day_two

March 15, 2016

1 Day 2 - Getting Data with Python

1.1 Introduction

A vast amount of data exists on the web and is now publicly available. In this section, we give an overview of popular ways to retrieve data from the web, and walk through some important concerns and considerations.
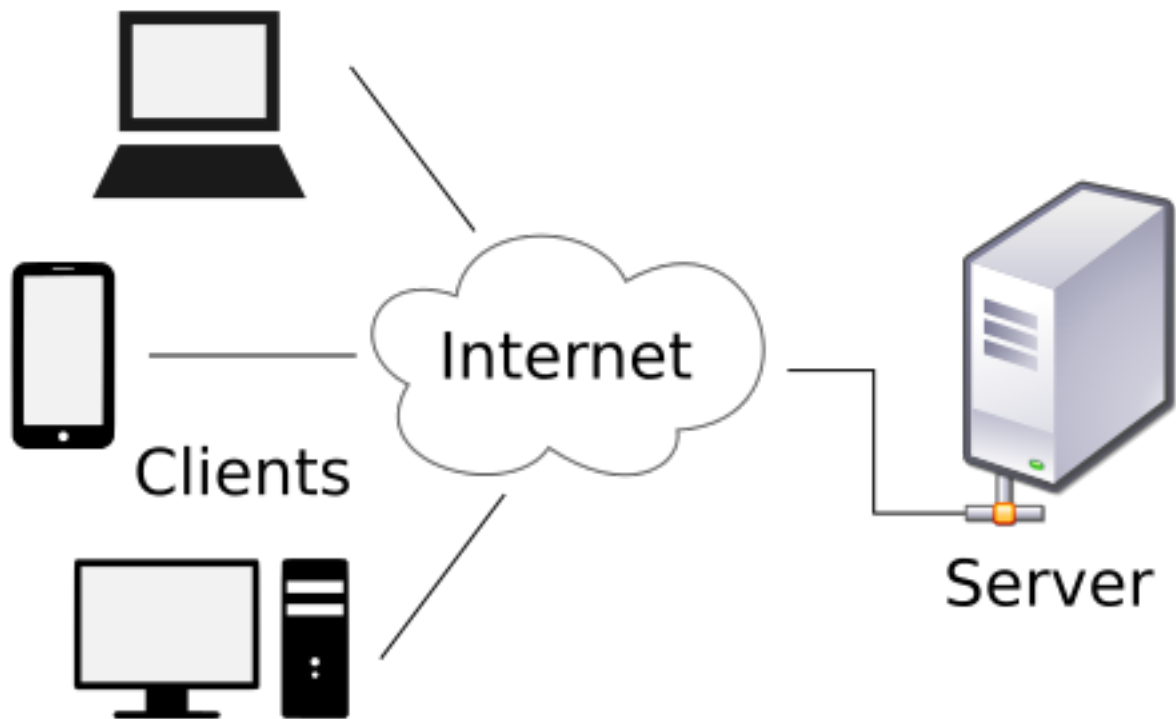
Figure 1: An extremely simplified model of the web

The internet follows a client-server architecture, where clients (e.g. you) ask servers to do things.

The most common way that you experience this is through a browser, where you enter a URL and a server sends your computer a page for your browser to render. Most of what you think about as the internet are stored documents (web pages) that are given out to anyone who asks.

You probably also have a program on your computer like Outlook or Thunderbird that sends emails to a server and asks it to forward them along to someone else. You may also have proprietary software that's protected by a license, and needs to connect to a license server to verify that you are an authenticated user.

Ultimately, the internet is just connecting to computers that you don't own and passing data back and forth. Because the data transfer protocol (`http`) and typical data formats (`html`) are not native to Python, we're going to leave Python just for a little bit.

1.2 Intro to HTTP requests

You can view the request sent by your browser by:

1) Opening a new tab in your browser

2) Enabling developer tools (**View -> Developer -> Developer Tools in Chrome** and **Tools -> Web Developer -> Toggle Tools in Firefox**)

3) Loading or reloading a web page (etc. www.google.com)

4) Navigating to the Network tab in the panel that appears at the bottom of the page.

These requests you send follow the HTTP protocol (Hypertext Transfer Protocol), part of which defines the information (along with the format) the server needs to receive to return the right resources. Your HTTP request contains **headers**, which contains information that the server needs to know in order to return the right information to you.

But we're not here to wander around the web (you probably do this a lot, all on your own). You're here because you want Python to do it for you.

In order to get web pages, we're going to use a python library called `requests`, which takes a lot of the fuss out of contacting servers.

```
In [1]: import requests

        r = requests.get("http://en.wikipedia.org/wiki/Main_Page")
```

This response object contains various information about the request you sent to the server, the resources returned, and information about the response the server returned to you, among other information. These are accessible through the **request** attribute, the **content** attribute and the **headers** attribute respectively, which we'll each examine below.

```
In [2]: type(r.request), type(r.content), type(r.headers)

Out[2]: (requests.models.PreparedRequest,
         bytes,
         requests.structures.CaseInsensitiveDict)
```

Here, we can see that **request** is an object with a custom type, **content** is a str value and **headers** is an object with "dict" in its name, suggesting we can interact with it like we would with a dictionary.

The content is the actual resource returned to us - let's take a look at the content first before examining the request and response objects more carefully. (We select the first 1000 characters b/c of the display limits of Jupyter/python notebook.)

```
In [3]: from pprint import pprint
        pprint(r.content[0:1000])

(b'<!DOCTYPE html>\n<html lang="en" dir="ltr" class="client-nojs">\n<head>\n<m'
 b'eta charset="UTF-8" />\n<title>Wikipedia, the free encyclopedia</title>\n<'
 b'script>document.documentElement.className = document.documentElement.classNa'
 b'me.replace( /(^|\\s)client-nojs(\\s|$)/, "$1client-js$2" );</script>\n<scri'
 b'pt>(window.RLQ = window.RLQ || []).push(function () {\nmw.config.set({"wg'
 b'CanonicalNamespace":"","wgCanonicalSpecialPageName":false,"wgNamespaceNumber'
```
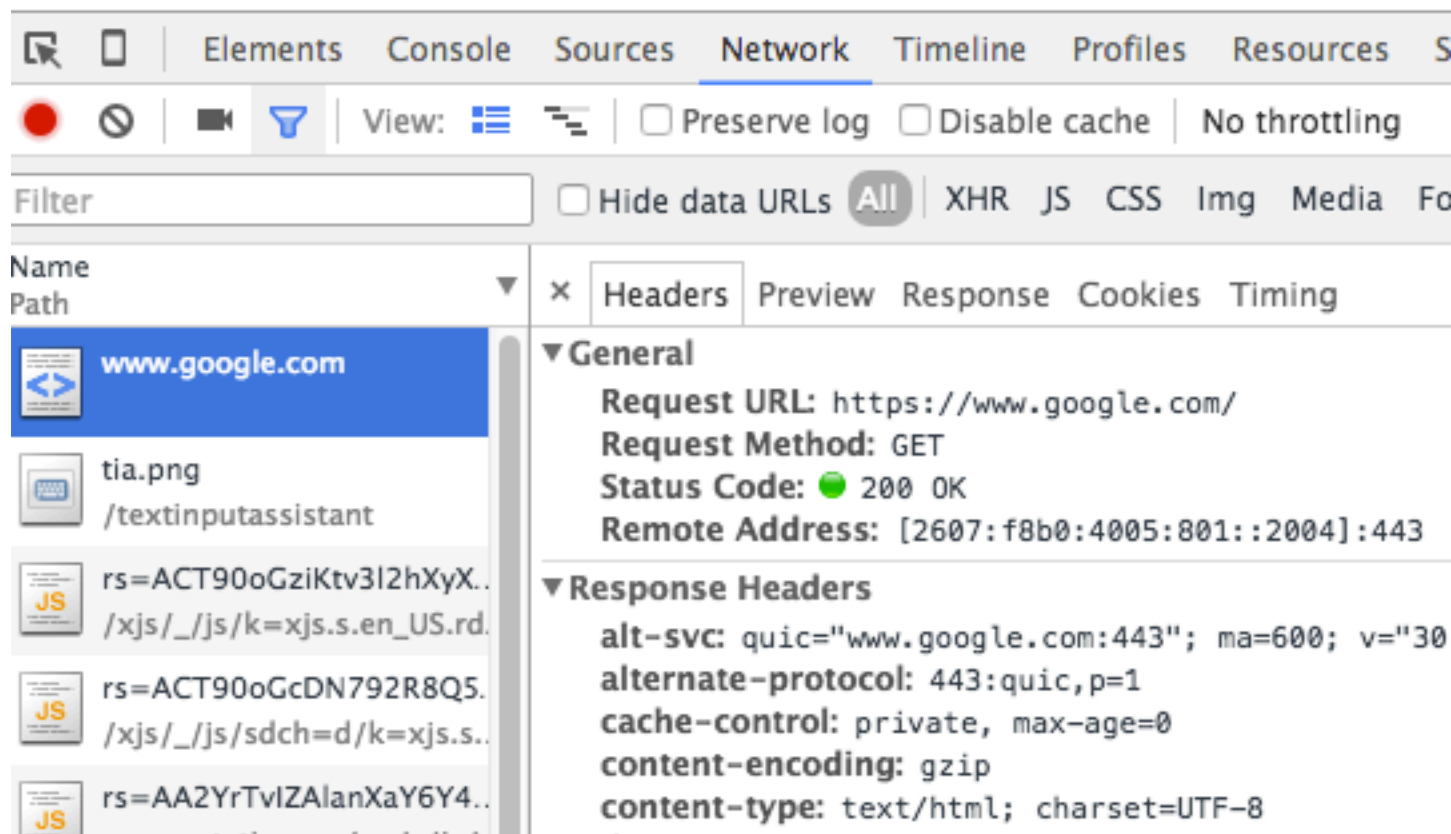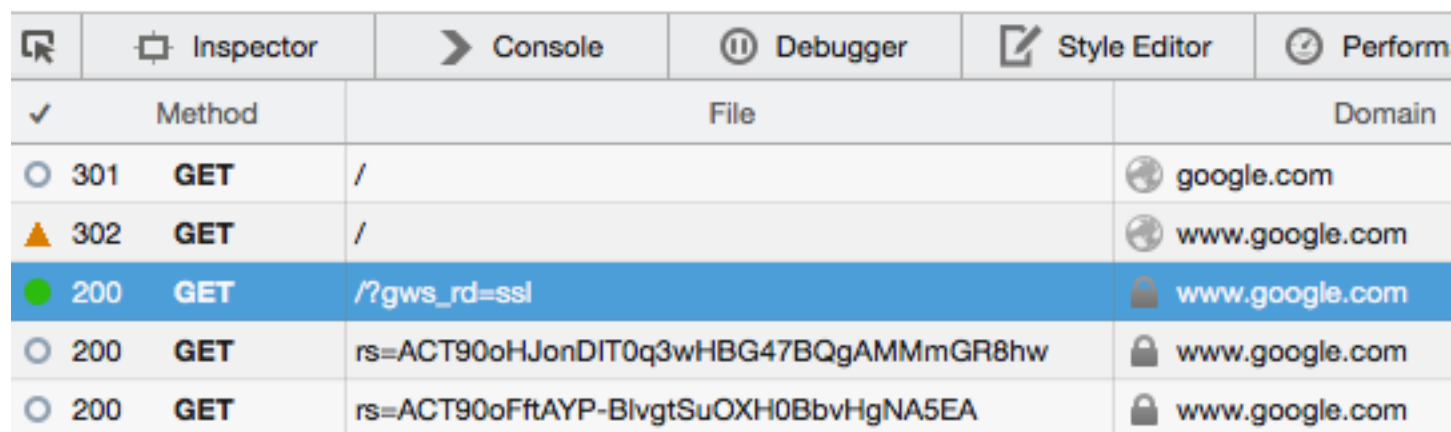
Figure 2: Chrome Examine Request Example

✓	Method	File	Domain
○ 301	**GET**	/	🌐 google.com
🔺 302	**GET**	/	🌐 www.google.com
🟢 200	**GET**	/?gws_rd=ssl	🔒 www.google.com
○ 200	**GET**	rs=ACT90oHJonDIT0q3wHBG47BQgAMMmGR8hw	🔒 www.google.com
○ 200	**GET**	rs=ACT90oFftAYP-BlvgtSuOXH0BbvHgNA5EA	🔒 www.google.com

Figure 3: Firefox Examine Request Example

```
b'":0,"wgPageName":"Main_Page","wgTitle":"Main Page","wgCurRevisionId":6968469'
b'20,"wgRevisionId":696846920,"wgArticleId":15580374,"wgIsArticle":true,"wgIsR'
b'edirect":false,"wgAction":"view","wgUserName":null,"wgUserGroups":["*"],"wgC'
b'ategories":[],"wgBreakFrames":false,"wgPageContentLanguage":"en","wgPageCont'
b'entModel":"wikitext","wgSeparatorTransformTable":["",""],"wgDigitTransformTa'
b'ble":["",""],"wgDefaultDateFormat":"dmy","wgMonthNames":["","January","Febru'
b'ary","March","April","May","June","July","August","September","October","Nov'
b'ember","December"],"wgMonthN')
```

The content returned is written in HTML (_H_yper _T_ext _M_arkup _L_anguage), which is the default format in which web pages are returned. The content looks like gibberish at first, with little to no spacing. The reason for this is that some of the formatting rules for the document, like its hierarchical structure, are saved in text along with the text in the document.

> note - this is called the _D_ocument _O_bject _M_odel (DOM) and is the same way that markdown and LaTeX documents are written

If you save a web page as a ".html" file, and open the file in a text editor like Notepad++ or Sublime Text, this is the same format you'll see. Opening the file in a browser (i.e. by double-clicking it) gives you the Google home page you are familiar with.

You can inspect the information you sent to Wikipedia long with your request

In [4]: `r.request.headers`

Out[4]: `{'Accept-Encoding': 'gzip, deflate', 'User-Agent': 'python-requests/2.9.1', 'Accept': '*/*', 'Co`

Along with the additional info that Wikipedia sent back:

In [5]: `r.headers`

Out[5]: `{'P3P': 'CP="This is not a P3P policy! See https://en.wikipedia.org/wiki/Special:CentralAutoLog`

But you will probably not ever need this information.

Most of what you'll be doing is sending what are called `GET` requests (this is why we typed in `requests.get` above). This is an `HTTP` protocol for asking a server to send you some stuff. We asked Wikipedia to `GET` us their main page. Things like queries (searching Wikipedia) also fall under `GET`.

From time to time, you may also want to send information to a server (we'll do this later today). These are called `POST` requests, because you are posting something to the server (and not asking for data back).

> note - From the server's perspective, the request it receives from your browser is not so different from the request received from your console (though some servers use a range of methods to determine if the request comes from a "valid" person using a browser, versus an automated program.)

