

# Wikipedia API

Dr. H

For questions or bugs, contact Dr. H

**Why?** → Wikipedia is a good source of text, network, and Image data

**Note:** Wikipedia is an **Open API** (no API key required)

For more see: <https://wikipedia.readthedocs.io/en/latest/code.html>

## Installation

From the command line

```
conda deactivate
conda active dsan5000
conda install -c conda-forge wikipedia
python -m pip install wikipedia_sections
```

To generate the wordcloud you need

```
conda install -c conda-forge wordcloud
```

To use a conda enviroment inside VSC run, if you haven't already

```
conda install ipykernel --update-deps --force-reinstall
```

## Import

► Code

# Search

**command:** wikipedia.search(query, results=10, suggestion=False)

- Does a Wikipedia search for query.
- results - the maximum number of results returned
- suggestion - if True, return results and suggestion (if any) in a tuple

**returns:** list of titles of wikipedia page titles relevant to search

## ► Code

```
['Association football',  
 'Maria Soccor',  
 'Darrin Henson',  
 'Soccor Velho',  
 'Palm Beach SC',  
 'Robert Clivillés',  
 'Mt Gravatt Hawks FC',  
 'Taringa Rovers SFC',  
 'United States women's national soccer team',  
 'Brisbane City FC']
```

# Summary

**command:** wikipedia.summary(query, sentences=0, chars=0, auto\_suggest=True, redirect=True)

- *arguments*:
- sentences - if set, return the first sentences sentences (can be no greater than 10).
- chars - if set, return only the first chars characters (actual text returned may be slightly longer).
- auto\_suggest - let Wikipedia find a valid page title for the query
- redirect - allow redirection without raising RedirectError

**Returns:** Plain text summary of the page.

## ► Code

```
"Palm Beach Soccer Club is an Australian amateur soccer club based in Palm Beach, Queensland. The club was formed in 1966. The club plays in the Football Queensland Premier League 3 - South Coast, the top flight of the Football Queensland South Coast administrative division and the fifth flight of the Australian soccer league system. The club has won 6 premierships, 7 championships and 4 President's Cups within the division."
```

## Page

**command:** wikipedia.page(title=None, pageid=None, auto\_suggest=True, redirect=True, preload=False)

- *Arguments:*
- title - the title of the page to load
- pageid - the numeric pageid of the page to load
- auto\_suggest - let Wikipedia find a valid page title for the query
- redirect - allow redirection without raising RedirectError
- preload - load content, summary, images, references, and links during initialization

**Returns:** a WikipediaPage object for the page with title or the pageid (mutually exclusive).

See the link above for more info on objects of the class “wikipedia.WikipediaPage”

► Code

## Geosearch

Do a wikipedia geo search for latitude and longitude using HTTP API described in <http://www.mediawiki.org/wiki/Extension:GeoData>

- Arguments:
- latitude (float or decimal.Decimal)
- longitude (float or decimal.Decimal)
- Keyword arguments:
- title - The title of an article to search for
- results - the maximum number of results returned
- radius - Search radius in meters. The value must be between 10 and 10000

## ► Code

```
['Zero Milestone',  
 'National Christmas Tree (United States)',  
 'White House vegetable garden',  
 "President's Park",  
 'White House tennis court',  
 'The Ellipse',  
 'South Lawn',  
 '1974 White House helicopter incident',  
 'General William Tecumseh Sherman Monument',  
 'Office of Trade and Manufacturing Policy']
```

## Explore the WikipediaPage object

## ► Code

"Palm Beach Soccer Club is an Australian amateur soccer club based in Palm Beach, Queensland. The club was formed in 1966. The club plays in the Football Queensland Premier League 3 - South Coast, the top flight of the Football Queensland South Coast administrative division and the fifth flight of the Australian soccer league system. The club has won 6 premierships, 7 championships and 4 President's Cups within the division. \n\n\n== History ==\n\n\n=== Palm Beach ===\n\nIn 1964 Eddie Wardell organised a group of boys, one of which was his son, to play football out of Palm Beach and compete in a three team football competition against Kingscliff and Twin Towns. The first game for Palm Beach was held at Salk Oval and the team members wore plain green T-shirts but the team would later move to a council paddock also located in Palm Beach. By 1965 the local Gold Coast Association football competition wanted a presence in Palm Beach and invited the team to join the local junior competitions. The first official playing strip was debuted in 1965 which was a plain T-shirt that embossed with a single blue stripe.\n\nIn 1966 the team incorporated three teams; under 8s, under 10s and under 12s. During sign on time it was decided the team needed to unofficially become a club and held its inaugural meeting in the Currumbin RSL. Neville Cripps, a father of a Palm Beach player, was elected the first president and names for the club began being thrown around. Names such as Palm Beach Hotspurs and Palm Beach Pilchards were put forward at the meeting but Palm Beach Currumbin Soccer Club was voted the winner. In 1966 the club began wearing playing strip which was a white shirt emblazoned with a large blue vee.\n\n1971 saw the council reclaim the playing field of the club and urgent calls to council were made which led to a field behind the Tugun Bowls Club becoming their new home ground. With the team being located in Tugun their official name was changed to the Palm Beach Currumbin Tugun Soccer Club. The following year while John Neumann was developing the suburb of Palm Beach he suggested to the club that they should attempt to be given two recreational fields in Mallawa Drive, Palm Beach. Club president Mal Sutherland met with the council which would eventually lead to the club being given two fields at their present location on Mallawa Drive. The Palm Beach Currumbin Lions Football Club donated the blue

