U5VGKQ2SY : Not enough flour or corn for that.
U1BP42MRS : Check the intro doc :slightly_smiling_face:
U1BP42MRS : Also the topic
U3G7RJP61 : taco bot is pretty annoying when it comes to DM's though :joy:
U3G7RJP61 : I've turned off DM notifications about receiving / handing out tacos each week about 6 times now and it *still* does it
U0LSCQQNR : so, got a question about python string casing. I have a large number (25K) PDF documents indexed in SOLR via `pdfminer`. Problem is, text casing is irregular, to say the least
U0LSCQQNR : is there a library I can use to apply grammar rules in regard to text casing?
U5VGKQ2SY : Just tread it :joy: great sense of humor. Thanks for the link.
U5VGKQ2SY : <@U1BP42MRS> :taco:
U5LNXQHN3 : What would be the intended end result?
U0LSCQQNR : ```PrinCiPleS oF bioloGyi

PrinCiPleS oF bioloGy ii
```

to

```Principles of Biology I
Principles of Biology II```

U5LNXQHN3 : Okay, so you're seeking some sort of normalised form
U0LSCQQNR : correct
U0LSCQQNR : based on regular english rules governing capitalization
U0LSCQQNR : I can find things for spelling correction
U5LNXQHN3 : You're potentially looking at PhD level amounts of NLP here to do it properly, but maybe you don't need _properly_
U0LSCQQNR : but not with capitalization
U0LSCQQNR : yeah, agreed
U0LSCQQNR : I mean, I could lowercase them all
U5LNXQHN3 : The example you gave is a good edge case... because capitalisation works differently if the context is a heading
U0LSCQQNR : correct
U0LSCQQNR : all this is text extracted from a PDF
U0LSCQQNR : so, source is pretty questionable, in a structured format perspective
U1BP42MRS : I'd probably just lowercase them to make it simple
U5LNXQHN3 : What is the purpose of changing the case? Is it for display to end users, or for facilitating search, or...?
U0LSCQQNR : display to end user
U60JLTH0W : <https://pypi.python.org/pypi/gingerit>
U0LSCQQNR : basically, enter in a search query, and list of results shows
U60JLTH0W : <https://pypi.python.org/pypi/language-check>
U0LSCQQNR : with click handlers for selection
U60JLTH0W : <https://pypi.python.org/pypi/grammar-check/1.3.1>
U60JLTH0W : <@U0LSCQQNR> check those 3 modules
U60JLTH0W : Haven't tried any of them actually but would love to hear your feedback :slightly_smiling_face:
U1BP42MRS : I would presume you'd index on `lower(thing)` and search that way, but find in the document fuzzily based on the string for display (granted not sure if solr supports that)
U0LSCQQNR : well, solr search is case insensitive by default, I believe
U0LSCQQNR : <@U60JLTH0W> thanks for the tips!
U1BP42MRS : Oh I see, it's really just what is contained in the doc isn't necessarily correct. :whoosh:
U5LNXQHN3 : Personally I'd just do `[sentence.lower().capitalize() for sentence in sentences]` and wait to see if that was sufficient. :slightly_smiling_face:
U5LNXQHN3 : Finding where one sentence ends and another starts, in badly scraped data... well, that's another problem
U0LSCQQNR : oh agreed
U0LSCQQNR : text is tricky
U60JLTH0W : <@U5LNXQHN3> may be a script can be written to search for `\n` as well as `.` and `,` in the data to detect the start of new sentences?
U5LNXQHN3 : It's better than nothing

U60JLTH0W : Hope any of them would be usefull :smiley:

U5K8B5ETE : Hello guys, has anyone tried Dan Bader's Python Tricks The Book? I wanted to know if it is really good as it is claimed on the author's page

U0NRYQNAZ : i've heard his stuff is good but i have no experience with it personally.

U4BMZ90T0 : You could always look for ends of sentences, capitalize next alpha character. Search against databases for pronouns and such

U5JF1KD18 : Hello guys I got a huge problem

U4BMZ90T0 : <@U5JF1KD18> ask away

U5JF1KD18 : We've been ask as assignment the implementation of Dijkstra's and Bellman Ford's algorithm for calculating the shortest path in a given graph

U4BMZ90T0 : So what's the issue?