To have a look at the content of the web page, we can ask for the content:

In [6]: `r.content[:1000]`

Out[6]: `b'<!DOCTYPE html>\n<html lang="en" dir="ltr" class="client-nojs">\n<head>\n<meta charset="UTF-8'`

which gives us the response in bytes, or text:

In [7]: `r.text[:1000]`

Out[7]: `'<!DOCTYPE html>\n<html lang="en" dir="ltr" class="client-nojs">\n<head>\n<meta charset="UTF-8"`

1.3 Parsing HTML in Python

Trying to parse this `str` by hand is basically a nightmare. Instead, we'll use a Python library called Beautiful
Soup to turn it into something that is still confusing, but less of a nightmare.

```
In [8]: from bs4 import BeautifulSoup

        page = BeautifulSoup(r.content)
        page
```

```
/Users/dillon/anaconda/lib/python3.5/site-packages/bs4/__init__.py:166: UserWarning: No parser was explic

To get rid of this warning, change this:

 BeautifulSoup([your markup])

to this:

 BeautifulSoup([your markup], "lxml")

  markup_type=markup_type))
```

```
Out[8]: <!DOCTYPE html>
        <html class="client-nojs" dir="ltr" lang="en">
        <head>
        <meta charset="utf-8"/>
        <title>Wikipedia, the free encyclopedia</title>
        <script>document.documentElement.className = document.documentElement.className.replace( /(^|\s)
        <script>(window.RLQ = window.RLQ || []).push(function () {
        mw.config.set({"wgCanonicalNamespace":"","wgCanonicalSpecialPageName":false,"wgNamespaceNumber"
        mw.user.tokens.set({"editToken":"+\\","patrolToken":"+\\","watchToken":"+\\","csrfToken":"+\\"})

        });mw.loader.load(["mediawiki.page.startup","mediawiki.legacy.wikibits","ext.centralauth.central
        } );</script>
        <link href="/w/load.php?debug=false&amp;lang=en&amp;modules=ext.gadget.DRN-wizard%2CReferenceTo
        <meta content="" name="ResourceLoaderDynamicStyles"/>
        <link href="/w/load.php?debug=false&amp;lang=en&amp;modules=site&amp;only=styles&amp;skin=vecto
        <script async="" src="/w/load.php?debug=false&amp;lang=en&amp;modules=startup&amp;only=scripts&a
        <meta content="MediaWiki 1.27.0-wmf.16" name="generator"/>
        <meta content="origin-when-cross-origin" name="referrer"/>
        <link href="android-app://org.wikipedia/http/en.m.wikipedia.org/wiki/Main_Page" rel="alternate"/
        <link href="/w/api.php?action=featuredfeed&amp;feed=potd&amp;feedformat=atom" rel="alternate" t
        <link href="/w/api.php?action=featuredfeed&amp;feed=featured&amp;feedformat=atom" rel="alternate
        <link href="/w/api.php?action=featuredfeed&amp;feed=onthisday&amp;feedformat=atom" rel="alterna
        <link href="/static/apple-touch/wikipedia.png" rel="apple-touch-icon"/>
        <link href="/static/favicon/wikipedia.ico" rel="shortcut icon"/>
        <link href="/w/opensearch_desc.php" rel="search" title="Wikipedia (en)" type="application/opense
        <link href="//en.wikipedia.org/w/api.php?action=rsd" rel="EditURI" type="application/rsd+xml"/>
        <link href="//creativecommons.org/licenses/by-sa/3.0/" rel="copyright"/>
        <link href="https://en.wikipedia.org/wiki/Main_Page" rel="canonical"/>
        <link href="//meta.wikimedia.org" rel="dns-prefetch"/>
        </head>
        <body class="mediawiki ltr sitedir-ltr ns-0 ns-subject page-Main_Page skin-vector action-view">
        <div class="noprint" id="mw-page-base"></div>
        <div class="noprint" id="mw-head-base"></div>
```

```
<div class="mw-body" id="content" role="main">
<a id="top"></a>
<div id="siteNotice"><!-- CentralNotice --></div>
<div class="mw-indicators">
</div>
<h1 class="firstHeading" id="firstHeading" lang="en">Main Page</h1>
<div class="mw-body-content" id="bodyContent">
<div id="siteSub">From Wikipedia, the free encyclopedia</div>
<div id="contentSub"></div>
<div class="mw-jump" id="jump-to-nav">
                                                Jump to:                                              <a href=
</div>
<div class="mw-content-ltr" dir="ltr" id="mw-content-text" lang="en"><table id="mp-topbanner" s
<tr>
<td style="width:61%; color:#000;">
<table style="width:280px; border:none; background:none;">
<tr>
<td style="width:280px; text-align:center; white-space:nowrap; color:#000;">
<div style="font-size:162%; border:none; margin:0; padding:.1em; color:#000;">Welcome to <a hre
<div style="top:+0.2em; font-size:95%;">the <a href="/wiki/Free_content" title="Free content">fr
<div id="articlecount" style="font-size:85%;"><a href="/wiki/Special:Statistics" title="Special
</td>
</tr>
</table>
</td>
<td style="width:13%; font-size:95%;">
<ul>
<li><a href="/wiki/Portal:Arts" title="Portal:Arts">Arts</a></li>
<li><a href="/wiki/Portal:Biography" title="Portal:Biography">Biography</a></li>
<li><a href="/wiki/Portal:Geography" title="Portal:Geography">Geography</a></li>
</ul>
</td>
<td style="width:13%; font-size:95%;">
<ul>
<li><a href="/wiki/Portal:History" title="Portal:History">History</a></li>
<li><a href="/wiki/Portal:Mathematics" title="Portal:Mathematics">Mathematics</a></li>
<li><a href="/wiki/Portal:Science" title="Portal:Science">Science</a></li>
</ul>
</td>
<td style="width:13%; font-size:95%;">
<ul>
<li><a href="/wiki/Portal:Society" title="Portal:Society">Society</a></li>
<li><a href="/wiki/Portal:Technology" title="Portal:Technology">Technology</a></li>
<li><b><a href="/wiki/Portal:Contents/Portals" title="Portal:Contents/Portals">All portals</a><,
</ul>
</td>
</tr>
</table>
<table id="mp-upper" style="width: 100%; margin:4px 0 0 0; background:none; border-spacing: 0px
<tr>
<td class="MainPageBG" style="width:55%; border:1px solid #cef2e0; background:#f5fffa; vertical-
<table id="mp-left" style="width:100%; vertical-align:top; background:#f5fffa;">
<tr>
<td style="padding:2px;">
```

```html
<h2 id="mp-tfa-h2" style="margin:3px; background:#cef2e0; font-family:inherit; font-size:120%;
</td>
</tr>
<tr>
<td style="color:#000;">
<div id="mp-tfa" style="padding:2px 5px">
<div id="mp-tfa-img" style="float: left; margin: 0.5em 0.9em 0.4em 0em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-w
</div>
<p><b><a href="/wiki/California_State_Route_78" title="California State Route 78">State Route 78
<ul style="list-style:none; margin-left:0; text-align:right;">
<li>Recently featured:
<div class="hlist inline">
<ul>
<li><i><a href="/wiki/Sarcoscypha_coccinea" title="Sarcoscypha coccinea">Sarcoscypha coccinea</a
<li><a href="/wiki/Japanese_battleship_Asahi" title="Japanese battleship Asahi">Japanese battles
<li><a href="/wiki/Isabella_Beeton" title="Isabella Beeton">Isabella Beeton</a></li>
</ul>
</div>
</li>
</ul>
<div class="hlist noprint" id="mp-tfa-footer" style="text-align: right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Today%27s_featured_article/March_2016" title="Wikipedia:Today's
<li><b><a class="extiw" href="https://lists.wikimedia.org/mailman/listinfo/daily-article-l" titl
<li><b><a href="/wiki/Wikipedia:Featured_articles" title="Wikipedia:Featured articles">More feat
</ul>
</div>
</div>
</td>
</tr>
<tr>
<td style="padding:2px;">
<h2 id="mp-dyk-h2" style="margin:3px; background:#cef2e0; font-family:inherit; font-size:120%;
</td>
</tr>
<tr>
<td style="color:#000; padding:2px 5px 5px;">
<div id="mp-dyk">
<div id="mp-dyk-img" style="float:right; margin-left:0.5em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-w
<div class="thumbcaption" style="padding: 0.25em 0; word-wrap: break-word;">Bilikiss Adebiyi</di
</div>
</div>
<ul>
<li>... that <b><a href="/wiki/Bilikiss_Adebiyi_Abiola" title="Bilikiss Adebiyi Abiola">Bilikiss
<li>... that the <a href="/wiki/BBC" title="BBC">BBC</a> re-launched its former television chann
<li>... that the Sanskrit text <b><a href="/wiki/Manasollasa" title="Manasollasa">Manasollasa</a
<li>... that the species name for <i><b><a href="/wiki/Burmaleon" title="Burmaleon">Burmaleon ma
<li>... that the documentary film <i><b><a href="/wiki/No_Land%27s_Song" title="No Land's Song">
<li>... that uninjured reporters commandeered a <a href="/wiki/Medical_evacuation" title="Medica
<li>... that of an estimated 100,000 <b><a href="/wiki/German_Jewish_military_personnel_of_World_
</ul>
<p><b>Correction</b>: we erroneously claimed here that in 1964 <a href="/wiki/Jim_Hazelton" titl
```

```html
<div class="hlist noprint" id="mp-dyk-footer" style="text-align:right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Recent_additions" title="Wikipedia:Recent additions">Recently in
<li><b><a href="/wiki/Wikipedia:Your_first_article" title="Wikipedia:Your first article">Start a
<li><b><a href="/wiki/Template_talk:Did_you_know" title="Template talk:Did you know">Nominate an
</ul>
</div>
</div>
</td>
</tr>
</table>
</td>
<td style="border:1px solid transparent;"></td>
<td class="MainPageBG" style="width:45%; border:1px solid #cedff2; background:#f5faff; vertical-
<table id="mp-right" style="width:100%; vertical-align:top; background:#f5faff;">
<tr>
<td style="padding:2px;">
<h2 id="mp-itn-h2" style="margin:3px; background:#cedff2; font-family:inherit; font-size:120%;
</td>
</tr>
<tr>
<td style="color:#000; padding:2px 5px;">
<div id="mp-itn">
<div id="mp-itn-img" style="float:right;margin-left:0.5em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-w
<div class="thumbcaption" style="padding: 0.25em 0; word-wrap: break-word;">Total solar eclipse
</div>
</div>
<ul>
<li><b><a href="/wiki/March_2016_Ankara_bombing" title="March 2016 Ankara bombing">An explosion<
<li>At least 18 people are killed in <b><a href="/wiki/2016_Grand-Bassam_shootings" title="2016
<li><a href="/wiki/Google_DeepMind" title="Google DeepMind">Google DeepMind</a>'s <a href="/wiki
<li>A total <a href="/wiki/Solar_eclipse" title="Solar eclipse">solar eclipse</a> <b><a href="/w
<li>In the <b><a href="/wiki/Slovak_parliamentary_election,_2016" title="Slovak parliamentary el
<li>The <a href="/wiki/Human_Rights_Protection_Party" title="Human Rights Protection Party">Human
</ul>
<ul style="list-style:none; margin-left:0;">
<li><b><a href="/wiki/Portal:Current_events" title="Portal:Current events">Ongoing events</a></b
<div class="hlist inline">
<ul>
<li><a href="/wiki/Zika_virus_outbreak_(2015%E2%80%93present)" title="Zika virus outbreak (2015p
<li><a href="/wiki/European_migrant_crisis" title="European migrant crisis">European migrant cri
</ul>
</div>
</li>
<li><b><a href="/wiki/Deaths_in_2016" title="Deaths in 2016">Recent deaths</a></b>:
<div class="hlist inline">
<ul>
<li><a href="/wiki/Hilary_Putnam" title="Hilary Putnam">Hilary Putnam</a></li>
<li><a href="/wiki/Lloyd_Shapley" title="Lloyd Shapley">Lloyd Shapley</a></li>
<li><a href="/wiki/Iolanda_Bala%C8%99" title="Iolanda Bala">Iolanda Bala</a></li>
</ul>
</div>
</li>
```

```
</ul>
</div>
</td>
</tr>
<tr>
<td style="padding:2px;">
<h2 id="mp-otd-h2" style="margin:3px; background:#cedff2; font-family:inherit; font-size:120%; 
</td>
</tr>
<tr>
<td style="color:#000; padding:2px 5px 5px;">
<div id="mp-otd">
<p><b><a href="/wiki/March_15" title="March 15">March 15</a></b>: <b><a href="/wiki/Ides_of_Marc
<div id="mp-otd-img" style="float:right;margin-left:0.5em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-w
<div class="thumbcaption" style="padding: 0.25em 0; word-wrap: break-word;">Pancho Villa</div>
</div>
</div>
<ul>
<li><a href="/wiki/1783" title="1783">1783</a>  A <b><a href="/wiki/Newburgh_Conspiracy" title="
<li><a href="/wiki/1892" title="1892">1892</a>  <b><a href="/wiki/Liverpool_F.C." title="Liverpo
<li><a href="/wiki/1916" title="1916">1916</a>  Six days after <a href="/wiki/Pancho_Villa" titl
<li><a href="/wiki/1941" title="1941">1941</a>  <b><a href="/wiki/Philippine_Airlines" title="Ph
<li><a href="/wiki/2011" title="2011">2011</a>  <a href="/wiki/Arab_Spring" title="Arab Spring">
</ul>
<ul style="list-style:none; margin-left:0;">
<li>More anniversaries:
<div class="hlist inline nowraplinks">
<ul>
<li><a href="/wiki/March_14" title="March 14">March 14</a></li>
<li><b><a href="/wiki/March_15" title="March 15">March 15</a></b></li>
<li><a href="/wiki/March_16" title="March 16">March 16</a></li>
</ul>
</div>
</li>
</ul>
<div class="hlist noprint" id="mp-otd-footer" style="text-align: right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Selected_anniversaries/March" title="Wikipedia:Selected annivers
<li><b><a class="extiw" href="https://lists.wikimedia.org/mailman/listinfo/daily-article-l" titl
<li><b><a href="/wiki/List_of_historical_anniversaries" title="List of historical anniversaries"
</ul>
<div style="font-size:smaller;">
<ul>
<li>Current date: <span class="nowrap">March 15, 2016</span> (<a href="/wiki/Coordinated_Univers
<li><span class="plainlinks" id="otd-purgelink"><span class="nowrap"><a class="external text" h
</ul>
</div>
</div>
</div>
</td>
</tr>
</table>
</td>
```

```html
</tr>
</table>
<table id="mp-lower" style="margin:4px 0 0 0; width:100%; background:none; border-spacing: 0px;
<tr>
<td class="MainPageBG" style="width:100%; border:1px solid #ddcef2; background:#faf5ff; vertical
<table id="mp-bottom" style="width:100%; vertical-align:top; background:#faf5ff; color:#000;">
<tr>
<td style="padding:2px;">
<h2 id="mp-tfp-h2" style="margin:3px; background:#ddcef2; font-family:inherit; font-size:120%;
</td>
</tr>
<tr>
<td style="color:#000; padding:2px;">
<div id="mp-tfp">
<table style="margin:0 3px 3px; width:100%; text-align:left; background-color:transparent; bord
<tr>
<td style="padding:0 0.9em 0 0;"><a class="image" href="/wiki/File:Ash_in_Yogyakarta_during_the_2
<td style="padding:0 6px 0 0">
<p>A man sweeping <a href="/wiki/Volcanic_ash" title="Volcanic ash">volcanic ash</a> in <a href=
<p><small>Photograph: <a href="/wiki/User:Crisco_1492" title="User:Crisco 1492">Chris Woodrich</
<ul style="list-style:none; margin-left:0; text-align:right;">
<li>Recently featured:
<div class="hlist inline">
<ul>
<li><a href="/wiki/Template:POTD/2016-03-14" title="Template:POTD/2016-03-14"><i>Homme au bain</
<li><a href="/wiki/Template:POTD/2016-03-13" title="Template:POTD/2016-03-13">Wagner VI project
<li><a href="/wiki/Template:POTD/2016-03-12" title="Template:POTD/2016-03-12">Lynx (constellati
</ul>
</div>
</li>
</ul>
<div class="hlist noprint" style="text-align:right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Picture_of_the_day/March_2016" title="Wikipedia:Picture of the d
<li><b><a href="/wiki/Wikipedia:Featured_pictures" title="Wikipedia:Featured pictures">More feat
</ul>
</div>
</td>
</tr>
</table>
</div>
</td>
</tr>
</table>
</td>
</tr>
</table>
<div id="mp-other" style="padding-top:4px; padding-bottom:2px;">
<h2><span class="mw-headline" id="Other_areas_of_Wikipedia">Other areas of Wikipedia</span></h2>
<ul>
<li><b><a href="/wiki/Wikipedia:Community_portal" title="Wikipedia:Community portal">Community p
<li><b><a href="/wiki/Wikipedia:Help_desk" title="Wikipedia:Help desk">Help desk</a></b>  Ask qu
<li><b><a href="/wiki/Wikipedia:Local_Embassy" title="Wikipedia:Local Embassy">Local embassy</a>
<li><b><a href="/wiki/Wikipedia:Reference_desk" title="Wikipedia:Reference desk">Reference desk<
```

