

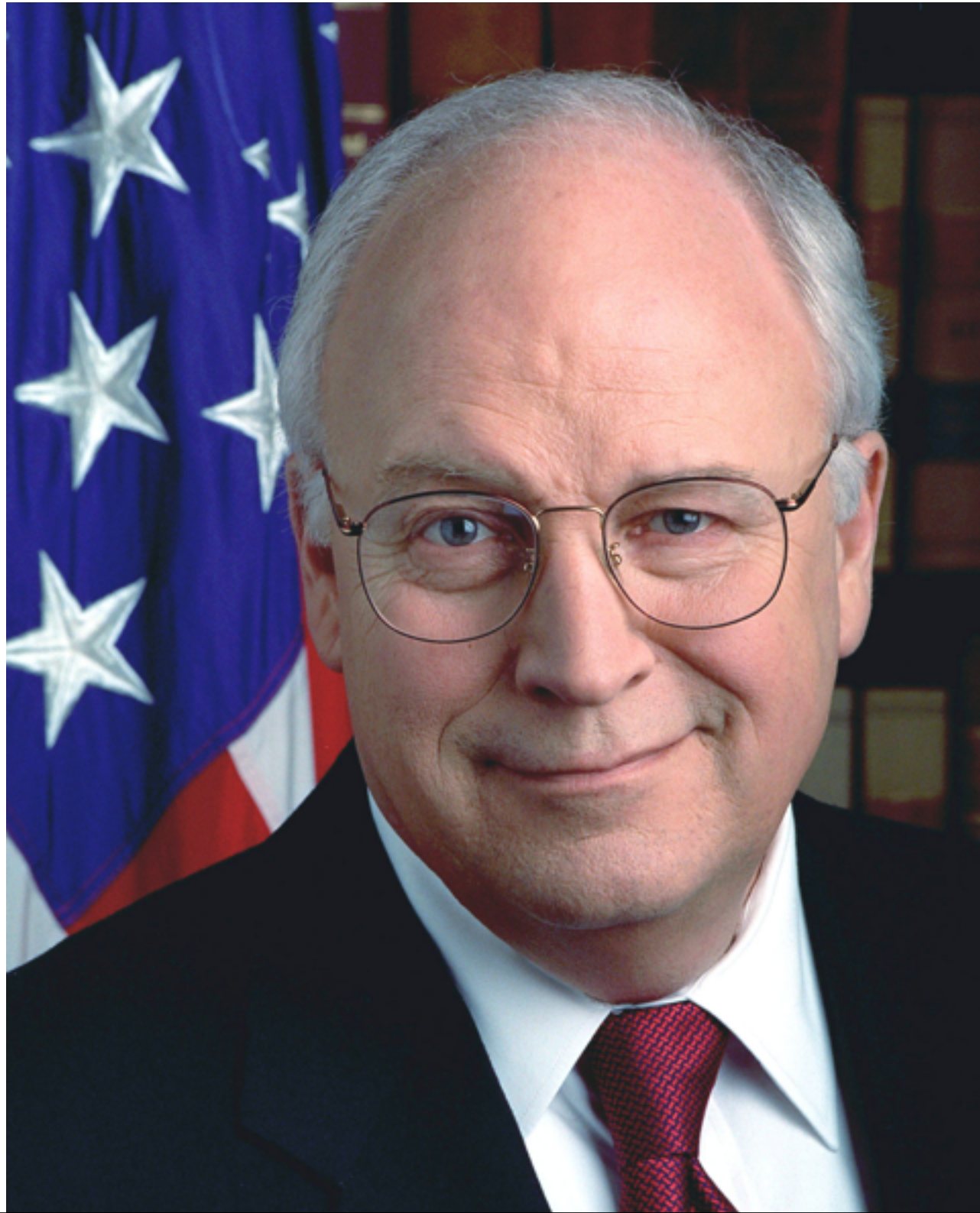
# Is Dick Cheney Dead yet? (spoiler: probably not))

---

website that uses information from Wikipedia

# Who?

---



# Set Up

---

- Use Wikipedia as a source
- Write a general python script that takes a name and checks if person is:
  - Dead
  - Alive
  - Ambiguous name / Not a Person / Not a wikipedia page
- Write a specific python script for Cheney

# Getting information from Wikipedia

---

- Get content from Wikipedia
- First (naive) Approach - get HTML and parse HTML (use Requests Module and BeautifulSoup)
- Second Approach - use API and get JSON file



