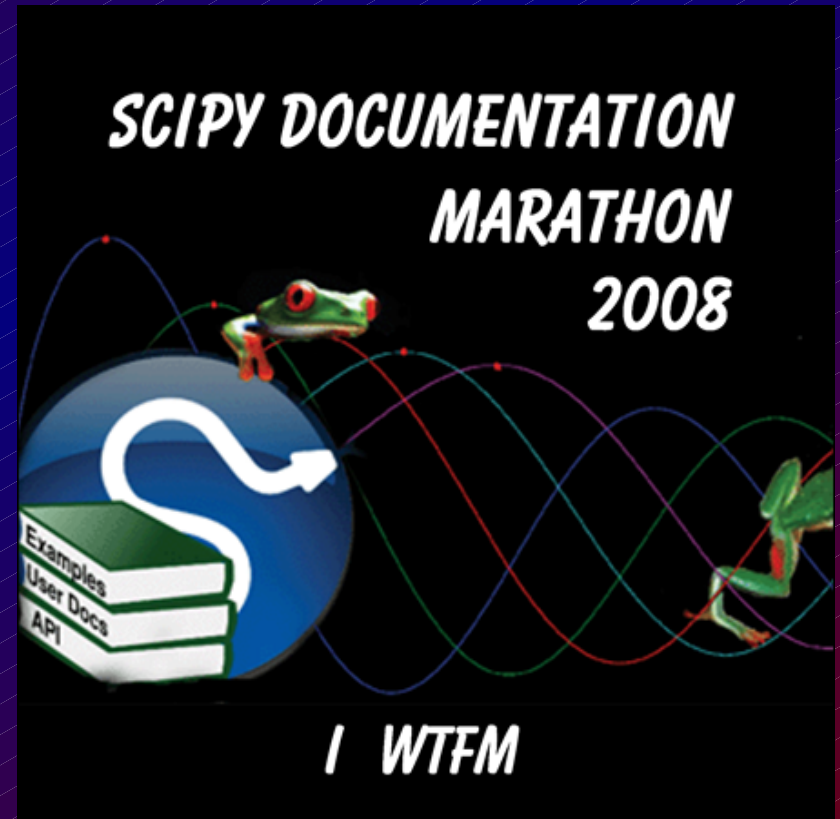
36 Community Heroes and counting

?	Bjørn Ådlandsvik	!	David Huard
	René Bastian	!	Alan Jackson
	Nathan Bell		Teresa Jeffcott
	Joshua Bloom		Samuel John
	Patrick Bouffard	!	Robert Kern
	Matthew Brett		Roban Kramer
	Christopher Burns		Vincent Noel
!	Tim Cera		Travis Oliphant
	Johann Cohen-Tanugi	?	Scott Sinclair
?	Neil Crighton	?	Bruce Southey
	Arnar Flatberg		Andrew Straw
	Pierre Gerard-Marquardt	!	Janet Swisher
!	Ralf Gommers		Theodore Test
!	Keith Goodman		James Turner
!	Perry Greenfield	!	Gael Varoquaux
!	Emmanuelle Guillart	!	Pauli Virtanen
!	Joe Harrington		Nicky van Foreest
	Robert Hetland	!	Stéfan van der Walt
			...who owes me ice cream!

!= Earned a T-shirt! Hooray and thanks!

?= May have earned a T-shirt, talk to Stefan

Results – Summer 2008

Weekly

87%	84%	82%	74%	73%	68%	60%	57%	50%	48%	45%	43%	43%
										30%	32%	33%
								32%	30%			
						25%	25%					
			19%	18%	20%							
12%	15%	17%				11%	15%	15%	17%	19%	20%	20%
			5%	7%	10%							

| 2008-05-26 | 2008-06-02 | 2008-06-09 | 2008-06-16 | 2008-06-23 | 2008-06-30 | 2008-07-07 | 2008-07-14 | 2008-07-21 | 2008-07-28 | 2008-08-04 | 2008-08-11 | 2008-08-18 |

Current

	%	Count
Needs editing	43	358
Being written / Changed	33	273
Needs review	20	164
Needs review (revised)	1	10
Needs work (reviewed)	0	2
Reviewed (needs proof)	2	15
Proofed	2	18
Unimportant	–	1531

- Quality is high! Lots of examples, and they work.
- PDF doc has 309 printed pages
- Doc wiki is your best source for docstrings!
  - scipy.org ➔ Documentation ➔ follow link

The Doc Challenge

- "Normal" professional (free or not) software writes docs in parallel with code development
- GOOD software projects write docs FIRST and use them as specs!
- NumPy had to start fast to unite Numeric and numarray.
- That era is now over (Thank Goodness!)
- CHALLENGE: No new functions or classes in NumPy or SciPy without reviewed docstrings.
- URGING: Lack of a docstring is a major bug. If you patch a function, *please* write a docstring!

Now and Future

- Go to doc wiki now, make an ID, post to scipy-dev!
- More writers!
  - More experts, especially in "uninteresting" pages
  - More code authors may be needed for SciPy
- Review
  - Separate technical and accessibility reviews?
  - Should reviewers be a small, trained group?
- Different approach for user guides?
- We need you. Yes, YOU, with the laptop! Sign up now!