```
<li><b><a href="/wiki/Wikipedia:News" title="Wikipedia:News">Site news</a></b>  Announcements, u
<li><b><a href="/wiki/Wikipedia:Village_pump" title="Wikipedia:Village pump">Village pump</a></b
</ul>
</div>
<div id="mp-sister">
<h2><span class="mw-headline" id="Wikipedia.27s_sister_projects">Wikipedia's sister projects</sp
<p>Wikipedia is hosted by the <a href="/wiki/Wikimedia_Foundation" title="Wikimedia Foundation">
<table class="layout plainlinks" style="width:100%; margin:auto; text-align:left; background:tra
<tr>
<td style="text-align:center; padding:4px;"><a href="//commons.wikimedia.org/wiki/" title="Comm
<td style="width:33%; padding:4px;"><b><a class="external text" href="//commons.wikimedia.org/">
Free media repository</td>
<td style="text-align:center; padding:4px;"><a href="//www.mediawiki.org/wiki/" title="MediaWik
<td style="width:33%; padding:4px;"><b><a class="external text" href="//mediawiki.org/">MediaWi
Wiki software development</td>
<td style="text-align:center; padding:4px;"><a href="//meta.wikimedia.org/wiki/" title="Meta-Wi
<td style="width:33%; padding:4px;"><b><a class="external text" href="//meta.wikimedia.org/">Me
Wikimedia project coordination</td>
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikibooks.org/wiki/" title="Wikibooks
<td style="padding:4px;"><b><a class="external text" href="//en.wikibooks.org/">Wikibooks</a></
Free textbooks and manuals</td>
<td style="text-align:center; padding:3px;"><a href="//www.wikidata.org/wiki/" title="Wikidata"
<td style="padding:4px;"><b><a class="external text" href="//www.wikidata.org/">Wikidata</a></b
Free knowledge base</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikinews.org/wiki/" title="Wikinews">
<td style="padding:4px;"><b><a class="external text" href="//en.wikinews.org/">Wikinews</a></b>
Free-content news</td>
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikiquote.org/wiki/" title="Wikiquote
<td style="padding:4px;"><b><a class="external text" href="//en.wikiquote.org/">Wikiquote</a></
Collection of quotations</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikisource.org/wiki/" title="Wikisour
<td style="padding:4px;"><b><a class="external text" href="//en.wikisource.org/">Wikisource</a>
Free-content library</td>
<td style="text-align:center; padding:4px;"><a href="//species.wikimedia.org/wiki/" title="Wiki
<td style="padding:4px;"><b><a class="external text" href="//species.wikimedia.org/">Wikispecie
Directory of species</td>
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikiversity.org/wiki/" title="Wikivers
<td style="padding:4px;"><b><a class="external text" href="//en.wikiversity.org/">Wikiversity</a
Free learning materials and activities</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikivoyage.org/wiki/" title="Wikivoyag
<td style="padding:4px;"><b><a class="external text" href="//en.wikivoyage.org/">Wikivoyage</a>
Free travel guide</td>
<td style="text-align:center; padding:4px;"><a href="//en.wiktionary.org/wiki/" title="Wiktionar
<td style="padding:4px;"><b><a class="external text" href="//en.wiktionary.org/">Wiktionary</a>
Dictionary and thesaurus</td>
</tr>
</table>
</div>
```

```html
<div id="mp-lang">
<h2><span class="mw-headline" id="Wikipedia_languages">Wikipedia languages</span></h2>
<div class="nowraplinks nourlexpansion plainlinks" id="lang">
<p>This Wikipedia is written in <a href="/wiki/English_language" title="English language">Englis
<ul>
<li id="lang-3">More than 1,000,000 articles:
<div class="hlist inline">
<ul>
<li><a class="external text" href="//de.wikipedia.org/wiki/"><span class="autonym" lang="de" ti
<li><a class="external text" href="//es.wikipedia.org/wiki/"><span class="autonym" lang="es" ti
<li><a class="external text" href="//fr.wikipedia.org/wiki/"><span class="autonym" lang="fr" ti
<li><a class="external text" href="//it.wikipedia.org/wiki/"><span class="autonym" lang="it" ti
<li><a class="external text" href="//nl.wikipedia.org/wiki/"><span class="autonym" lang="nl" ti
<li><a class="external text" href="//ja.wikipedia.org/wiki/"><span class="autonym" lang="ja" ti
<li><a class="external text" href="//pl.wikipedia.org/wiki/"><span class="autonym" lang="pl" ti
<li><a class="external text" href="//ru.wikipedia.org/wiki/"><span class="autonym" lang="ru" ti
<li><a class="external text" href="//sv.wikipedia.org/wiki/"><span class="autonym" lang="sv" ti
<li><a class="external text" href="//vi.wikipedia.org/wiki/"><span class="autonym" lang="vi" ti
</ul>
</div>
</li>
<li id="lang-2">More than 250,000 articles:
<div class="hlist inline">
<ul>
<li><a class="external text" href="//ar.wikipedia.org/wiki/"><span class="autonym" lang="ar" ti
<li><a class="external text" href="//id.wikipedia.org/wiki/"><span class="autonym" lang="id" ti
<li><a class="external text" href="//ms.wikipedia.org/wiki/"><span class="autonym" lang="ms" ti
<li><a class="external text" href="//ca.wikipedia.org/wiki/"><span class="autonym" lang="ca" ti
<li><a class="external text" href="//cs.wikipedia.org/wiki/"><span class="autonym" lang="cs" ti
<li><a class="external text" href="//fa.wikipedia.org/wiki/"><span class="autonym" lang="fa" ti
<li><a class="external text" href="//ko.wikipedia.org/wiki/"><span class="autonym" lang="ko" ti
<li><a class="external text" href="//hu.wikipedia.org/wiki/"><span class="autonym" lang="hu" ti
<li><a class="external text" href="//no.wikipedia.org/wiki/"><span class="autonym" lang="no" ti
<li><a class="external text" href="//pt.wikipedia.org/wiki/"><span class="autonym" lang="pt" ti
<li><a class="external text" href="//ro.wikipedia.org/wiki/"><span class="autonym" lang="ro" ti
<li><a class="external text" href="//sr.wikipedia.org/wiki/"><span class="autonym" lang="sr" ti
<li><a class="external text" href="//sh.wikipedia.org/wiki/"><span class="autonym" lang="sh" ti
<li><a class="external text" href="//fi.wikipedia.org/wiki/"><span class="autonym" lang="fi" ti
<li><a class="external text" href="//tr.wikipedia.org/wiki/"><span class="autonym" lang="tr" ti
<li><a class="external text" href="//uk.wikipedia.org/wiki/"><span class="autonym" lang="uk" ti
<li><a class="external text" href="//zh.wikipedia.org/wiki/"><span class="autonym" lang="zh" ti
</ul>
</div>
</li>
<li id="lang-1">More than 50,000 articles:
<div class="hlist inline">
<ul>
<li><a class="external text" href="//bs.wikipedia.org/wiki/"><span class="autonym" lang="bs" ti
<li><a class="external text" href="//bg.wikipedia.org/wiki/"><span class="autonym" lang="bg" ti
<li><a class="external text" href="//da.wikipedia.org/wiki/"><span class="autonym" lang="da" ti
<li><a class="external text" href="//et.wikipedia.org/wiki/"><span class="autonym" lang="et" ti
<li><a class="external text" href="//el.wikipedia.org/wiki/"><span class="autonym" lang="el" ti
<li><a class="external text" href="//simple.wikipedia.org/wiki/"><span class="autonym" lang="sim
<li><a class="external text" href="//eo.wikipedia.org/wiki/"><span class="autonym" lang="eo" ti
```

```
<li><a class="external text" href="//eu.wikipedia.org/wiki/"><span class="autonym" lang="eu" tit
<li><a class="external text" href="//gl.wikipedia.org/wiki/"><span class="autonym" lang="gl" tit
<li><a class="external text" href="//he.wikipedia.org/wiki/"><span class="autonym" lang="he" tit
<li><a class="external text" href="//hr.wikipedia.org/wiki/"><span class="autonym" lang="hr" tit
<li><a class="external text" href="//lv.wikipedia.org/wiki/"><span class="autonym" lang="lv" tit
<li><a class="external text" href="//lt.wikipedia.org/wiki/"><span class="autonym" lang="lt" tit
<li><a class="external text" href="//nn.wikipedia.org/wiki/"><span class="autonym" lang="nn" tit
<li><a class="external text" href="//sk.wikipedia.org/wiki/"><span class="autonym" lang="sk" tit
<li><a class="external text" href="//sl.wikipedia.org/wiki/"><span class="autonym" lang="sl" tit
<li><a class="external text" href="//th.wikipedia.org/wiki/"><span class="autonym" lang="th" tit
</ul>
</div>
</li>
</ul>
</div>
<div class="plainlinks" id="metalink" style="text-align:center;"><b><a class="extiw" href="//me
</div>
<!--
NewPP limit report
Parsed by mw1096
Cached time: 20160315215020
Cache expiry: 3600
Dynamic content: true
CPU time usage: 0.363 seconds
Real time usage: 0.444 seconds
Preprocessor visited node count: 3198/1000000
Preprocessor generated node count: 0/1500000
Postexpand include size: 101448/2097152 bytes
Template argument size: 6661/2097152 bytes
Highest expansion depth: 14/40
Expensive parser function count: 5/500
Lua time usage: 0.112/10.000 seconds
Lua memory usage: 2.14 MB/50 MB
Number of Wikibase entities loaded: 0-->
<!--
Transclusion expansion time report (%,ms,calls,template)
100.00%  328.474      1 - -total
 49.05%  161.113      8 - Template:Main_page_image
 44.60%  146.499      1 - Wikipedia:Main_Page/Tomorrow
 18.27%   60.025      2 - Template:Wikipedia_languages
 17.20%   56.502      2 - Template:In_the_news
 16.99%   55.798      1 - Wikipedia:Today's_featured_article/March_15,_2016
 16.19%   53.167      1 - Template:Did_you_know/Queue/2
 15.83%   51.986      8 - Template:Str_number/trim
 14.65%   48.124      2 - Template:In_the_news/image
 13.74%   45.122     24 - Template:If_empty
-->
<!-- Saved in parser cache with key enwiki:pcache:idhash:15580374-0!*!0!!*!4!* and timestamp 20
 -->
<noscript><img alt="" height="1" src="//en.wikipedia.org/wiki/Special:CentralAutoLogin/start?typ
                                     Retrieved from "<a dir="ltr" href="https://en.w
<div class="catlinks catlinks-allhidden" data-mw="interface" id="catlinks"></div> <div class="v
</div>
</div>
```

```
<div id="mw-navigation">
<h2>Navigation menu</h2>
<div id="mw-head">
<div aria-labelledby="p-personal-label" class="" id="p-personal" role="navigation">
<h3 id="p-personal-label">Personal tools</h3>
<ul>
<li id="pt-anonuserpage">Not logged in</li><li id="pt-anontalk"><a accesskey="n" href="/wiki/Sp
</div>
<div id="left-navigation">
<div aria-labelledby="p-namespaces-label" class="vectorTabs" id="p-namespaces" role="navigation"
<h3 id="p-namespaces-label">Namespaces</h3>
<ul>
<li class="selected" id="ca-nstab-main"><span><a accesskey="c" href="/wiki/Main_Page" title="Vie
<li id="ca-talk"><span><a accesskey="t" href="/wiki/Talk:Main_Page" rel="discussion" title="Disc
</ul>
</div>
<div aria-labelledby="p-variants-label" class="vectorMenu emptyPortlet" id="p-variants" role="na
<h3 id="p-variants-label">
<span>Variants</span><a href="#"></a>
</h3>
<div class="menu">
<ul>
</ul>
</div>
</div>
</div>
<div id="right-navigation">
<div aria-labelledby="p-views-label" class="vectorTabs" id="p-views" role="navigation">
<h3 id="p-views-label">Views</h3>
<ul>
<li class="selected" id="ca-view"><span><a href="/wiki/Main_Page">Read</a></span></li>
<li id="ca-viewsource"><span><a accesskey="e" href="/w/index.php?title=Main_Page&amp;action=edit
You can view its source [e]">View source</a></span></li>
<li class="collapsible" id="ca-history"><span><a accesskey="h" href="/w/index.php?title=Main_Pag
</ul>
</div>
<div aria-labelledby="p-cactions-label" class="vectorMenu emptyPortlet" id="p-cactions" role="na
<h3 id="p-cactions-label"><span>More</span><a href="#"></a></h3>
<div class="menu">
<ul>
</ul>
</div>
</div>
<div id="p-search" role="search">
<h3>
<label for="searchInput">Search</label>
</h3>
<form action="/w/index.php" id="searchform">
<div id="simpleSearch">
<input accesskey="f" id="searchInput" name="search" placeholder="Search" title="Search Wikipedia
</form>
</div>
</div>
</div>
```

```
<div id="mw-panel">
<div id="p-logo" role="banner"><a class="mw-wiki-logo" href="/wiki/Main Page" title="Visit the m
<div aria-labelledby="p-navigation-label" class="portal" id="p-navigation" role="navigation">
<h3 id="p-navigation-label">Navigation</h3>
<div class="body">
<ul>
<li id="n-mainpage-description"><a accesskey="z" href="/wiki/Main Page" title="Visit the main pa
</div>
</div>
<div aria-labelledby="p-interaction-label" class="portal" id="p-interaction" role="navigation">
<h3 id="p-interaction-label">Interaction</h3>
<div class="body">
<ul>
<li id="n-help"><a href="/wiki/Help:Contents" title="Guidance on how to use and edit Wikipedia">
</div>
</div>
<div aria-labelledby="p-tb-label" class="portal" id="p-tb" role="navigation">
<h3 id="p-tb-label">Tools</h3>
<div class="body">
<ul>
<li id="t-whatlinkshere"><a accesskey="j" href="/wiki/Special:WhatLinksHere/Main Page" title="Li
</div>
</div>
<div aria-labelledby="p-coll-print_export-label" class="portal" id="p-coll-print_export" role="n
<h3 id="p-coll-print_export-label">Print/export</h3>
<div class="body">
<ul>
<li id="coll-create_a_book"><a href="/w/index.php?title=Special:Book&amp;bookcmd=book_creator&am
</div>
</div>
<div aria-labelledby="p-wikibase-otherprojects-label" class="portal" id="p-wikibase-otherproject
<h3 id="p-wikibase-otherprojects-label">In other projects</h3>
<div class="body">
<ul>
<li class="wb-otherproject-link wb-otherproject-commons"><a href="https://commons.wikimedia.org/
</div>
</div>
<div aria-labelledby="p-lang-label" class="portal" id="p-lang" role="navigation">
<h3 id="p-lang-label">Languages</h3>
<div class="body">
<ul>
<li class="interlanguage-link interwiki-simple"><a href="//simple.wikipedia.org/wiki/" hreflang=
</div>
</div>
</div>
</div>
<div id="footer" role="contentinfo">
<ul id="footer-info">
<li id="footer-info-lastmod"> This page was last modified on 26 December 2015, at 10:03.</li>
<li id="footer-info-copyright">Text is available under the <a href="//en.wikipedia.org/wiki/Wik
additional terms may apply.  By using this site, you agree to the <a href="//wikimediafoundatio
</ul>
<ul id="footer-places">
<li id="footer-places-privacy"><a href="//wikimediafoundation.org/wiki/Privacy policy" title="wm
```

```
<li id="footer-places-about"><a href="/wiki/Wikipedia:About" title="Wikipedia:About">About Wiki
<li id="footer-places-disclaimer"><a href="/wiki/Wikipedia:General_disclaimer" title="Wikipedia:
<li id="footer-places-contact"><a href="//en.wikipedia.org/wiki/Wikipedia:Contact_us">Contact Wi
<li id="footer-places-developers"><a href="https://www.mediawiki.org/wiki/Special:MyLanguage/Ho
<li id="footer-places-cookiestatement"><a href="//wikimediafoundation.org/wiki/Cookie_statement'
<li id="footer-places-mobileview"><a class="noprint stopMobileRedirectToggle" href="//en.m.wiki
</ul>
<ul class="noprint" id="footer-icons">
<li id="footer-copyrightico">
<a href="//wikimediafoundation.org/"><img alt="Wikimedia Foundation" height="31" src="/static/i
<li id="footer-poweredbyico">
<a href="//www.mediawiki.org/"><img alt="Powered by MediaWiki" height="31" src="/w/resources/as
</ul>
<div style="clear:both"></div>
</div>
<script>(window.RLQ = window.RLQ || []).push(function () {
mw.loader.state({"ext.globalCssJs.site":"ready","ext.globalCssJs.user":"ready","user":"ready","u
} );</script><script>(window.RLQ = window.RLQ || []).push(function () {
mw.config.set({"wgBackendResponseTime":61,"wgHostname":"mw1250"}); /* @nomin */
} );</script>
</body>
</html>
```