# Get and Parse HTML

```
(test)🐧 nono ~ python
Python 2.7.10 (default, Jul 13 2015, 12:18:59)
[GCC 4.2.1 Compatible Apple LLVM 6.0 (clang-600.0.57)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
```

```
>>> import requests
>>> from BeautifulSoup import BeautifulSoup
>>> url = 'https://en.wikipedia.org/wiki/Dick_Cheney'
>>> r = requests.get(url)
>>> soup = BeautifulSoup(r.text)
```

```
>>> soup
<!DOCTYPE html>
<html lang="en" dir="ltr" class="client-nojs">
<head>
<meta charset="UTF-8" />
<title>Dick Cheney - Wikipedia, the free encyclopedia</title>
<meta name="generator" content="MediaWiki 1.26wmf14" />
<link rel="alternate" href="android-app://org.wikipedia/http/en.m.wikipedia.org/wiki/Dick_Cheney" />
<link rel="apple-touch-icon" href="/static/apple-touch/wikipedia.png" />
<link rel="shortcut icon" href="/static/favicon/wikipedia.ico" />
<link rel="search" type="application/opensearchdescription+xml" href="/w/opensearch_desc.php" title="Wikipedia (en)" />
<link rel="EditURI" type="application/rsd+xml" href="//en.wikipedia.org/w/api.php?action=rsd" />
<link rel="copyright" href="//creativecommons.org/licenses/by-sa/3.0/" />
<link rel="alternate" type="application/atom+xml" title="Wikipedia Atom feed" href="/w/index.php?title=Special:RecentChanges&feed=atom" />
<link rel="canonical" href="https://en.wikipedia.org/wiki/Dick_Cheney" />
<link rel="stylesheet" href="//en.wikipedia.org/w/load.php?debug=false&lang=en&modules=ext.uls.nojs%7Cext.visualEditor.viewPageTarget.noscript%7Cext.wikihihero%7Cmediawiki.legacy.commonPrint%2Cshared%7Cmediawiki.sectionAnchor%7Cmediawiki.skinning.interface%7Cmediawiki.ui.button%7Cskins.vector.styles%7Cwikibase.client.init&only=styles&skin=vector&*"/>
<meta name="ResourceLoaderDynamicStyles" content="" />
<link rel="stylesheet" href="//en.wikipedia.org/w/load.php?debug=false&lang=en&modules=site&only=styles&skin=vector&*" />
<style>a:lang(ar),a:lang(kk-arab),a:lang(mzn),a:lang(ps),a:lang(ur){text-decoration:none}
/* cache key: global:resourceloader:filter:minify-css:7:de1ab5287c9076b96eadd3f97a84a7b6 */</style>
<script src="//en.wikipedia.org/w/load.php?debug=false&lang=en&modules=startup&only=scripts&skin=vector&*"></script>
<script>if(window.mw){
mw.config.set({
"wgCanonicalNamespace":"","wgCanonicalSpecialPageName":false,"wgNamespaceNumber":0,"wgPageName":"Dick_Cheney","wgTitle":"Dick Cheney","wgCurRevisionId":671354046,"wgRevisionId":671354046,"wgArticleId":5058628,"wgIsArticle":true,"wgIsRedirect":false,"wgAction":"view","wgUserName":null,"wgUserGroups":["*"],"wgCategories":["Pages using duplicate arguments in template calls","Pages containing links to subscription-only content","All articles with dead external links","Articles with dead external links from August 2014","Wikipedia indefinitely semi-protected biographies of living people","Wikipedia articles with VIAF identifiers","Wikipedia articles with LCCN identifiers","Wikipedia articles with ISNI identifiers","Wikipedia articles with GND identifiers","Use mdy dates from April 2012","Good articles","Dick Cheney","1941 births","Living people","20th-century American writers","21st-century American writers","American anti-communists","American chief executives","American Enterprise Institute","American male writers","American memoirists","American Methodists","American people of English descent","American people of French descent","American people of Huguenot descent","American people of Welsh descent","Cheney family","Ford Administration personnel","George H. W. Bush administration cabinet members","George W. Bush Administration personnel","Halliburton","Members of the United States House of Representatives from Wyoming","Organ transplant recipients","People from Casper, Wyoming","People from Jackson, Wyoming","People from Lincoln, Nebraska","Presidential Medal of Freedom recipients","Republican Party (United States) vice presidential nominees","Republican Party Vice Presidents of the United States","United Methodists","United States congressional aides","United States Secretaries of Defense","United States vice-presidential candidates, 2000","United States vice-presidential candidates, 2004","University of Wisconsin-Madison alumni","University of Wyoming alumni","Vice Presidents of the United States","White House Chiefs of Staff","Wyoming Republicans","Republican Party members of the United States House of Representatives","Writers from Wyoming","Writers from Nebraska","American political writers"],"wgBreakFrames":false,"wgPageContentLanguage":"en","wgPageContentModel":"wikitext","wgSeparatorTransformTable":["",""],"wgDigitTransformTable":["",""],"wgDefaultDateFormat":"dmy","wgMonthNames":["","January","February","March","April","May","June","July","August","September","October","November","December"],"wgMonthNamesShort":["","Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec"],"wgRelevantPageName":"Dick_Cheney","wgRelevantArticleId":5058628,"wgRestrictionEdit":false,"wgRestrictionMove":["autoconfirmed"],"wgRestrictionMove":["autoconfirmed"],"wgMediaViewerOnClick":true,"wgMediaViewerEnabledByDefault":true,"wikilove-enabled":"","wikilove-...
}
```