U4BMZ90T0 : run into a problem?

U5JF1KD18 : I don't really know how to start and that's my problem

U4BMZ90T0 : Well were you given code to start off with as far as "i will give you this input" in order to build the graph you have with the path weights and what not?

U5JF1KD18 : The Input to the algorithms is an oriented graph with weights

U4BMZ90T0 : So then you have to build the graph in the code, yes?

U5JF1KD18 : More precisely compute the shortest path from a given source to all the other nodes

U4BMZ90T0 : I understand that, I am just asking if you have anything to build off of like code given to you to complete the task or if you have to start from scratch and build the graph in code then calculate it.

U5JF1KD18 : You need to code the algorithm from scratch and supply it with a graph in a text file and it'll calculate the shortest path in it

U4BMZ90T0 : can you give an example input

U0LSCQQNR : <@U5JF1KD18> how much experience do you have writing code?

U0LSCQQNR : for example, there are quite a number of existing examples of the algorithms you're talking about

U5JF1KD18 : basic i'm just starting

U0LSCQQNR : ok

U0LSCQQNR : can you describe the steps on how you execute the algorithm?

U0LSCQQNR : and do you understand why those steps are necessary?

U0LSCQQNR : if so, then the next step you take is translating your written description of the process into pseudocode

U0LSCQQNR : once you have a reasonable sequence of actions, you then implement the pseudocode in your language of choice

U0LSCQQNR : frankly, the first two items are always the most difficult

U0LSCQQNR : because it requires you to understand the problem domain

U0LSCQQNR : once you understand it, making it work is usually much less effort

U1NSCAY6R : oof graph theory for a beginner. do you understand how those algorithms work, <@U5JF1KD18> ?

U5JF1KD18 : Yes I understand how those work

U0LSCQQNR : just having trouble translating described steps to code?

U2BS4M1RV : I saw this today, <https://github.com/stackimpact/stackimpact-python>. Looks nice, but don't think we could use that internally. Anyone have any experience with Stack Impact, or know of something that would provide similar data without sending it to an outside source? Even just having the data print out at the end of a run would be great.

U1UFZTD4J : <@U2BS4M1RV> I dont know of any tool that does all of that. But I know of many tools that do peices

U1UFZTD4J : What data are you mainly interested in?

U2BS4M1RV : The charts are pretty, but I would just like to see which calls are blocking and for how long.

U2BS4M1RV : For most of my code it's pretty clear anyway, and I throw in timing decorators at the really long ones.

U1UFZTD4J : so, if you just want CPU profiling, then vprof works really well.

U2BS4M1RV : Having a monitoring system for users who use my scripts would be nice too though. Right now I just have them email the logs if somethign goes wrong.

U1UFZTD4J : So, for "errors", I highly recommend sentry

U1UFZTD4J : basically, anyone not using sentry, should be doing it right now

U1UFZTD4J : its a game changer

U2BS4M1RV : Having a syslog server for loggign would be amazing, however the logs get extremely large for some scripts.

U2BS4M1RV : One of them produces a log so large the last step in the script is to zip and delete its own logs.

U1UFZTD4J : So, what do you actually use the logs for though?

U1UFZTD4J : just trying to figure out what went wrong?

U2BS4M1RV : Determine why some data was missed or not parsed correctly, why communication failed on a device, why an exception occurred.

U2BS4M1RV : Yeah

U1UFZTD4J : I would try using sentry instead of logging
U1UFZTD4J : it might change your life
U2BS4M1RV : The worst of them runs as a cron job and downloads a few hundred mb of text from 60 routers.
U2BS4M1RV : I'll take a look, tahnks.
U1UFZTD4J : here is the page: <https://sentry.io/welcome/>
U1UFZTD4J : but its an Open source project, so you can host it yourself and dont have to use the SaaS
U5ZSB9UDP : Is anyone aware of a standard library utility related to the `ast` module that can extract the source of a function without evaluation? Or more specifically, the ability to trace the execution path to a function without evaluation?
U1UFZTD4J : what are you really looking for?
U1UFZTD4J : there are various ones for different things
U1UFZTD4J : if you just want to look at the function, etc, then `inspect` works pretty well
U1UFZTD4J : &gt; the ability to trace the execution path to a function without evaluationWhat does this mean? you want to look and see what the "stack" might look like had the function ran?