Beautiful Soup creates a linked tree, where the root of the tree is the whole HTML document. It has children, which are all the elements of the HTML document. Each of those has children, which are any elements they have. Each element of the tree is aware of its parent and children.

You probably don't want to iterate through each child of the whole HTML document - you want a specific thing or things in it. In some cases, you want to seach for html tags. Common tages include:

tag	function
<title>	The title of the web page (shows up in your browser header)
<meta>	Information about the web page that is not shown to the user
<a>	Links to other web pages
<p>	Paragraph of text

In other cases, you want to look for IDs. These are optional information added to a tag to help developers or other code on the web page know which tag is for which purpose. Unlike tags, these are not standardized, so they will change from site to site and author to author. They will look something like:

    <div id="banner" class="MyBanner">

With the advent of CSS (__C__ascading __S__tyle __S__heets), it is also common for people to define their own HTML styling tags. So, while things like lists (<ol>) and tables (<table>, <tr>, and <td>) are in the HTML specification, it's not safe to assume they'll be used when you expect.

As a general strategy, when web scraping, you should have the page you want to scrape open in a browser with either the Developer Tools window open, or the HTML source displayed.

We can pull out elements by tag with:

In [9]: page.p

Out[9]: <p><b><a href="/wiki/California_State_Route_78" title="California State Route 78">State Route 78

This is grabbing the paragraph tag from the page. If we want the first link from the first paragraph, we can try:

In [10]: page.p.a

17

Out[10]: State Route 78

But what if we want all the links? We are going to use a method of bs4's elements called `find_all`.

In [11]: page.p.findAll('a')

Out[11]: [State Route 78
          state highway,
          California,
          Oceanside,
          Blythe,
          Arizona,
          <a class="mw-redirect" href="/wiki/Interstate_5_(California)" title="Interstate 5 (California)
          <a href="/wiki/San_Diego_County,_California" title="San Diego County, California">San Diego Co
          <a class="mw-redirect" href="/wiki/Interstate_10_(California)" title="Interstate 10 (Californi
          <a href="/wiki/Riverside_County,_California" title="Riverside County, California">Riverside Co
          Cuyamaca Mountains,
          Julian,
          <a href="/wiki/Imperial_County,_California" title="Imperial County, California">Imperial Count
          Salton Sea,
          Brawley,
          <a class="mw-redirect" href="/wiki/San_Diego_North_County,_California" title="San Diego North C
          San Diego,
          Escondido,
          <a href="/wiki/California_State_Route_78" title="California State Route 78"><b>Fullarticle...<

What if you want all the elements in that paragraph, and not just the links? bs4 has an iterator for children:

In [12]: for element in page.p.children:
             print(element)

```
<b><a href="/wiki/California_State_Route_78" title="California State Route 78">State Route 78</a></b>
 is a
<a href="/wiki/State_highway" title="State highway">state highway</a>
 in
<a href="/wiki/California" title="California">California</a>
 that runs from
<a href="/wiki/Oceanside,_California" title="Oceanside, California">Oceanside</a>
 east to
<a href="/wiki/Blythe,_California" title="Blythe, California">Blythe</a>
, a few miles from
<a href="/wiki/Arizona" title="Arizona">Arizona</a>
. Its western terminus is at
<a class="mw-redirect" href="/wiki/Interstate_5_(California)" title="Interstate 5 (California)">Intersta
 in
<a href="/wiki/San_Diego_County,_California" title="San Diego County, California">San Diego County</a>
 and its eastern terminus is at
<a class="mw-redirect" href="/wiki/Interstate_10_(California)" title="Interstate 10 (California)">Inters
 in
<a href="/wiki/Riverside_County,_California" title="Riverside County, California">Riverside County</a>
. The route is a freeway through the heavily populated cities of northern San Diego County and a two-lan
<a href="/wiki/Cuyamaca_Mountains" title="Cuyamaca Mountains">Cuyamaca Mountains</a>
 to
<a href="/wiki/Julian,_California" title="Julian, California">Julian</a>
```

```
. In
<a href="/wiki/Imperial_County,_California" title="Imperial County, California">Imperial County</a>
, it travels through the desert near the
<a href="/wiki/Salton_Sea" title="Salton Sea">Salton Sea</a>
 and passes through the city of
<a href="/wiki/Brawley,_California" title="Brawley, California">Brawley</a>
 before turning north into an area of sand dunes on the way to its terminus in Blythe. Portions of the 
<a class="mw-redirect" href="/wiki/San_Diego_North_County,_California" title="San Diego North County, Cal
 of
<a href="/wiki/San_Diego" title="San Diego">San Diego</a>
 that connects Oceanside and
<a href="/wiki/Escondido,_California" title="Escondido, California">Escondido</a>
 was built in the middle of the 20th century in several stages, including a transitory stage known as t
<a href="/wiki/California_State_Route_78" title="California State Route 78"><b>Fullarticle...</b></a>
)
```

HTML elements can be nested, but children only iterates at one level below the element. If you want everything, you can iterate with `descendants`

```
In [13]: for element in page.p.descendants:
             print(element)

<b><a href="/wiki/California_State_Route_78" title="California State Route 78">State Route 78</a></b>
<a href="/wiki/California_State_Route_78" title="California State Route 78">State Route 78</a>
State Route 78
 is a
<a href="/wiki/State_highway" title="State highway">state highway</a>
state highway
 in
<a href="/wiki/California" title="California">California</a>
California
 that runs from
<a href="/wiki/Oceanside,_California" title="Oceanside, California">Oceanside</a>
Oceanside
 east to
<a href="/wiki/Blythe,_California" title="Blythe, California">Blythe</a>
Blythe
, a few miles from
<a href="/wiki/Arizona" title="Arizona">Arizona</a>
Arizona
. Its western terminus is at
<a class="mw-redirect" href="/wiki/Interstate_5_(California)" title="Interstate 5 (California)">Intersta
Interstate 5
 in
<a href="/wiki/San_Diego_County,_California" title="San Diego County, California">San Diego County</a>
San Diego County
 and its eastern terminus is at
<a class="mw-redirect" href="/wiki/Interstate_10_(California)" title="Interstate 10 (California)">Inters
Interstate 10
 in
<a href="/wiki/Riverside_County,_California" title="Riverside County, California">Riverside County</a>
Riverside County
. The route is a freeway through the heavily populated cities of northern San Diego County and a two-la
<a href="/wiki/Cuyamaca_Mountains" title="Cuyamaca Mountains">Cuyamaca Mountains</a>
Cuyamaca Mountains
```

```
 to
<a href="/wiki/Julian,_California" title="Julian, California">Julian</a>
Julian
. In
<a href="/wiki/Imperial_County,_California" title="Imperial County, California">Imperial County</a>
Imperial County
, it travels through the desert near the
<a href="/wiki/Salton_Sea" title="Salton Sea">Salton Sea</a>
Salton Sea
 and passes through the city of
<a href="/wiki/Brawley,_California" title="Brawley, California">Brawley</a>
Brawley
 before turning north into an area of sand dunes on the way to its terminus in Blythe. Portions of the :
<a class="mw-redirect" href="/wiki/San_Diego_North_County,_California" title="San Diego North County, Cal:
North County
 of
<a href="/wiki/San_Diego" title="San Diego">San Diego</a>
San Diego
 that connects Oceanside and
<a href="/wiki/Escondido,_California" title="Escondido, California">Escondido</a>
Escondido
 was built in the middle of the 20th century in several stages, including a transitory stage known as tl
<a href="/wiki/California_State_Route_78" title="California State Route 78"><b>Fullarticle...</b></a>
<b>Fullarticle...</b>
Fullarticle...
)
```

This splits out formatting tags that we *probably* don't care about, like bold-faced text, and so we probably won't use it again.