# Finding the right information in Soup


## James Brown

From Wikipedia, the free encyclopedia

For other people named James Brown, see [James Brown \(disambiguation\)](#).

**James Joseph Brown**<sup>[1]</sup> (May 3, 1933 – December 25, 2006) was an American singer and dancer. The founding father of [funk](#) music and a major figure of 20th-century popular music and dance, he is often referred to as "The Godfather of Soul". In a career that spanned six decades, Brown influenced the development of several music genres.<sup>[2]</sup>

Brown began his career as a [gospel](#) singer in [Toccoa, Georgia](#). Joining an [R&B](#) vocal group called the Avons that later evolved to become [The Famous Flames](#), Brown served as the group's lead singer.<sup>[3][4]</sup> First coming to national public attention in the late 1950s with the hit ballads "Please, Please, Please" and "Try Me", Brown built a reputation as a tireless live performer with the Famous Flames and his backing band, sometimes known as the James Brown Band or the James Brown Orchestra. Brown's success peaked in the 1960s with the live album *Live at the Apollo* and hit singles such as "Papa's Got a Brand New Bag", "I Got You (I Feel Good)" and "It's a Man's Man's Man's World". During the late 1960s, Brown moved from a continuum of blues and gospel-based forms and styles to a profoundly "[Africanized](#)" approach to music-making that influenced the development of funk music.<sup>[5]</sup> By the early 1970s, Brown had fully established the funk sound after the formation of [The J.B.'s](#) with records such as "Get Up (I Feel Like Being a) Sex Machine" and "The Payback". Brown also became notable for songs of social commentary, including the 1969 hit "Say It Loud – I'm Black and I'm Proud". Brown continued



James Brown, 1966.

Background information	
<b>Birth name</b>	James Joseph Brown
<b>Born</b>	May 3, 1933 Barnwell, South Carolina, United States
<b>Origin</b>	Toccoa, Georgia
<b>Died</b>	December 25, 2006 (aged 73) Atlanta, Georgia
<b>Genres</b>	Funk, soul, R&B
<b>Occupation(s)</b>	Singer, producer, dancer, bandleader

late 1960s, Brown moved from a continuum of blues and gospel-based forms and styles to a profoundly "[Africanized](#)" approach to music-making that influenced the development of funk music.<sup>[5]</sup> By the early 1970s, Brown had fully established the funk sound after the formation of [The J.B.'s](#) with records such as "Get Up (I Feel Like Being a) Sex Machine" and "The Payback". Brown also became notable for songs of social commentary, including the 1969 hit "Say It Loud – I'm Black and I'm Proud". Brown continued

Original	Toccoa, Georgia
Died	December 25, 2006 (aged 73) Atlanta, Georgia
Genres	Funk, soul, R&B
Occupation(s)	Singer, producer, dancer, bandleader

Cite this page  
Print/export  
Create a book  
Download as PDF  
Printable version  
Languages  
Alemannisch

Inspector  
tbody  
tr  
th  
Died

86.4833x38.9667 static  
84.4833x36.9667

```
def not_exist(name):  
    return name.find("div", {"class" : "noarticletext"})  
  
def is_dead(name):  
    return name.find("th", text="Died")  
  
def is_ambiguous(name):  
    return name.find("a", text="disambiguation")
```