## ► Code

```
<class 'str'>
<div class="mw-parser-output"><style data-mw-deduplicate="TemplateStyles:r1097763485">.mw-parser-output
.ambox{border:1px solid #a2a9b1;border-left:10px solid #36c;background-color:#fbfbfb;box-sizing:border-box}.mw-
parser-output .ambox+link+.ambox,.mw-parser-output .ambox+link+style+.ambox,.mw-parser-output
.ambox+link+link+.ambox,.mw-parser-output .ambox+.mw-empty-elt+link+.ambox,.mw-parser-output .ambox+.mw-empty-
elt+link+style+.ambox,.mw-parser-output .ambox+.mw-empty-elt+link+link+.ambox{margin-top:-1px}html
body.mediawiki .mw-parser-output .ambox.mbox-small-left{margin:4px 1em 4px 0;overflow:hidden;width:238px;border-
collapse:collapse;font-size:88%;line-height:1.25em}.mw-parser-output .ambox-speedy{border-left:10px solid
#b32424;background-color:#fee7e6}.mw-parser-output .ambox-delete{border-left:10px solid #b32424}.mw-parser-
output .ambox-content{border-left:10px solid #f28500}.mw-parser-output .ambox-style{border-left:10px solid
#fc3}.mw-parser-output .ambox-move{border-left:10p
```

## ► Code

```
['1966 establishments in Australia',  
 'All Wikipedia articles written in Australian English',  
 'All articles lacking in-text citations',  
 'Articles lacking in-text citations from December 2020',  
 'Articles with short description',  
 'Association football clubs established in 1966',  
 'Brisbane Premier League teams',  
 'National Premier Leagues clubs',  
 'Short description matches Wikidata',  
 'Soccer teams on the Gold Coast, Queensland',  
 'Use Australian English from June 2020',  
 'Use dmy dates from June 2018']
```

## ► Code

```
['https://upload.wikimedia.org/wikipedia/commons/8/88/Flag_of_Australia_%28converted%29.svg',  
 'https://upload.wikimedia.org/wikipedia/commons/0/04/Flag_of_Queensland.svg',  
 'https://upload.wikimedia.org/wikipedia/commons/8/83/Kit_body.svg',  
 'https://upload.wikimedia.org/wikipedia/commons/7/77/Kit_body_lightbluestripes.png',  
 'https://upload.wikimedia.org/wikipedia/commons/7/7f/Kit_left_arm.svg',  
 'https://upload.wikimedia.org/wikipedia/commons/b/b5/Kit_left_arm_lightblue_stripes.png',  
 'https://upload.wikimedia.org/wikipedia/commons/c/cd/Kit_right_arm.svg',  
 'https://upload.wikimedia.org/wikipedia/commons/e/ec/Kit_right_arm_lightblue_stripes.png',  
 'https://upload.wikimedia.org/wikipedia/commons/a/af/Kit_shorts.svg',  
 'https://upload.wikimedia.org/wikipedia/commons/9/9b/Kit_socks_long.svg',  
 'https://upload.wikimedia.org/wikipedia/commons/a/a4/Text_document_with_red_question_mark.svg',  
 'https://upload.wikimedia.org/wikipedia/en/8/8d/Palm_Beach_SC.jpg']
```

## ► Code

```
['1884 Anglo-Queensland FA Cup',  
 '1991 in Queensland soccer',  
 '1992 in Queensland Soccer',  
 '1993 in Queensland Soccer',  
 '1994 in Queensland Soccer',  
 '1995 in Queensland Soccer',  
 '1996 in Queensland Soccer',
```



```
'1997 in Queensland Soccer',  
'1998 in Queensland Soccer',  
'1999 in Queensland Soccer',  
'2000 in Queensland Soccer',  
'2001 in Queensland Soccer',  
'2002 in Queensland Soccer',  
'2003 Football Queensland South Coast Season',  
'2003 in Queensland Soccer',  
'2004 Football Queensland South Coast Season',  
'2004 in Queensland Soccer',  
'2005 Football Queensland South Coast Season',  
'2005 in Queensland Soccer',  
'2006 Football Queensland South Coast Season',  
'2006 in Queensland Soccer',  
'2007 Football Queensland South Coast Season'
```

## ► Code

```
['http://www.gccfc.com.au/',  
 'http://www.palmbeachsoccerclub.com.au/',  
 'http://www.socceraust.co.uk/2014/QLD/QLD.html']
```