In reality, you won't be inspecting things yourself, so you'll want to get in the habit of using your knowledge from day 2 about looping and control structures to make decisions for you. For example, what if we wanted to look at every link in the page, then print it's neighbor but only if the link is not to a media file? We could do something like:

```
In [14]: for link in page.find_all('a'):
             if link.attrs.get('class') != 'mw-redirect':
                 print(link.find_next())

<div id="siteNotice"><!-- CentralNotice --></div>
<a href="#p-search">search</a>
<div class="mw-content-ltr" dir="ltr" id="mw-content-text" lang="en"><table id="mp-topbanner" style="wic
<tr>
<td style="width:61%; color:#000;">
<table style="width:280px; border:none; background:none;">
<tr>
<td style="width:280px; text-align:center; white-space:nowrap; color:#000;">
<div style="font-size:162%; border:none; margin:0; padding:.1em; color:#000;">Welcome to <a href="/wiki,
<div style="top:+0.2em; font-size:95%;">the <a href="/wiki/Free_content" title="Free content">free</a> <
<div id="articlecount" style="font-size:85%;"><a href="/wiki/Special:Statistics" title="Special:Statist:
</td>
</tr>
</table>
</td>
<td style="width:13%; font-size:95%;">
<ul>
```

```html
<li><a href="/wiki/Portal:Arts" title="Portal:Arts">Arts</a></li>
<li><a href="/wiki/Portal:Biography" title="Portal:Biography">Biography</a></li>
<li><a href="/wiki/Portal:Geography" title="Portal:Geography">Geography</a></li>
</ul>
</td>
<td style="width:13%; font-size:95%;">
<ul>
<li><a href="/wiki/Portal:History" title="Portal:History">History</a></li>
<li><a href="/wiki/Portal:Mathematics" title="Portal:Mathematics">Mathematics</a></li>
<li><a href="/wiki/Portal:Science" title="Portal:Science">Science</a></li>
</ul>
</td>
<td style="width:13%; font-size:95%;">
<ul>
<li><a href="/wiki/Portal:Society" title="Portal:Society">Society</a></li>
<li><a href="/wiki/Portal:Technology" title="Portal:Technology">Technology</a></li>
<li><b><a href="/wiki/Portal:Contents/Portals" title="Portal:Contents/Portals">All portals</a></b></li>
</ul>
</td>
</tr>
</table>
<table id="mp-upper" style="width: 100%; margin:4px 0 0 0; background:none; border-spacing: 0px;">
<tr>
<td class="MainPageBG" style="width:55%; border:1px solid #cef2e0; background:#f5fffa; vertical-align:t
<table id="mp-left" style="width:100%; vertical-align:top; background:#f5fffa;">
<tr>
<td style="padding:2px;">
<h2 id="mp-tfa-h2" style="margin:3px; background:#cef2e0; font-family:inherit; font-size:120%; font-wei
</td>
</tr>
<tr>
<td style="color:#000;">
<div id="mp-tfa" style="padding:2px 5px">
<div id="mp-tfa-img" style="float: left; margin: 0.5em 0.9em 0.4em 0em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-width: 1
</div>
<p><b><a href="/wiki/California_State_Route_78" title="California State Route 78">State Route 78</a></b>
<ul style="list-style:none; margin-left:0; text-align:right;">
<li>Recently featured:
<div class="hlist inline">
<ul>
<li><i><a href="/wiki/Sarcoscypha_coccinea" title="Sarcoscypha coccinea">Sarcoscypha coccinea</a></i></l
<li><a href="/wiki/Japanese_battleship_Asahi" title="Japanese battleship Asahi">Japanese battleship <i>A
<li><a href="/wiki/Isabella_Beeton" title="Isabella Beeton">Isabella Beeton</a></li>
</ul>
</div>
</li>
</ul>
<div class="hlist noprint" id="mp-tfa-footer" style="text-align: right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Today%27s_featured_article/March_2016" title="Wikipedia:Today's featured
<li><b><a class="extiw" href="https://lists.wikimedia.org/mailman/listinfo/daily-article-l" title="mail
<li><b><a href="/wiki/Wikipedia:Featured_articles" title="Wikipedia:Featured articles">More featured art
</ul>
```

```
</div>
</div>
</td>
</tr>
<tr>
<td style="padding:2px;">
<h2 id="mp-dyk-h2" style="margin:3px; background:#cef2e0; font-family:inherit; font-size:120%; font-wei
</td>
</tr>
<tr>
<td style="color:#000; padding:2px 5px 5px;">
<div id="mp-dyk">
<div id="mp-dyk-img" style="float:right; margin-left:0.5em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-width: 1
<div class="thumbcaption" style="padding: 0.25em 0; word-wrap: break-word;">Bilikiss Adebiyi</div>
</div>
</div>
<ul>
<li>... that <b><a href="/wiki/Bilikiss_Adebiyi_Abiola" title="Bilikiss Adebiyi Abiola">Bilikiss Adebiyi
<li>... that the <a href="/wiki/BBC" title="BBC">BBC</a> re-launched its former television channel <a h
<li>... that the Sanskrit text <b><a href="/wiki/Manasollasa" title="Manasollasa">Manasollasa</a></b> i
<li>... that the species name for <i><b><a href="/wiki/Burmaleon" title="Burmaleon">Burmaleon magnificu
<li>... that the documentary film <i><b><a href="/wiki/No_Land%27s_Song" title="No Land's Song">No Land'
<li>... that uninjured reporters commandeered a <a href="/wiki/Medical_evacuation" title="Medical evacua
<li>... that of an estimated 100,000 <b><a href="/wiki/German_Jewish_military_personnel_of_World_War_I" t
</ul>
<p><b>Correction</b>: we erroneously claimed here that in 1964 <a href="/wiki/Jim_Hazelton" title="Jim H
<div class="hlist noprint" id="mp-dyk-footer" style="text-align:right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Recent_additions" title="Wikipedia:Recent additions">Recently improved a
<li><b><a href="/wiki/Wikipedia:Your_first_article" title="Wikipedia:Your first article">Start a new art
<li><b><a href="/wiki/Template_talk:Did_you_know" title="Template talk:Did you know">Nominate an article
</ul>
</div>
</div>
</td>
</tr>
</table>
</td>
<td style="border:1px solid transparent;"></td>
<td class="MainPageBG" style="width:45%; border:1px solid #cedff2; background:#f5faff; vertical-align:to
<table id="mp-right" style="width:100%; vertical-align:top; background:#f5faff;">
<tr>
<td style="padding:2px;">
<h2 id="mp-itn-h2" style="margin:3px; background:#cedff2; font-family:inherit; font-size:120%; font-wei
</td>
</tr>
<tr>
<td style="color:#000; padding:2px 5px;">
<div id="mp-itn">
<div id="mp-itn-img" style="float:right;margin-left:0.5em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-width: 1
<div class="thumbcaption" style="padding: 0.25em 0; word-wrap: break-word;">Total solar eclipse, viewed
</div>
```

```
</div>
<ul>
<li><b><a href="/wiki/March_2016_Ankara_bombing" title="March 2016 Ankara bombing">An explosion</a></b>
<li>At least 18 people are killed in <b><a href="/wiki/2016_Grand-Bassam_shootings" title="2016 Grand-Ba
<li><a href="/wiki/Google_DeepMind" title="Google DeepMind">Google DeepMind</a>'s <a href="/wiki/AlphaGc
<li>A total <a href="/wiki/Solar_eclipse" title="Solar eclipse">solar eclipse</a> <b><a href="/wiki/Sola
<li>In the <b><a href="/wiki/Slovak_parliamentary_election,_2016" title="Slovak parliamentary election,_
<li>The <a href="/wiki/Human_Rights_Protection_Party" title="Human Rights Protection Party">Human Rights
</ul>
<ul style="list-style:none; margin-left:0;">
<li><b><a href="/wiki/Portal:Current_events" title="Portal:Current events">Ongoing events</a></b>:
<div class="hlist inline">
<ul>
<li><a href="/wiki/Zika_virus_outbreak_(2015%E2%80%93present)" title="Zika virus outbreak (2015present)"
<li><a href="/wiki/European_migrant_crisis" title="European migrant crisis">European migrant crisis</a><
</ul>
</div>
</li>
<li><b><a href="/wiki/Deaths_in_2016" title="Deaths in 2016">Recent deaths</a></b>:
<div class="hlist inline">
<ul>
<li><a href="/wiki/Hilary_Putnam" title="Hilary Putnam">Hilary Putnam</a></li>
<li><a href="/wiki/Lloyd_Shapley" title="Lloyd Shapley">Lloyd Shapley</a></li>
<li><a href="/wiki/Iolanda_Bala%C8%99" title="Iolanda Bala">Iolanda Bala</a></li>
</ul>
</div>
</li>
</ul>
</div>
</td>
</tr>
<tr>
<td style="padding:2px;">
<h2 id="mp-otd-h2" style="margin:3px; background:#cedff2; font-family:inherit; font-size:120%; font-weig
</td>
</tr>
<tr>
<td style="color:#000; padding:2px 5px 5px;">
<div id="mp-otd">
<p><b><a href="/wiki/March_15" title="March 15">March 15</a></b>: <b><a href="/wiki/Ides_of_March" title
<div id="mp-otd-img" style="float:right;margin-left:0.5em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-width: 10
<div class="thumbcaption" style="padding: 0.25em 0; word-wrap: break-word;">Pancho Villa</div>
</div>
</div>
<ul>
<li><a href="/wiki/1783" title="1783">1783</a>  A <b><a href="/wiki/Newburgh_Conspiracy" title="Newburgh
<li><a href="/wiki/1892" title="1892">1892</a>  <b><a href="/wiki/Liverpool_F.C." title="Liverpool F.C."
<li><a href="/wiki/1916" title="1916">1916</a>  Six days after <a href="/wiki/Pancho_Villa" title="Panch
<li><a href="/wiki/1941" title="1941">1941</a>  <b><a href="/wiki/Philippine_Airlines" title="Philippine
<li><a href="/wiki/2011" title="2011">2011</a>  <a href="/wiki/Arab_Spring" title="Arab Spring">Arab Spr
</ul>
<ul style="list-style:none; margin-left:0;">
<li>More anniversaries:
```

```
<div class="hlist inline nowraplinks">
<ul>
<li><a href="/wiki/March_14" title="March 14">March 14</a></li>
<li><b><a href="/wiki/March_15" title="March 15">March 15</a></b></li>
<li><a href="/wiki/March_16" title="March 16">March 16</a></li>
</ul>
</div>
</li>
</ul>
<div class="hlist noprint" id="mp-otd-footer" style="text-align: right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Selected_anniversaries/March" title="Wikipedia:Selected anniversaries/Ma
<li><b><a class="extiw" href="https://lists.wikimedia.org/mailman/listinfo/daily-article-l" title="mail
<li><b><a href="/wiki/List_of_historical_anniversaries" title="List of historical anniversaries">List of
</ul>
<div style="font-size:smaller;">
<ul>
<li>Current date: <span class="nowrap">March 15, 2016</span> (<a href="/wiki/Coordinated_Universal_Time"
<li><span class="plainlinks" id="otd-purgelink"><span class="nowrap"><a class="external text" href="//en
</ul>
</div>
</div>
</div>
</td>
</tr>
</table>
</td>
</tr>
</table>
<table id="mp-lower" style="margin:4px 0 0 0; width:100%; background:none; border-spacing: 0px;">
<tr>
<td class="MainPageBG" style="width:100%; border:1px solid #ddcef2; background:#faf5ff; vertical-align:
<table id="mp-bottom" style="width:100%; vertical-align:top; background:#faf5ff; color:#000;">
<tr>
<td style="padding:2px;">
<h2 id="mp-tfp-h2" style="margin:3px; background:#ddcef2; font-family:inherit; font-size:120%; font-wei
</td>
</tr>
<tr>
<td style="color:#000; padding:2px;">
<div id="mp-tfp">
<table style="margin:0 3px 3px; width:100%; text-align:left; background-color:transparent; border-colla
<tr>
<td style="padding:0 0.9em 0 0;"><a class="image" href="/wiki/File:Ash_in_Yogyakarta_during_the_2014_erup
<td style="padding:0 6px 0 0">
<p>A man sweeping <a href="/wiki/Volcanic_ash" title="Volcanic ash">volcanic ash</a> in <a href="/wiki/Y
<p><small>Photograph: <a href="/wiki/User:Crisco_1492" title="User:Crisco 1492">Chris Woodrich</a></smal
<ul style="list-style:none; margin-left:0; text-align:right;">
<li>Recently featured:
<div class="hlist inline">
<ul>
<li><a href="/wiki/Template:POTD/2016-03-14" title="Template:POTD/2016-03-14"><i>Homme au bain</i></a></
<li><a href="/wiki/Template:POTD/2016-03-13" title="Template:POTD/2016-03-13">Wagner VI projection</a></
<li><a href="/wiki/Template:POTD/2016-03-12" title="Template:POTD/2016-03-12">Lynx (constellation)</a></
```

```
</ul>
</div>
</li>
</ul>
<div class="hlist noprint" style="text-align:right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Picture_of_the_day/March_2016" title="Wikipedia:Picture of the day/March
<li><b><a href="/wiki/Wikipedia:Featured_pictures" title="Wikipedia:Featured pictures">More featured pic
</ul>
</div>
</td>
</tr>
</table>
</div>
</td>
</tr>
</table>
</td>
</tr>
</table>
<div id="mp-other" style="padding-top:4px; padding-bottom:2px;">
<h2><span class="mw-headline" id="Other_areas_of_Wikipedia">Other areas of Wikipedia</span></h2>
<ul>
<li><b><a href="/wiki/Wikipedia:Community_portal" title="Wikipedia:Community portal">Community portal</a
<li><b><a href="/wiki/Wikipedia:Help_desk" title="Wikipedia:Help desk">Help desk</a></b>  Ask questions
<li><b><a href="/wiki/Wikipedia:Local_Embassy" title="Wikipedia:Local Embassy">Local embassy</a></b>  Fo
<li><b><a href="/wiki/Wikipedia:Reference_desk" title="Wikipedia:Reference desk">Reference desk</a></b>
<li><b><a href="/wiki/Wikipedia:News" title="Wikipedia:News">Site news</a></b>  Announcements, updates,
<li><b><a href="/wiki/Wikipedia:Village_pump" title="Wikipedia:Village pump">Village pump</a></b>  For c
</ul>
</div>
<div id="mp-sister">
<h2><span class="mw-headline" id="Wikipedia.27s_sister_projects">Wikipedia's sister projects</span></h2>
<p>Wikipedia is hosted by the <a href="/wiki/Wikimedia_Foundation" title="Wikimedia Foundation">Wikimedi
<table class="layout plainlinks" style="width:100%; margin:auto; text-align:left; background:transparen
<tr>
<td style="text-align:center; padding:4px;"><a href="//commons.wikimedia.org/wiki/" title="Commons"><img
<td style="width:33%; padding:4px;"><b><a class="external text" href="//commons.wikimedia.org/">Commons
Free media repository</td>
<td style="text-align:center; padding:4px;"><a href="//www.mediawiki.org/wiki/" title="MediaWiki"><img a
<td style="width:33%; padding:4px;"><b><a class="external text" href="//mediawiki.org/">MediaWiki</a></
Wiki software development</td>
<td style="text-align:center; padding:4px;"><a href="//meta.wikimedia.org/wiki/" title="Meta-Wiki"><img
<td style="width:33%; padding:4px;"><b><a class="external text" href="//meta.wikimedia.org/">Meta-Wiki<
Wikimedia project coordination</td>
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikibooks.org/wiki/" title="Wikibooks"><img a
<td style="padding:4px;"><b><a class="external text" href="//en.wikibooks.org/">Wikibooks</a></b><br/>
Free textbooks and manuals</td>
<td style="text-align:center; padding:3px;"><a href="//www.wikidata.org/wiki/" title="Wikidata"><img al
<td style="padding:4px;"><b><a class="external text" href="//www.wikidata.org/">Wikidata</a></b><br/>
Free knowledge base</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikinews.org/wiki/" title="Wikinews"><img alt=
```

```
<td style="padding:4px;"><b><a class="external text" href="//en.wikinews.org/">Wikinews</a></b><br/>
Free-content news</td>
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikiquote.org/wiki/" title="Wikiquote"><img a
<td style="padding:4px;"><b><a class="external text" href="//en.wikiquote.org/">Wikiquote</a></b><br/>
Collection of quotations</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikisource.org/wiki/" title="Wikisource"><img
<td style="padding:4px;"><b><a class="external text" href="//en.wikisource.org/">Wikisource</a></b><br/>
Free-content library</td>
<td style="text-align:center; padding:4px;"><a href="//species.wikimedia.org/wiki/" title="Wikispecies">
<td style="padding:4px;"><b><a class="external text" href="//species.wikimedia.org/">Wikispecies</a></b
Directory of species</td>
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikiversity.org/wiki/" title="Wikiversity"><i
<td style="padding:4px;"><b><a class="external text" href="//en.wikiversity.org/">Wikiversity</a></b><b
Free learning materials and activities</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikivoyage.org/wiki/" title="Wikivoyage"><img
<td style="padding:4px;"><b><a class="external text" href="//en.wikivoyage.org/">Wikivoyage</a></b><br/>
Free travel guide</td>
<td style="text-align:center; padding:4px;"><a href="//en.wiktionary.org/wiki/" title="Wiktionary"><img
<td style="padding:4px;"><b><a class="external text" href="//en.wiktionary.org/">Wiktionary</a></b><br/>
Dictionary and thesaurus</td>
</tr>
</table>
</div>
<div id="mp-lang">
<h2><span class="mw-headline" id="Wikipedia_languages">Wikipedia languages</span></h2>
<div class="nowraplinks nourlexpansion plainlinks" id="lang">
<p>This Wikipedia is written in <a href="/wiki/English_language" title="English language">English</a>. S
<ul>
<li id="lang-3">More than 1,000,000 articles:
<div class="hlist inline">
<ul>
<li><a class="external text" href="//de.wikipedia.org/wiki/"><span class="autonym" lang="de" title="Ger
<li><a class="external text" href="//es.wikipedia.org/wiki/"><span class="autonym" lang="es" title="Spar
<li><a class="external text" href="//fr.wikipedia.org/wiki/"><span class="autonym" lang="fr" title="Frer
<li><a class="external text" href="//it.wikipedia.org/wiki/"><span class="autonym" lang="it" title="Ital
<li><a class="external text" href="//nl.wikipedia.org/wiki/"><span class="autonym" lang="nl" title="Dute
<li><a class="external text" href="//ja.wikipedia.org/wiki/"><span class="autonym" lang="ja" title="Japa
<li><a class="external text" href="//pl.wikipedia.org/wiki/"><span class="autonym" lang="pl" title="Poli
<li><a class="external text" href="//ru.wikipedia.org/wiki/"><span class="autonym" lang="ru" title="Russ
<li><a class="external text" href="//sv.wikipedia.org/wiki/"><span class="autonym" lang="sv" title="Swec
<li><a class="external text" href="//vi.wikipedia.org/wiki/"><span class="autonym" lang="vi" title="Viet
</ul>
</div>
</li>
<li id="lang-2">More than 250,000 articles:
<div class="hlist inline">
<ul>
<li><a class="external text" href="//ar.wikipedia.org/wiki/"><span class="autonym" lang="ar" title="Aral
<li><a class="external text" href="//id.wikipedia.org/wiki/"><span class="autonym" lang="id" title="Indc
<li><a class="external text" href="//ms.wikipedia.org/wiki/"><span class="autonym" lang="ms" title="Mala
```

```
<li><a class="external text" href="//ca.wikipedia.org/wiki/"><span class="autonym" lang="ca" title="Cat
<li><a class="external text" href="//cs.wikipedia.org/wiki/"><span class="autonym" lang="cs" title="Cze
<li><a class="external text" href="//fa.wikipedia.org/wiki/"><span class="autonym" lang="fa" title="Per
<li><a class="external text" href="//ko.wikipedia.org/wiki/"><span class="autonym" lang="ko" title="Kor
<li><a class="external text" href="//hu.wikipedia.org/wiki/"><span class="autonym" lang="hu" title="Hun
<li><a class="external text" href="//no.wikipedia.org/wiki/"><span class="autonym" lang="no" title="Nor
<li><a class="external text" href="//pt.wikipedia.org/wiki/"><span class="autonym" lang="pt" title="Por
<li><a class="external text" href="//ro.wikipedia.org/wiki/"><span class="autonym" lang="ro" title="Rom
<li><a class="external text" href="//sr.wikipedia.org/wiki/"><span class="autonym" lang="sr" title="Ser
<li><a class="external text" href="//sh.wikipedia.org/wiki/"><span class="autonym" lang="sh" title="Ser
<li><a class="external text" href="//fi.wikipedia.org/wiki/"><span class="autonym" lang="fi" title="Fin
<li><a class="external text" href="//tr.wikipedia.org/wiki/"><span class="autonym" lang="tr" title="Tur
<li><a class="external text" href="//uk.wikipedia.org/wiki/"><span class="autonym" lang="uk" title="Ukr
<li><a class="external text" href="//zh.wikipedia.org/wiki/"><span class="autonym" lang="zh" title="Chi
</ul>
</div>
</li>
<li id="lang-1">More than 50,000 articles:
<div class="hlist inline">
<ul>
<li><a class="external text" href="//bs.wikipedia.org/wiki/"><span class="autonym" lang="bs" title="Bos
<li><a class="external text" href="//bg.wikipedia.org/wiki/"><span class="autonym" lang="bg" title="Bul
<li><a class="external text" href="//da.wikipedia.org/wiki/"><span class="autonym" lang="da" title="Dan
<li><a class="external text" href="//et.wikipedia.org/wiki/"><span class="autonym" lang="et" title="Est
<li><a class="external text" href="//el.wikipedia.org/wiki/"><span class="autonym" lang="el" title="Gre
<li><a class="external text" href="//simple.wikipedia.org/wiki/"><span class="autonym" lang="simple" ti
<li><a class="external text" href="//eo.wikipedia.org/wiki/"><span class="autonym" lang="eo" title="Esp
<li><a class="external text" href="//eu.wikipedia.org/wiki/"><span class="autonym" lang="eu" title="Bas
<li><a class="external text" href="//gl.wikipedia.org/wiki/"><span class="autonym" lang="gl" title="Gal
<li><a class="external text" href="//he.wikipedia.org/wiki/"><span class="autonym" lang="he" title="Heb
<li><a class="external text" href="//hr.wikipedia.org/wiki/"><span class="autonym" lang="hr" title="Cro
<li><a class="external text" href="//lv.wikipedia.org/wiki/"><span class="autonym" lang="lv" title="Lat
<li><a class="external text" href="//lt.wikipedia.org/wiki/"><span class="autonym" lang="lt" title="Lit
<li><a class="external text" href="//nn.wikipedia.org/wiki/"><span class="autonym" lang="nn" title="Nor
<li><a class="external text" href="//sk.wikipedia.org/wiki/"><span class="autonym" lang="sk" title="Slo
<li><a class="external text" href="//sl.wikipedia.org/wiki/"><span class="autonym" lang="sl" title="Slo
<li><a class="external text" href="//th.wikipedia.org/wiki/"><span class="autonym" lang="th" title="Tha
</ul>
</div>
</li>
</ul>
</div>
<div class="plainlinks" id="metalink" style="text-align:center;"><b><a class="extiw" href="//meta.wikim
</div>
<!--
NewPP limit report
Parsed by mw1096
Cached time: 20160315215020
Cache expiry: 3600
Dynamic content: true
CPU time usage: 0.363 seconds
Real time usage: 0.444 seconds
Preprocessor visited node count: 3198/1000000
Preprocessor generated node count: 0/1500000
```