# Running the script - Ouch

---

```
(test)🐞 nono dead_yet python dead_html.py
Please enter a name (e.g. "James Brown" or "Mark E. Smith")
Scraping HTML with BS took 3.40938568115e-05 seconds
(test)🐞 nono dead_yet python dead_html.py dick cheney
Dick Cheney is still alive
Scraping HTML with BS took 1.26621294022 seconds
(test)🐞 nono dead_yet python dead_html.py james brown
James Brown is dead
Scraping HTML with BS took 0.954394102097 seconds
(test)🐞 nono dead_yet python dead_html.py madonna
Name is ambiguous
Scraping HTML with BS took 0.165064811707 seconds
(test)🐞 nono dead_yet python dead_html.py james clown
I don't know this person
Scraping HTML with BS took 0.34019613266 seconds
🐞🐞
```

# Specific script - Ouch

---

```
def dick_dead():  
    r = requests.get('https://en.wikipedia.org/wiki/Dick_Cheney')  
    soup = BeautifulSoup(r.text)  
    if is_dead(soup):  
        print "Yes, Dick is dead!"  
    else:  
        print "Not yet..."
```

---

```
(test)🐞 nono dead_yet python dick_dead_html.py  
Not yet...  
Scraping HTML with BS took 0.774477005005 seconds  
(test)🐞 nono dead_yet python dick_dead_html.py  
Not yet...  
Scraping HTML with BS took 0.787503004074 seconds  
(test)🐞 nono dead_yet python dick_dead_html.py  
Not yet...  
Scraping HTML with BS took 0.79307103157 seconds  
(test)🐞 nono dead_yet python dick_dead_html.py  
Not yet...  
Scraping HTML with BS took 0.786529064178 seconds
```



# Using the Wikipedia API

- Whole Page:

<https://en.wikipedia.org/w/api.php?>

[action=query&titles=<<name>>&prop=revisions&rvprop=content&format=json](https://en.wikipedia.org/w/api.php?action=query&titles=<<name>>&prop=revisions&rvprop=content&format=json)

```
{
  "batchcomplete": "",
  "query": {
    "pages": {
      "15766": {
        "pageid": 15766,
        "ns": 0,
        "title": "James Brown",
        "revisions": [
          {
            "contentformat": "text/x-wiki",
            "contentmodel": "wikitext",
            "*": "{{other people|James Brown}}\n{{Use mdy dates|date=December 2013}}\n{{Infobox
musical artist\n|background          = solo_singer\n|image              = JamesBrown1986Pic.jpg\n|image_size      =
<!-- if smaller than 220px, otherwise leave blank -->\n|caption          = James Brown, 1986.\n|birth_name
= James Joseph Brown\n|birth_date    = {{birth date|1933|5|3}}\n|birth_place    = [[Barnwell, South
Carolina]], [[United States]]\n|origin      = [[Toccoa, Georgia]]\n|death_date    = {{death date and
age|2006|12|25|1933|5|3}}\n|death_place    = [[Atlanta]], [[Georgia (U.S. state)|Georgia]]\n|genre
= [[Funk]], [[soul music|soul]], [[Rhythm and blues|R&B]]\n|occupation    = [[Singer]], [[Record
producer|producer]], [[dancer]], [[bandleader]]\n|instrument    = [[Vocals]], [[drums]], [[percussion]],
[[organ (music)|organ]], [[keyboards]]\n|years_active    = 1954\u201320131988; 1991\u20132006\n|label
= [[Federal Records|Federal]], [[King Records (USA)|King]], Dade, [[Try Me Records|Try Me]], [[Smash
Records|Smash]], [[People Records|People]], [[Polydor Records|Polydor]], [[Scotti Brothers Records|Scotti
Bros.]]\n|associated_acts = [[The Famous Flames]], [[The J.B.'s]], [[The Dapps]], [[Bobby Byrd]], The Soul
Generals, [[Lyn Collins]], [[Bobby Bennett (The Famous Flames)|Bobby Bennett]], [[Bootsy
Collins]]\n|website
= {{URL|www.jamesbrown.com/}}\n}}\n''James Joseph Brown''<ref name=jbbirthname/>
(May 3, 1933&nbsp;\u2013December 25, 2006) was an American singer and dancer. The founding father of [[funk]]
```