U5ZSB9UDP : <@U1UFZTD4J> Right, I'm essentially trying to get the stack without actually executing the code.
U5CRANMFV : Hello. How in Data Frame Pandas to remove the top line. So the second line became the title?
U5TNDB5V3 : can you display what your dataframe looks like and what you want it to look like?
U5TNDB5V3 : i.e. are you trying to just drop a row?
U5CRANMFV : <https://stackoverflow.com/questions/31328861/python-pandas-replacing-header-with-top-row>Thank you, <@U5TNDB5V3> help me.

U5CRANMFV : `df.to_csv("file_name.csv", sep=',', encoding='utf-8')`does why is that so
`,a,b,c,d,e,f`
and <http://imgur.com/zDdawCWl.png> this is what I get

U5TNDB5V3 : not really understanding what the issue is, is it upon export you're getting the indices?
U5ZSB9UDP : <@U5CRANMFV> try `df.to_csv("file_name.csv", sep=',', encoding='utf-8', index=False)`
U5CRANMFV : Thank you very much, guys. I'm happy.
U5TNDB5V3 : (Y)
U5CRANMFV : <@U5TNDB5V3> it <https://en.wikipedia.org/wiki/Thumb_signal> ?
U5TNDB5V3 : exactly
U5CRANMFV : <http://imgur.com/ob9JKs8l.png> - reminded
U5CRANMFV : `df = pd.read_html(URL)`I can not understand, he analyzes the page and looks for a table?
That is what I have used XPath was a mistake?

U1UFZTD4J : <@U5ZSB9UDP> thats a really really hard problem. Especially with the dynamicness of python
U5ZSB9UDP : Yeah, I think I'm just going to take a crack a dumbed down approach that just works for simple functions.
U1UFZTD4J : So, `inspect` will give you the source code, then you can keep going down that path
U5ZSB9UDP : <@U5CRANMFV> I believe `read_html` tries to find a `&lt;table&gt;` tag in the HTML, so if you want to load an HTML table into a DataFrame, then that would probably be a good usage instead of trying to extract it yourself with XPath or another XML parsing library.
U5ZSB9UDP : Thanks for your help!
U5UQKCC06 : Can anyone help me parse this string?
U5UQKCC06 : `u'["Create Content"],["Facebook", "Twitter"],["Photography"],One sweet vision'`
U5UQKCC06 : I'd like to return a list of 3 strings and a list
U5UQKCC06 : each of the strings can have more than one string, but the last item is always just one string
U5UQKCC06 : ast.literal_eval fails because the last string item doesn't have quotes
U1UFZTD4J : <@U5UQKCC06> show us what you have and we will help you solve the issues you run into
U5UQKCC06 : This is ugly
U5UQKCC06 : returning `['Create Content', 'Facebook, Twitter', 'Photography', 'One sweet visio']` atm
U60KNBMPX : <@U5UQKCC06> why don't you do a sequence of tokenization steps with "split()" ?
U1UFZTD4J : Is it always going to be that format?
U5UQKCC06 : It's always going to be &lt;list&gt;, &lt;list&gt;, &lt;list&gt;, &lt;string&gt;
U5UQKCC06 : not exactly sure what you mean when you say tokenization steps
U5UQKCC06 : <@U60KNBMPX> **
U60KNBMPX : <@U5UQKCC06> if this is a string, then you can tell that there is a "],[" motiff that would split it into the logical pieces.
U60KNBMPX : You just have to examine it, think, and then act.
U5UQKCC06 : let me try
U5UQKCC06 : returns `['Create Content', ('Facebook', 'Twitter'), 'Photography', 'One sweet vision']`