Postexpand include size: 101448/2097152 bytes
Template argument size: 6661/2097152 bytes
Highest expansion depth: 14/40
Expensive parser function count: 5/500
Lua time usage: 0.112/10.000 seconds
Lua memory usage: 2.14 MB/50 MB
Number of Wikibase entities loaded: 0-->
<!--
Transclusion expansion time report (%,ms,calls,template)
100.00% 328.474   1 - -total
 49.05% 161.113   8 - Template:Main_page_image
 44.60% 146.499   1 - Wikipedia:Main_Page/Tomorrow
 18.27%   60.025   2 - Template:Wikipedia_languages
 17.20%   56.502   2 - Template:In_the_news
 16.99%   55.798   1 - Wikipedia:Today's_featured_article/March_15,_2016
 16.19%   53.167   1 - Template:Did_you_know/Queue/2
 15.83%   51.986   8 - Template:Str_number/trim
 14.65%   48.124   2 - Template:In_the_news/image
 13.74%   45.122   24 - Template:If_empty
-->
<!-- Saved in parser cache with key enwiki:pcache:idhash:15580374-0!*!0!!*!4!* and timestamp 2016031521!
 -->
<noscript><img alt="" height="1" src="//en.wikipedia.org/wiki/Special:CentralAutoLogin/start?type=1x1"
<div style="top:+0.2em; font-size:95%;">the free <
encyclopedia
anyone can edit
<div id="articlecount" style="font-size:85%;"><a href="/wiki/Special:Statistics" title="Special:Statist
English
<td style="width:13%; font-size:95%;">