# Using the Wikipedia API - Categories

```
{
  "continue": {
    "clcontinue": "15766|All_articles_with_unsourced_statements",
    "continue": ""
  },
  "query": {
    "pages": {
      "15766": {
        "pageid": 15766,
        "ns": 0,
        "title": "James Brown",
        "categories": [
          {
            "ns": 14,
            "title": "Category:"
          },
          {
            "ns": 14,
            "title": "Category:"
          },
          {
            "ns": 14,
            "title": "Category:"
          },
          {
            "ns": 14,
            "title": "Category:"
          },
          {
            "ns": 14,
            "title": "Category:African-American composers"
          },
          {
            "ns": 14,
            "title": "Category:African-American male dancers"
          },
          {
            "ns": 14,
            "title": "Category:African-American male singers"
          },
          {
            "ns": 14,
            "title": "Category:African-American rock musicians"
          },
          {
            "ns": 14,
            "title": "Category:African-Americans' civil rights activists"
          },
          {
            "ns": 14,
            "title": "Category:All articles with dead external links"
          }
        ]
      }
    }
  }
}
```

```
    "ns": 0,
    "title": "James Brown",
    "categories": [
      {
        "ns": 14,
        "title": "Category:1933 births"
      },
      {
        "ns": 14,
        "title": "Category:2006 deaths"
      },
      {
        "ns": 14,
        "title": "Category:20th-century American singers"
      }
    ]
  }
}
```

# Using the API - 3

---

- Dead people have a category XXXX deaths
- Living people have got at XXXX births category but not a XXXX deaths category
- If there is an ambiguous category, name is ambiguous (will not have birth or death category)
- If there isn't a birth category I'm assuming it's not a person
- Non existing page - will not have categories

# Using the Api - selecting categories

---

```
5
6 def categories(response):
7     return response.json()["query"]["pages"].values()[0]["categories"]
8
9 def is_ambiguous(dic):
10    for item in dic:
11        if 'disambiguation' in item["title"]:
12            return True
13    return False
14
15 def is_dead(dic):
16    for item in dic:
17        if 'deaths' in item["title"]:
18            return True
19    return False
20
21 def was_born(dic):
22    for item in dic:
23        if 'births' in item["title"]:
24            return True
25    return False
26
27
```



# Running general script

---

```
(test)👤👤 nono dead_yet python dead_json.py dick cheney
```

```
Dick Cheney is alive
```

```
Calling Wikipedia API took 0.119389057159 seconds
```

**was 1.2!**

```
(test)👤👤 nono dead_yet python dead_json.py james brown
```

```
James Brown is dead
```

```
Calling Wikipedia API took 0.113177061081 seconds
```

**was 0.9!**

```
(test)👤👤 nono dead_yet python dead_json.py madonna
```

```
Not sure who you're talking about
```

```
Calling Wikipedia API took 0.141903877258 seconds
```

```
(test)👤👤 nono dead_yet python dead_json.py james clown
```

```
Person/Wikipediapage does not exist
```

```
Calling Wikipedia API took 0.131882190704 seconds
```

# Specific script

```
def dick_dead_json():  
    r = requests.get('https://en.wikipedia.org/w/api.php?action=query&titles=Dick%20  
    text = r.json()["query"]["pages"].values()[0]["categories"][1]  
    if 'deaths' in text["title"]:  
        print "He's dead!"  
    else:  
        print "Not yet!"
```

```
-----  
(test)🐞 nono dead_yet python dick_dead_json.py  
Not yet!  
Using JSON API took 0.121052980423 seconds  
(test)🐞 nono dead_yet python dick_dead_json.py  
Not yet!  
Using JSON API took 0.113301992416 seconds  
(test)🐞 nono dead_yet python dick_dead_json.py  
Not yet!  
Using JSON API took 0.113208055496 seconds  
(test)🐞 nono dead_yet
```

**was 0.75!**

# Making it into a website

---

- Create a flask app
- Deploy to Heroku
- Add custom domain

# Flask - single page

---

- Flask is a python microframework
  - (<http://flask.pocoo.org/>)
- Single HTML Page

```
<!DOCTYPE html>
<html>
  <head>
    <title>Is Dick Dead Yet?</title>
    <link rel="stylesheet" type="text/css" href="{{ url_for('static', filename='style.css') }}">
  </head>
  <body class={{ name }}>
    <p>{{ result }}</p>
  </body>
</html>
```



# Flash App

---

```
1  from flask import Flask, render_template
2  import requests
3  from wiki_request import categories, is_dead
4
5  app = Flask(__name__)
6
7  @app.route('/', methods=['GET', 'POST'])
8  def index():
9      result = "Test"
10     url = 'https://en.wikipedia.org/w/api.php?action=query&titles=Dick Cheney'
11     r = requests.get(url)
12     d = categories(r)
13     if is_dead(d):
14         name = "dead"
15         result = "Dick Cheney is Dead!"
16     else:
17         name = "alive"
18         result = "Dick Cheney is not dead yet..."
19     return render_template('index.html', name=name, result=result)
20
21 if __name__ == '__main__':
22     app.run()
23
```

# Heroku

---

- simple deploy with heroku - updating easy as using git
- added custom domain ( <http://www.isdickdeadyet.xyz/> )