U5UQKCC06 : I guess I was hoping there was a way to do it more pythonic than this brute force route I took

U60KNBMPX : sometimes it just has to be ugly

U1BP42MRS : Can anyone see any major drawbacks of having a simple boolean 'can_write' field on a user object for a very limited scope permissions (literally for the forseeable future: read/only and read/write)? We have a different user object per subdomain (not my call) - and rather than porting a whole permissions framework over, to me this seems reasonable

U2UP91YLE : well , it's like the 'is_admin' field ,right?,isn't the best choice,but it's fine , i don't see any drawbacks "for a very limited scope permissions"

U1BP42MRS : I figure that's enough to not lock us in too much if we need more. A simple migration could get us to a more involved permissions model

U4UHWHXC0 : Hi guys. There is one site that returns response in string like: `{"ConsignorName": "first part "second part""}` which should be json. Result should be `{'ConsignorName': 'first part "second part"'}`. And I can't just `json.loads(data)` because it fails with `json.decoder.JSONDecodeError: Expecting ',' delimiter:...`.  So I somehow need to cast this string in dict. Maybe u know some easy fast pythonic way to do this? So far I came up with blunt operations with split and check for double quotes inside double quotes and u can see it's already getting boring and I'm gonna stop right here :sweat:

U5LNXQHN3 : It is actually broken JSON, so I'm surprised that it returns it. You can `strip` the { and }, then use `split` on the ":", and you should have the key and the value

U4UHWHXC0 : Ikr, I was like "how does it even work??" Someone definetaly needs to reformat their data... :sigh:

U5ZPMJA06 : Probably they're not using a json library to generate that output. I guess they're just concatenating strings and hope for the best.

U5Y9Z9FPS : they can replace unicode quotes after converting it into string

U5ZPMJA06 : These people would write `&lt;consignorname&gt;&lt;first&gt;Jan&lt;second&gt;de Vries&lt;/first&gt;&lt;/second&gt;&lt;/consignorname&gt;` if they were to output XML ;-)

U4UHWHXC0 : U know the funny thing? They return xml and inside that xml is a string that should represent json. Like wtf :laughing::sweat_smile:

U4UHWHXC0 : I'm surprised `requests` can even parse that

U5ZPMJA06 : I recently scraped an angular website which also did something like that: a response which is a HTML fragment of a page, with a DIV in it that contained JSON. And in that json data was a string field filled with serialized json! It's like a <https://en.wikipedia.org/wiki/Matryoshka_doll>

U5ZPMJA06 : Anyway, with a few calls to BeautifulSoup and json.loads() I finally got to the data that was needed.

U4UHWHXC0 : At least u had valid json ._.

U5Y9Z9FPS : <@U4UHWHXC0> ```def replace(m):  s = m.groups()[0].replace('\"', '\\"')
  return ': "{}"'.format(s)


re.sub(r': "([^,}]+)"', replace,'{"first": "aaa "bb"", "ConsignorName": "first part "second part"", "second": ""ddd""}')

```  madskilz :slightly_smiling_face:

U4UHWHXC0 : Oh my god... U must be a regex guru! :pray: Thank u so much!

U5Y9Z9FPS : urw ^^ though i'm not sure that it can process all cases %)

U5LNXQHN3 : Only tangentially Python related, but does anyone know good ways of generating code-related documents offline? I'm struggling with static blog generators like Pelican because trying to coerce Pelican and Pygments and ReStructuredText to cooperate to do what I want is twisting my mind

U1NSCAY6R : hmm, not python but have you looked at jekyll?

U5LNXQHN3 : I'm on Windows and Jekyll is a bit awkward on that platform

U1NSCAY6R : ah, okay, just thought id toss that one out there

U0NRYQNAZ : sphinx should be able take care of the pygments and ReStructuredText parts.

U5LNXQHN3 : I think my problem is that ReStructuredText is really not that good

U5LNXQHN3 : If you want more control over the formatting, for instance. Rules like "Multiple successive blank lines are equivalent to a single blank line" make it awkward to introduce whitespace for readability