Arts
Biography
Geography

</td>
Biography
Geography
<td style="width:13%; font-size:95%;">

History
Mathematics
Science

</td>
Mathematics
Science
<td style="width:13%; font-size:95%;">

Society
Technology
All portals

</td>
Technology

```
<li><b><a href="/wiki/Portal:Contents/Portals" title="Portal:Contents/Portals">All portals</a></b></li>
<table id="mp-upper" style="width: 100%; margin:4px 0 0 0; background:none; border-spacing: 0px;">
<tr>
<td class="MainPageBG" style="width:55%; border:1px solid #cef2e0; background:#f5fffa; vertical-align:t
<table id="mp-left" style="width:100%; vertical-align:top; background:#f5fffa;">
<tr>
<td style="padding:2px;">
<h2 id="mp-tfa-h2" style="margin:3px; background:#cef2e0; font-family:inherit; font-size:120%; font-wei
</td>
</tr>
<tr>
<td style="color:#000;">
<div id="mp-tfa" style="padding:2px 5px">
<div id="mp-tfa-img" style="float: left; margin: 0.5em 0.9em 0.4em 0em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-width: 1
</div>
<p><b><a href="/wiki/California_State_Route_78" title="California State Route 78">State Route 78</a></b>
<ul style="list-style:none; margin-left:0; text-align:right;">
<li>Recently featured:
<div class="hlist inline">
<ul>
<li><i><a href="/wiki/Sarcoscypha_coccinea" title="Sarcoscypha coccinea">Sarcoscypha coccinea</a></i></l
<li><a href="/wiki/Japanese_battleship_Asahi" title="Japanese battleship Asahi">Japanese battleship <i>A
<li><a href="/wiki/Isabella_Beeton" title="Isabella Beeton">Isabella Beeton</a></li>
</ul>
</div>
</li>
</ul>
<div class="hlist noprint" id="mp-tfa-footer" style="text-align: right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Today%27s_featured_article/March_2016" title="Wikipedia:Today's featured
<li><b><a class="extiw" href="https://lists.wikimedia.org/mailman/listinfo/daily-article-l" title="mail
<li><b><a href="/wiki/Wikipedia:Featured_articles" title="Wikipedia:Featured articles">More featured art
</ul>
</div>
</div>
</td>
</tr>
<tr>
<td style="padding:2px;">
<h2 id="mp-dyk-h2" style="margin:3px; background:#cef2e0; font-family:inherit; font-size:120%; font-wei
</td>
</tr>
<tr>
<td style="color:#000; padding:2px 5px 5px;">
<div id="mp-dyk">
<div id="mp-dyk-img" style="float:right; margin-left:0.5em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-width: 1
<div class="thumbcaption" style="padding: 0.25em 0; word-wrap: break-word;">Bilikiss Adebiyi</div>
</div>
</div>
<ul>
<li>... that <b><a href="/wiki/Bilikiss_Adebiyi_Abiola" title="Bilikiss Adebiyi Abiola">Bilikiss Adebiyi
<li>... that the <a href="/wiki/BBC" title="BBC">BBC</a> re-launched its former television channel <a hi
```

```html
<li>... that the Sanskrit text <b><a href="/wiki/Manasollasa" title="Manasollasa">Manasollasa</a></b> is
<li>... that the species name for <i><b><a href="/wiki/Burmaleon" title="Burmaleon">Burmaleon magnificus
<li>... that the documentary film <i><b><a href="/wiki/No_Land%27s_Song" title="No Land's Song">No Land'
<li>... that uninjured reporters commandeered a <a href="/wiki/Medical_evacuation" title="Medical evacua
<li>... that of an estimated 100,000 <b><a href="/wiki/German_Jewish_military_personnel_of_World_War_I" t
</ul>
<p><b>Correction</b>: we erroneously claimed here that in 1964 <a href="/wiki/Jim_Hazelton" title="Jim H
<div class="hlist noprint" id="mp-dyk-footer" style="text-align:right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Recent_additions" title="Wikipedia:Recent additions">Recently improved a
<li><b><a href="/wiki/Wikipedia:Your_first_article" title="Wikipedia:Your first article">Start a new art
<li><b><a href="/wiki/Template_talk:Did_you_know" title="Template talk:Did you know">Nominate an article
</ul>
</div>
</div>
</td>
</tr>
</table>
</td>
<td style="border:1px solid transparent;"></td>
<td class="MainPageBG" style="width:45%; border:1px solid #cedff2; background:#f5faff; vertical-align:to
<table id="mp-right" style="width:100%; vertical-align:top; background:#f5faff;">
<tr>
<td style="padding:2px;">
<h2 id="mp-itn-h2" style="margin:3px; background:#cedff2; font-family:inherit; font-size:120%; font-wei
</td>
</tr>
<tr>
<td style="color:#000; padding:2px 5px;">
<div id="mp-itn">
<div id="mp-itn-img" style="float:right;margin-left:0.5em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-width: 1
<div class="thumbcaption" style="padding: 0.25em 0; word-wrap: break-word;">Total solar eclipse, viewed
</div>
</div>
<ul>
<li><b><a href="/wiki/March_2016_Ankara_bombing" title="March 2016 Ankara bombing">An explosion</a></b>
<li>At least 18 people are killed in <b><a href="/wiki/2016_Grand-Bassam_shootings" title="2016 Grand-Ba
<li><a href="/wiki/Google_DeepMind" title="Google DeepMind">Google DeepMind</a>'s <a href="/wiki/AlphaGo
<li>A total <a href="/wiki/Solar_eclipse" title="Solar eclipse">solar eclipse</a> <b><a href="/wiki/Sola
<li>In the <b><a href="/wiki/Slovak_parliamentary_election,_2016" title="Slovak parliamentary election,
<li>The <a href="/wiki/Human_Rights_Protection_Party" title="Human Rights Protection Party">Human Rights
</ul>
<ul style="list-style:none; margin-left:0;">
<li><b><a href="/wiki/Portal:Current_events" title="Portal:Current events">Ongoing events</a></b>:
<div class="hlist inline">
<ul>
<li><a href="/wiki/Zika_virus_outbreak_(2015%E2%80%93present)" title="Zika virus outbreak (2015present)"
<li><a href="/wiki/European_migrant_crisis" title="European migrant crisis">European migrant crisis</a><
</ul>
</div>
</li>
<li><b><a href="/wiki/Deaths_in_2016" title="Deaths in 2016">Recent deaths</a></b>:
<div class="hlist inline">
```

```
<ul>
<li><a href="/wiki/Hilary_Putnam" title="Hilary Putnam">Hilary Putnam</a></li>
<li><a href="/wiki/Lloyd_Shapley" title="Lloyd Shapley">Lloyd Shapley</a></li>
<li><a href="/wiki/Iolanda_Bala%C8%99" title="Iolanda Bala">Iolanda Bala</a></li>
</ul>
</div>
</li>
</ul>
</div>
</td>
</tr>
<tr>
<td style="padding:2px;">
<h2 id="mp-otd-h2" style="margin:3px; background:#cedff2; font-family:inherit; font-size:120%; font-weig
</td>
</tr>
<tr>
<td style="color:#000; padding:2px 5px 5px;">
<div id="mp-otd">
<p><b><a href="/wiki/March_15" title="March 15">March 15</a></b>: <b><a href="/wiki/Ides_of_March" title
<div id="mp-otd-img" style="float:right;margin-left:0.5em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-width: 10
<div class="thumbcaption" style="padding: 0.25em 0; word-wrap: break-word;">Pancho Villa</div>
</div>
</div>
<ul>
<li><a href="/wiki/1783" title="1783">1783</a>  A <b><a href="/wiki/Newburgh_Conspiracy" title="Newburgh
<li><a href="/wiki/1892" title="1892">1892</a>  <b><a href="/wiki/Liverpool_F.C." title="Liverpool F.C."
<li><a href="/wiki/1916" title="1916">1916</a>  Six days after <a href="/wiki/Pancho_Villa" title="Panch
<li><a href="/wiki/1941" title="1941">1941</a>  <b><a href="/wiki/Philippine_Airlines" title="Philippine
<li><a href="/wiki/2011" title="2011">2011</a>  <a href="/wiki/Arab_Spring" title="Arab Spring">Arab Spr
</ul>
<ul style="list-style:none; margin-left:0;">
<li>More anniversaries:
<div class="hlist inline nowraplinks">
<ul>
<li><a href="/wiki/March_14" title="March 14">March 14</a></li>
<li><b><a href="/wiki/March_15" title="March 15">March 15</a></b></li>
<li><a href="/wiki/March_16" title="March 16">March 16</a></li>
</ul>
</div>
</li>
</ul>
<div class="hlist noprint" id="mp-otd-footer" style="text-align: right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Selected_anniversaries/March" title="Wikipedia:Selected anniversaries/Ma
<li><b><a class="extiw" href="https://lists.wikimedia.org/mailman/listinfo/daily-article-l" title="mail
<li><b><a href="/wiki/List_of_historical_anniversaries" title="List of historical anniversaries">List of
</ul>
<div style="font-size:smaller;">
<ul>
<li>Current date: <span class="nowrap">March 15, 2016</span> (<a href="/wiki/Coordinated_Universal_Time"
<li><span class="plainlinks" id="otd-purgelink"><span class="nowrap"><a class="external text" href="//en
</ul>
```

```
</div>
</div>
</div>
</td>
</tr>
</table>
</td>
</tr>
</table>
<img alt="SR 78 in Oceanside at the El Camino Real overpass" data-file-height="1080" data-file-width="19
<a href="/wiki/State_highway" title="State highway">state highway</a>
<a href="/wiki/California" title="California">California</a>
<a href="/wiki/Oceanside,_California" title="Oceanside, California">Oceanside</a>
<a href="/wiki/Blythe,_California" title="Blythe, California">Blythe</a>
<a href="/wiki/Arizona" title="Arizona">Arizona</a>
<a class="mw-redirect" href="/wiki/Interstate_5_(California)" title="Interstate 5 (California)">Intersta
<a href="/wiki/San_Diego_County,_California" title="San Diego County, California">San Diego County</a>
<a class="mw-redirect" href="/wiki/Interstate_10_(California)" title="Interstate 10 (California)">Inters
<a href="/wiki/Riverside_County,_California" title="Riverside County, California">Riverside County</a>
<a href="/wiki/Cuyamaca_Mountains" title="Cuyamaca Mountains">Cuyamaca Mountains</a>
<a href="/wiki/Julian,_California" title="Julian, California">Julian</a>
<a href="/wiki/Imperial_County,_California" title="Imperial County, California">Imperial County</a>
<a href="/wiki/Salton_Sea" title="Salton Sea">Salton Sea</a>
<a href="/wiki/Brawley,_California" title="Brawley, California">Brawley</a>
<a class="mw-redirect" href="/wiki/San_Diego_North_County,_California" title="San Diego North County, Cal
<a href="/wiki/San_Diego" title="San Diego">San Diego</a>
<a href="/wiki/Escondido,_California" title="Escondido, California">Escondido</a>
<a href="/wiki/California_State_Route_78" title="California State Route 78"><b>Fullarticle...</b></a>
<b>Fullarticle...</b>
<li><a href="/wiki/Japanese_battleship_Asahi" title="Japanese battleship Asahi">Japanese battleship <i>A
<i>Asahi</i>
<div class="hlist noprint" id="mp-tfa-footer" style="text-align: right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Today%27s_featured_article/March_2016" title="Wikipedia:Today's featured
<li><b><a class="extiw" href="https://lists.wikimedia.org/mailman/listinfo/daily-article-l" title="mail
<li><b><a href="/wiki/Wikipedia:Featured_articles" title="Wikipedia:Featured articles">More featured art
</ul>
</div>
<li><b><a class="extiw" href="https://lists.wikimedia.org/mailman/listinfo/daily-article-l" title="mail
<li><b><a href="/wiki/Wikipedia:Featured_articles" title="Wikipedia:Featured articles">More featured art
<tr>
<td style="padding:2px;">
<h2 id="mp-dyk-h2" style="margin:3px; background:#cef2e0; font-family:inherit; font-size:120%; font-wei
</td>
</tr>
<img alt="Bilikiss Adebiyi" data-file-height="500" data-file-width="447" height="133" src="//upload.wik
<i>(pictured)</i>
<a href="/wiki/Massachusetts_Institute_of_Technology" title="Massachusetts Institute of Technology">MIT<
<li>... that the <a href="/wiki/BBC" title="BBC">BBC</a> re-launched its former television channel <a h
<a href="/wiki/BBC_Three_(former)" title="BBC Three (former)">BBC Three</a>
<b><a href="/wiki/BBC_Three_(Internet_television)" title="BBC Three (Internet television)">Internet tele
<li>... that the Sanskrit text <b><a href="/wiki/Manasollasa" title="Manasollasa">Manasollasa</a></b> i
<li>... that the species name for <i><b><a href="/wiki/Burmaleon" title="Burmaleon">Burmaleon magnificus
<li>... that the documentary film <i><b><a href="/wiki/No_Land%27s_Song" title="No Land's Song">No Land'
```

```html
<li>... that uninjured reporters commandeered a <a href="/wiki/Medical_evacuation" title="Medical evacua
<b><a href="/wiki/Campaign_Z" title="Campaign Z">Campaign Z</a></b>
<li>... that of an estimated 100,000 <b><a href="/wiki/German_Jewish_military_personnel_of_World_War_I" t
<a href="/wiki/German_Army_(German_Empire)" title="German Army (German Empire)">German Army</a>
<a href="/wiki/World_War_I" title="World War I">World War I</a>
<p><b>Correction</b>: we erroneously claimed here that in 1964 <a href="/wiki/Jim_Hazelton" title="Jim H
<a href="/wiki/Charles_Kingsford_Smith" title="Charles Kingsford Smith">Charles Kingsford Smith</a>
<a href="/wiki/Gordon_Taylor_(aviator)" title="Gordon Taylor (aviator)">Gordon Taylor</a>
<a href="/wiki/Lockheed_Altair" title="Lockheed Altair">Lockheed Altair</a>
<i><a href="/wiki/Lady_Southern_Cross" title="Lady Southern Cross">Lady Southern Cross</a></i>
<div class="hlist noprint" id="mp-dyk-footer" style="text-align:right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Recent_additions" title="Wikipedia:Recent additions">Recently improved a
<li><b><a href="/wiki/Wikipedia:Your_first_article" title="Wikipedia:Your first article">Start a new art
<li><b><a href="/wiki/Template_talk:Did_you_know" title="Template talk:Did you know">Nominate an article
</ul>
</div>
<li><b><a href="/wiki/Wikipedia:Your_first_article" title="Wikipedia:Your first article">Start a new art
<li><b><a href="/wiki/Template_talk:Did_you_know" title="Template talk:Did you know">Nominate an article
<td style="border:1px solid transparent;"></td>
<img alt="Total solar eclipse, viewed from Balikpapan" data-file-height="388" data-file-width="396" heig
<ul>
<li><b><a href="/wiki/March_2016_Ankara_bombing" title="March 2016 Ankara bombing">An explosion</a></b>
<li>At least 18 people are killed in <b><a href="/wiki/2016_Grand-Bassam_shootings" title="2016 Grand-Ba
<li><a href="/wiki/Google_DeepMind" title="Google DeepMind">Google DeepMind</a>'s <a href="/wiki/AlphaGo
<li>A total <a href="/wiki/Solar_eclipse" title="Solar eclipse">solar eclipse</a> <b><a href="/wiki/Sola
<li>In the <b><a href="/wiki/Slovak_parliamentary_election,_2016" title="Slovak parliamentary election,_
<li>The <a href="/wiki/Human_Rights_Protection_Party" title="Human Rights Protection Party">Human Rights
</ul>
<a href="/wiki/Ankara" title="Ankara">Ankara</a>
<li>At least 18 people are killed in <b><a href="/wiki/2016_Grand-Bassam_shootings" title="2016 Grand-Ba
<a href="/wiki/Grand-Bassam" title="Grand-Bassam">Grand-Bassam</a>
<li><a href="/wiki/Google_DeepMind" title="Google DeepMind">Google DeepMind</a>'s <a href="/wiki/AlphaGo
<a href="/wiki/AlphaGo" title="AlphaGo">AlphaGo</a>
<b><a href="/wiki/AlphaGo_versus_Lee_Sedol" title="AlphaGo versus Lee Sedol">wins a series</a></b>
<a href="/wiki/Lee_Sedol" title="Lee Sedol">Lee Sedol</a>
<a href="/wiki/Go_(game)" title="Go (game)">Go</a>
<li>A total <a href="/wiki/Solar_eclipse" title="Solar eclipse">solar eclipse</a> <b><a href="/wiki/Sola
<b><a href="/wiki/Solar_eclipse_of_March_9,_2016" title="Solar eclipse of March 9, 2016">occurs</a></b>
<i>(pictured)</i>
<a href="/wiki/Direction_%E2%80%93_Social_Democracy" title="Direction  Social Democracy">Direction  Soci
<a href="/wiki/National_Council_(Slovakia)" title="National Council (Slovakia)">National Council</a>
<li>The <a href="/wiki/Human_Rights_Protection_Party" title="Human Rights Protection Party">Human Rights
<a href="/wiki/Tuilaepa_Aiono_Sailele_Malielegaoi" title="Tuilaepa Aiono Sailele Malielegaoi">Tuilaepa A
<b><a href="/wiki/Samoan_general_election,_2016" title="Samoan general election, 2016">Samoan general el
<ul style="list-style:none; margin-left:0;">
<li><b><a href="/wiki/Portal:Current_events" title="Portal:Current events">Ongoing events</a></b>:
<div class="hlist inline">
<ul>
<li><a href="/wiki/Zika_virus_outbreak_(2015%E2%80%93present)" title="Zika virus outbreak (2015present)"
<li><a href="/wiki/European_migrant_crisis" title="European migrant crisis">European migrant crisis</a><
</ul>
</div>
</li>
```

```
<li><b><a href="/wiki/Deaths_in_2016" title="Deaths in 2016">Recent deaths</a></b>:
<div class="hlist inline">
<ul>
<li><a href="/wiki/Hilary_Putnam" title="Hilary Putnam">Hilary Putnam</a></li>
<li><a href="/wiki/Lloyd_Shapley" title="Lloyd Shapley">Lloyd Shapley</a></li>
<li><a href="/wiki/Iolanda_Bala%C8%99" title="Iolanda Bala">Iolanda Bala</a></li>
</ul>
</div>
</li>
</ul>
<div class="hlist inline">
<ul>
<li><a href="/wiki/Zika_virus_outbreak_(2015%E2%80%93present)" title="Zika virus outbreak (2015present)"
<li><a href="/wiki/European_migrant_crisis" title="European migrant crisis">European migrant crisis</a><
</ul>
</div>
<li><a href="/wiki/European_migrant_crisis" title="European migrant crisis">European migrant crisis</a><
<li><b><a href="/wiki/Deaths_in_2016" title="Deaths in 2016">Recent deaths</a></b>:
<div class="hlist inline">
<ul>
<li><a href="/wiki/Hilary_Putnam" title="Hilary Putnam">Hilary Putnam</a></li>
<li><a href="/wiki/Lloyd_Shapley" title="Lloyd Shapley">Lloyd Shapley</a></li>
<li><a href="/wiki/Iolanda_Bala%C8%99" title="Iolanda Bala">Iolanda Bala</a></li>
</ul>
</div>
</li>
<div class="hlist inline">
<ul>
<li><a href="/wiki/Hilary_Putnam" title="Hilary Putnam">Hilary Putnam</a></li>
<li><a href="/wiki/Lloyd_Shapley" title="Lloyd Shapley">Lloyd Shapley</a></li>
<li><a href="/wiki/Iolanda_Bala%C8%99" title="Iolanda Bala">Iolanda Bala</a></li>
</ul>
</div>
<li><a href="/wiki/Lloyd_Shapley" title="Lloyd Shapley">Lloyd Shapley</a></li>
<li><a href="/wiki/Iolanda_Bala%C8%99" title="Iolanda Bala">Iolanda Bala</a></li>
<tr>
<td style="padding:2px;">
<h2 id="mp-otd-h2" style="margin:3px; background:#cedff2; font-family:inherit; font-size:120%; font-weig
</td>
</tr>
<b><a href="/wiki/Ides_of_March" title="Ides of March">Ides of March</a></b>
<b><a href="/wiki/Hungarian_Revolution_of_1848" title="Hungarian Revolution of 1848">National Day</a></b
<a href="/wiki/1848" title="1848">1848</a>
<div id="mp-otd-img" style="float:right;margin-left:0.5em;">
<div class="thumbinner mp-thumb" style="background: transparent; border: none; padding: 0; max-width: 10
<div class="thumbcaption" style="padding: 0.25em 0; word-wrap: break-word;">Pancho Villa</div>
</div>
</div>
<img alt="Pancho Villa" data-file-height="574" data-file-width="431" height="133" src="//upload.wikimedi
<b><a href="/wiki/Newburgh_Conspiracy" title="Newburgh Conspiracy">potential uprising</a></b>
<a href="/wiki/Newburgh_(city),_New_York" title="Newburgh (city), New York">Newburgh, New York</a>
<a href="/wiki/George_Washington" title="George Washington">George Washington</a>
<a href="/wiki/Continental_Army" title="Continental Army">Continental Army</a>
<a href="/wiki/United_States_Congress" title="United States Congress">Congress</a>
```

```html
<li><a href="/wiki/1892" title="1892">1892</a>  <b><a href="/wiki/Liverpool_F.C." title="Liverpool F.C."
<b><a href="/wiki/Liverpool_F.C." title="Liverpool F.C.">Liverpool F.C.</a></b>
<a href="/wiki/Association_football" title="Association football">football</a>
<li><a href="/wiki/1916" title="1916">1916</a>  Six days after <a href="/wiki/Pancho_Villa" title="Panch