## ► Code

```
[]
```

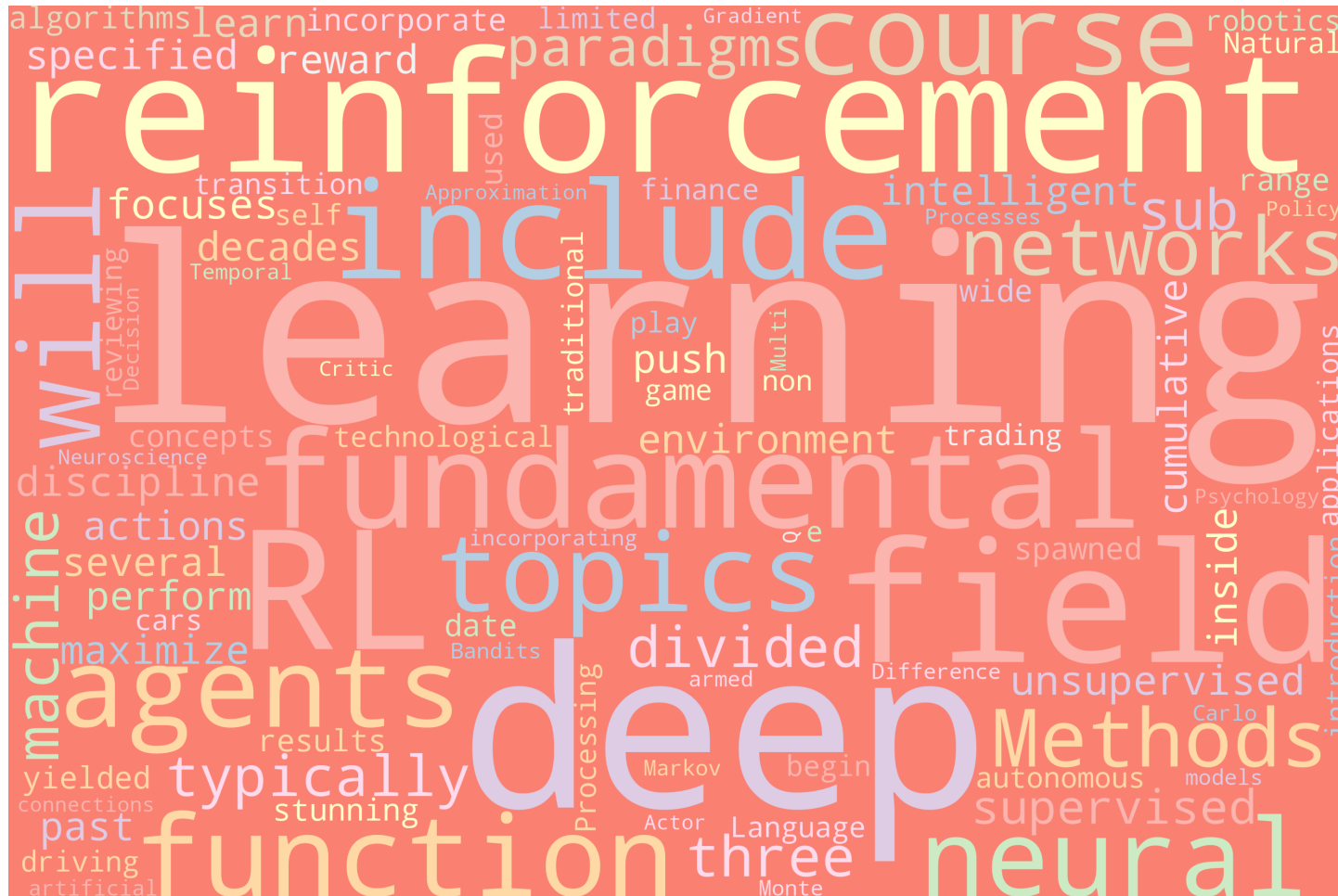
## ► Code

"Palm Beach Soccer Club is an Australian amateur soccer club based in Palm Beach, Queensland. The club was formed in 1966. The club plays in the Football Queensland Premier League 3 - South Coast, the top flight of the Football Queensland South Coast administrative division and the fifth flight of the Australian soccer league system. The club has won 6 premierships, 7 championships and 4 President's Cups within the division."

# Word cloud example

Search over composite topics

► Code



# Wikipedia text-mining example

- The following script does;
  - It loops over a list of countries and topics.
  - It searches Wikipedia for pages associated with those topics.
  - It scrapes the plaintext content from those pages.
  - It builds a word cloud with the text for each country.

## ► Code

```
----- usa -----
topic =  travel in usa
TITLES= ['Work and Travel USA', 'Travel agency']
-----
Work and Travel USA
https://en.wikipedia.org/wiki/Work_and_Travel_USA
[]
-----
Travel agency
https://en.wikipedia.org/wiki/Travel_agency
[]
topic =  tourism in usa
TITLES= ['Tourism in the United States', 'Tourism']
-----
Tourism in the United States
https://en.wikipedia.org/wiki/Tourism_in_the_United_States
[]
```

-----

Tourism  
<https://en.wikipedia.org/wiki/Tourism>  
[]  
f-----