<a href="/wiki/Pancho_Villa" title="Pancho Villa">Pancho Villa</a>
<i>(pictured)</i>
<a href="/wiki/John_J._Pershing" title="John J. Pershing">John J. Pershing</a>
<b><a href="/wiki/Pancho_Villa_Expedition" title="Pancho Villa Expedition">punitive expedition into Mexi
<li><a href="/wiki/1941" title="1941">1941</a>  <b><a href="/wiki/Philippine_Airlines" title="Philippine
<b><a href="/wiki/Philippine_Airlines" title="Philippine Airlines">Philippine Airlines</a></b>
<a href="/wiki/Flag_carrier" title="Flag carrier">flag carrier</a>
<li><a href="/wiki/2011" title="2011">2011</a>  <a href="/wiki/Arab_Spring" title="Arab Spring">Arab Spr
<a href="/wiki/Arab_Spring" title="Arab Spring">Arab Spring</a>
<b><a href="/wiki/Syrian_Civil_War" title="Syrian Civil War">across Syria</a></b>
<ul style="list-style:none; margin-left:0;">
<li>More anniversaries:
<div class="hlist inline nowraplinks">
<ul>
<li><a href="/wiki/March_14" title="March 14">March 14</a></li>
<li><b><a href="/wiki/March_15" title="March 15">March 15</a></b></li>
<li><a href="/wiki/March_16" title="March 16">March 16</a></li>
</ul>
</div>
</li>
</ul>
<li><b><a href="/wiki/March_15" title="March 15">March 15</a></b></li>
<li><a href="/wiki/March_16" title="March 16">March 16</a></li>
<div class="hlist noprint" id="mp-otd-footer" style="text-align: right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Selected_anniversaries/March" title="Wikipedia:Selected anniversaries/Ma
<li><b><a class="extiw" href="https://lists.wikimedia.org/mailman/listinfo/daily-article-l" title="mail
<li><b><a href="/wiki/List_of_historical_anniversaries" title="List of historical anniversaries">List of
</ul>
<div style="font-size:smaller;">
<ul>
<li>Current date: <span class="nowrap">March 15, 2016</span> (<a href="/wiki/Coordinated_Universal_Time"
<li><span class="plainlinks" id="otd-purgelink"><span class="nowrap"><a class="external text" href="//er
</ul>
</div>
</div>
<li><b><a class="extiw" href="https://lists.wikimedia.org/mailman/listinfo/daily-article-l" title="mail
<li><b><a href="/wiki/List_of_historical_anniversaries" title="List of historical anniversaries">List of
<div style="font-size:smaller;">
<ul>
<li>Current date: <span class="nowrap">March 15, 2016</span> (<a href="/wiki/Coordinated_Universal_Time"
<li><span class="plainlinks" id="otd-purgelink"><span class="nowrap"><a class="external text" href="//er
</ul>
</div>
<li><span class="plainlinks" id="otd-purgelink"><span class="nowrap"><a class="external text" href="//er
<table id="mp-lower" style="margin:4px 0 0 0; width:100%; background:none; border-spacing: 0px;">
<tr>
<td class="MainPageBG" style="width:100%; border:1px solid #ddcef2; background:#faf5ff; vertical-align:
<table id="mp-bottom" style="width:100%; vertical-align:top; background:#faf5ff; color:#000;">
<tr>
```

```html
<td style="padding:2px;">
<h2 id="mp-tfp-h2" style="margin:3px; background:#ddcef2; font-family:inherit; font-size:120%; font-weig
</td>
</tr>
<tr>
<td style="color:#000; padding:2px;">
<div id="mp-tfp">
<table style="margin:0 3px 3px; width:100%; text-align:left; background-color:transparent; border-colla
<tr>
<td style="padding:0 0.9em 0 0;"><a class="image" href="/wiki/File:Ash_in_Yogyakarta_during_the_2014_erup
<td style="padding:0 6px 0 0">
<p>A man sweeping <a href="/wiki/Volcanic_ash" title="Volcanic ash">volcanic ash</a> in <a href="/wiki/Y
<p><small>Photograph: <a href="/wiki/User:Crisco_1492" title="User:Crisco 1492">Chris Woodrich</a></smal
<ul style="list-style:none; margin-left:0; text-align:right;">
<li>Recently featured:
<div class="hlist inline">
<ul>
<li><a href="/wiki/Template:POTD/2016-03-14" title="Template:POTD/2016-03-14"><i>Homme au bain</i></a></
<li><a href="/wiki/Template:POTD/2016-03-13" title="Template:POTD/2016-03-13">Wagner VI projection</a></
<li><a href="/wiki/Template:POTD/2016-03-12" title="Template:POTD/2016-03-12">Lynx (constellation)</a></
</ul>
</div>
</li>
</ul>
<div class="hlist noprint" style="text-align:right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Picture_of_the_day/March_2016" title="Wikipedia:Picture of the day/March
<li><b><a href="/wiki/Wikipedia:Featured_pictures" title="Wikipedia:Featured pictures">More featured pic
</ul>
</div>
</td>
</tr>
</table>
</div>
</td>
</tr>
</table>
</td>
</tr>
</table>
<img alt="Man sweeping volcanic ash" data-file-height="1524" data-file-width="2246" height="258" src="/
<a href="/wiki/Yogyakarta" title="Yogyakarta">Yogyakarta</a>
<b><a href="/wiki/Kelud#2014_eruption" title="Kelud">2014 eruption</a></b>
<a href="/wiki/Kelud" title="Kelud">Kelud</a>
<a href="/wiki/East_Java" title="East Java">East Javan</a>
<a href="/wiki/Bandung" title="Bandung">Bandung</a>
<p><small>Photograph: <a href="/wiki/User:Crisco_1492" title="User:Crisco 1492">Chris Woodrich</a></smal
<ul style="list-style:none; margin-left:0; text-align:right;">
<li>Recently featured:
<div class="hlist inline">
<ul>
<li><a href="/wiki/Template:POTD/2016-03-14" title="Template:POTD/2016-03-14"><i>Homme au bain</i></a></
<li><a href="/wiki/Template:POTD/2016-03-13" title="Template:POTD/2016-03-13">Wagner VI projection</a></
<li><a href="/wiki/Template:POTD/2016-03-12" title="Template:POTD/2016-03-12">Lynx (constellation)</a></
```

```html
</ul>
</div>
</li>
</ul>
<i>Homme au bain</i>
<li><a href="/wiki/Template:POTD/2016-03-12" title="Template:POTD/2016-03-12">Lynx (constellation)</a><
<div class="hlist noprint" style="text-align:right;">
<ul>
<li><b><a href="/wiki/Wikipedia:Picture_of_the_day/March_2016" title="Wikipedia:Picture of the day/March
<li><b><a href="/wiki/Wikipedia:Featured_pictures" title="Wikipedia:Featured pictures">More featured pic
</ul>
</div>
<li><b><a href="/wiki/Wikipedia:Featured_pictures" title="Wikipedia:Featured pictures">More featured pic
<div id="mp-other" style="padding-top:4px; padding-bottom:2px;">
<h2><span class="mw-headline" id="Other_areas_of_Wikipedia">Other areas of Wikipedia</span></h2>
<ul>
<li><b><a href="/wiki/Wikipedia:Community_portal" title="Wikipedia:Community portal">Community portal</a
<li><b><a href="/wiki/Wikipedia:Help_desk" title="Wikipedia:Help desk">Help desk</a></b>  Ask questions
<li><b><a href="/wiki/Wikipedia:Local_Embassy" title="Wikipedia:Local Embassy">Local embassy</a></b>  Fo
<li><b><a href="/wiki/Wikipedia:Reference_desk" title="Wikipedia:Reference desk">Reference desk</a></b>
<li><b><a href="/wiki/Wikipedia:News" title="Wikipedia:News">Site news</a></b>  Announcements, updates,
<li><b><a href="/wiki/Wikipedia:Village_pump" title="Wikipedia:Village pump">Village pump</a></b>  For c
</ul>
</div>
<li><b><a href="/wiki/Wikipedia:Help_desk" title="Wikipedia:Help desk">Help desk</a></b>  Ask questions
<li><b><a href="/wiki/Wikipedia:Local_Embassy" title="Wikipedia:Local Embassy">Local embassy</a></b>  Fo
<li><b><a href="/wiki/Wikipedia:Reference_desk" title="Wikipedia:Reference desk">Reference desk</a></b>
<li><b><a href="/wiki/Wikipedia:News" title="Wikipedia:News">Site news</a></b>  Announcements, updates,
<li><b><a href="/wiki/Wikipedia:Village_pump" title="Wikipedia:Village pump">Village pump</a></b>  For c
<div id="mp-sister">
<h2><span class="mw-headline" id="Wikipedia.27s_sister_projects">Wikipedia's sister projects</span></h2>
<p>Wikipedia is hosted by the <a href="/wiki/Wikimedia_Foundation" title="Wikimedia Foundation">Wikimedi
<table class="layout plainlinks" style="width:100%; margin:auto; text-align:left; background:transparen
<tr>
<td style="text-align:center; padding:4px;"><a href="//commons.wikimedia.org/wiki/" title="Commons"><img
<td style="width:33%; padding:4px;"><b><a class="external text" href="//commons.wikimedia.org/">Commons-
Free media repository</td>
<td style="text-align:center; padding:4px;"><a href="//www.mediawiki.org/wiki/" title="MediaWiki"><img a
<td style="width:33%; padding:4px;"><b><a class="external text" href="//mediawiki.org/">MediaWiki</a></b
Wiki software development</td>
<td style="text-align:center; padding:4px;"><a href="//meta.wikimedia.org/wiki/" title="Meta-Wiki"><img
<td style="width:33%; padding:4px;"><b><a class="external text" href="//meta.wikimedia.org/">Meta-Wiki</
Wikimedia project coordination</td>
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikibooks.org/wiki/" title="Wikibooks"><img al
<td style="padding:4px;"><b><a class="external text" href="//en.wikibooks.org/">Wikibooks</a></b><br/>
Free textbooks and manuals</td>
<td style="text-align:center; padding:3px;"><a href="//www.wikidata.org/wiki/" title="Wikidata"><img alt
<td style="padding:4px;"><b><a class="external text" href="//www.wikidata.org/">Wikidata</a></b><br/>
Free knowledge base</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikinews.org/wiki/" title="Wikinews"><img alt=
<td style="padding:4px;"><b><a class="external text" href="//en.wikinews.org/">Wikinews</a></b><br/>
Free-content news</td>
```

```
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikiquote.org/wiki/" title="Wikiquote"><img al
<td style="padding:4px;"><b><a class="external text" href="//en.wikiquote.org/">Wikiquote</a></b><br/>
Collection of quotations</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikisource.org/wiki/" title="Wikisource"><img
<td style="padding:4px;"><b><a class="external text" href="//en.wikisource.org/">Wikisource</a></b><br/>
Free-content library</td>
<td style="text-align:center; padding:4px;"><a href="//species.wikimedia.org/wiki/" title="Wikispecies">
<td style="padding:4px;"><b><a class="external text" href="//species.wikimedia.org/">Wikispecies</a></b
Directory of species</td>
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikiversity.org/wiki/" title="Wikiversity"><in
<td style="padding:4px;"><b><a class="external text" href="//en.wikiversity.org/">Wikiversity</a></b><bi
Free learning materials and activities</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikivoyage.org/wiki/" title="Wikivoyage"><img
<td style="padding:4px;"><b><a class="external text" href="//en.wikivoyage.org/">Wikivoyage</a></b><br/:
Free travel guide</td>
<td style="text-align:center; padding:4px;"><a href="//en.wiktionary.org/wiki/" title="Wiktionary"><img
<td style="padding:4px;"><b><a class="external text" href="//en.wiktionary.org/">Wiktionary</a></b><br/:
Dictionary and thesaurus</td>
</tr>
</table>
</div>
<a class="extiw" href="//wikimediafoundation.org/wiki/Our_projects" title="wmf:Our projects">projects</a
<table class="layout plainlinks" style="width:100%; margin:auto; text-align:left; background:transparent
<tr>
<td style="text-align:center; padding:4px;"><a href="//commons.wikimedia.org/wiki/" title="Commons"><img
<td style="width:33%; padding:4px;"><b><a class="external text" href="//commons.wikimedia.org/">Commons
Free media repository</td>
<td style="text-align:center; padding:4px;"><a href="//www.mediawiki.org/wiki/" title="MediaWiki"><img
<td style="width:33%; padding:4px;"><b><a class="external text" href="//mediawiki.org/">MediaWiki</a></l
Wiki software development</td>
<td style="text-align:center; padding:4px;"><a href="//meta.wikimedia.org/wiki/" title="Meta-Wiki"><img
<td style="width:33%; padding:4px;"><b><a class="external text" href="//meta.wikimedia.org/">Meta-Wiki</
Wikimedia project coordination</td>
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikibooks.org/wiki/" title="Wikibooks"><img al
<td style="padding:4px;"><b><a class="external text" href="//en.wikibooks.org/">Wikibooks</a></b><br/>
Free textbooks and manuals</td>
<td style="text-align:center; padding:3px;"><a href="//www.wikidata.org/wiki/" title="Wikidata"><img alt
<td style="padding:4px;"><b><a class="external text" href="//www.wikidata.org/">Wikidata</a></b><br/>
Free knowledge base</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikinews.org/wiki/" title="Wikinews"><img alt=
<td style="padding:4px;"><b><a class="external text" href="//en.wikinews.org/">Wikinews</a></b><br/>
Free-content news</td>
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikiquote.org/wiki/" title="Wikiquote"><img al
<td style="padding:4px;"><b><a class="external text" href="//en.wikiquote.org/">Wikiquote</a></b><br/>
Collection of quotations</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikisource.org/wiki/" title="Wikisource"><img
```

```html
<td style="padding:4px;"><b><a class="external text" href="//en.wikisource.org/">Wikisource</a></b><br/>
Free-content library</td>
<td style="text-align:center; padding:4px;"><a href="//species.wikimedia.org/wiki/" title="Wikispecies">
<td style="padding:4px;"><b><a class="external text" href="//species.wikimedia.org/">Wikispecies</a></b>
Directory of species</td>
</tr>
<tr>
<td style="text-align:center; padding:4px;"><a href="//en.wikiversity.org/wiki/" title="Wikiversity"><im
<td style="padding:4px;"><b><a class="external text" href="//en.wikiversity.org/">Wikiversity</a></b><b
Free learning materials and activities</td>
<td style="text-align:center; padding:4px;"><a href="//en.wikivoyage.org/wiki/" title="Wikivoyage"><img
<td style="padding:4px;"><b><a class="external text" href="//en.wikivoyage.org/">Wikivoyage</a></b><br/>
Free travel guide</td>
<td style="text-align:center; padding:4px;"><a href="//en.wiktionary.org/wiki/" title="Wiktionary"><img
<td style="padding:4px;"><b><a class="external text" href="//en.wiktionary.org/">Wiktionary</a></b><br/>
Dictionary and thesaurus</td>
</tr>
</table>
<img alt="Commons" data-file-height="41" data-file-width="31" height="41" src="//upload.wikimedia.org/w
<br/>
<img alt="MediaWiki" data-file-height="102" data-file-width="135" height="26" src="//upload.wikimedia.o
<br/>
<img alt="Meta-Wiki" data-file-height="35" data-file-width="35" height="35" src="//upload.wikimedia.org/
<br/>
<img alt="Wikibooks" data-file-height="35" data-file-width="35" height="35" src="//upload.wikimedia.org/
<br/>
<img alt="Wikidata" data-file-height="590" data-file-width="1050" height="26" src="//upload.wikimedia.o
<br/>
<img alt="Wikinews" data-file-height="30" data-file-width="51" height="30" src="//upload.wikimedia.org/v
<br/>
<img alt="Wikiquote" data-file-height="41" data-file-width="51" height="41" src="//upload.wikimedia.org/
<br/>
<img alt="Wikisource" data-file-height="37" data-file-width="35" height="37" src="//upload.wikimedia.org
<br/>
<img alt="Wikispecies" data-file-height="41" data-file-width="35" height="41" src="//upload.wikimedia.o
<br/>
<img alt="Wikiversity" data-file-height="32" data-file-width="41" height="32" src="//upload.wikimedia.o
<br/>
<img alt="Wikivoyage" data-file-height="193" data-file-width="193" height="35" src="//upload.wikimedia.
<br/>
<img alt="Wiktionary" data-file-height="35" data-file-width="51" height="35" src="//upload.wikimedia.org
<br/>
<span style="display:none">(<span class="bday dtstart published updated">2001</span>)</span>
<ul>
<li id="lang-3">More than 1,000,000 articles:
<div class="hlist inline">
<ul>
<li><a class="external text" href="//de.wikipedia.org/wiki/"><span class="autonym" lang="de" title="Ger
<li><a class="external text" href="//es.wikipedia.org/wiki/"><span class="autonym" lang="es" title="Spa
<li><a class="external text" href="//fr.wikipedia.org/wiki/"><span class="autonym" lang="fr" title="Fre
<li><a class="external text" href="//it.wikipedia.org/wiki/"><span class="autonym" lang="it" title="Ital
<li><a class="external text" href="//nl.wikipedia.org/wiki/"><span class="autonym" lang="nl" title="Dut
<li><a class="external text" href="//ja.wikipedia.org/wiki/"><span class="autonym" lang="ja" title="Japa
<li><a class="external text" href="//pl.wikipedia.org/wiki/"><span class="autonym" lang="pl" title="Pol
```

```html
<li><a class="external text" href="//ru.wikipedia.org/wiki/"><span class="autonym" lang="ru" title="Russ
<li><a class="external text" href="//sv.wikipedia.org/wiki/"><span class="autonym" lang="sv" title="Swed
<li><a class="external text" href="//vi.wikipedia.org/wiki/"><span class="autonym" lang="vi" title="Viet
</ul>
</div>
</li>
<li id="lang-2">More than 250,000 articles:
<div class="hlist inline">
<ul>
<li><a class="external text" href="//ar.wikipedia.org/wiki/"><span class="autonym" lang="ar" title="Arab
<li><a class="external text" href="//id.wikipedia.org/wiki/"><span class="autonym" lang="id" title="Inde
<li><a class="external text" href="//ms.wikipedia.org/wiki/"><span class="autonym" lang="ms" title="Mala
<li><a class="external text" href="//ca.wikipedia.org/wiki/"><span class="autonym" lang="ca" title="Cata
<li><a cls="external text" href="//cs.wikipedia.org/wiki/"><span class="autonym" lang="cs" title="Czec
<li><a class="external text" href="//fa.wikipedia.org/wiki/"><span class="autonym" lang="fa" title="Pers
<li><a class="external text" href="//ko.wikipedia.org/wiki/"><span class="autonym" lang="ko" title="Kore
<li><a class="external text" href="//hu.wikipedia.org/wiki/"><span class="autonym" lang="hu" title="Hung
<li><a class="external text" href="//no.wikipedia.org/wiki/"><span class="autonym" lang="no" title="Norw
<li><a class="external text" href="//pt.wikipedia.org/wiki/"><span class="autonym" lang="pt" title="Port
<li><a class="external text" href="//ro.wikipedia.org/wiki/"><span class="autonym" lang="ro" title="Roma
<li><a class="external text" href="//sr.wikipedia.org/wiki/"><span class="autonym" lang="sr" title="Serl
<li><a class="external text" href="//sh.wikipedia.org/wiki/"><span class="autonym" lang="sh" title="Serl
<li><a class="external text" href="//fi.wikipedia.org/wiki/"><span class="autonym" lang="fi" title="Finn
<li><a class="external text" href="//tr.wikipedia.org/wiki/"><span class="autonym" lang="tr" title="Turl
<li><a class="external text" href="//uk.wikipedia.org/wiki/"><span class="autonym" lang="uk" title="Ukra
<li><a class="external text" href="//zh.wikipedia.org/wiki/"><span class="autonym" lang="zh" title="Chir
</ul>
</div>
</li>
<li id="lang-1">More than 50,000 articles:
<div class="hlist inline">
<ul>
<li><a class="external text" href="//bs.wikipedia.org/wiki/"><span class="autonym" lang="bs" title="Bosr
<li><a class="external text" href="//bg.wikipedia.org/wiki/"><span class="autonym" lang="bg" title="Bulg
<li><a class="external text" href="//da.wikipedia.org/wiki/"><span class="autonym" lang="da" title="Danr
<li><a class="external text" href="//et.wikipedia.org/wiki/"><span class="autonym" lang="et" title="Esto
<li><a class="external text" href="//el.wikipedia.org/wiki/"><span class="autonym" lang="el" title="Gree
<li><a class="external text" href="//simple.wikipedia.org/wiki/"><span class="autonym" lang="simple" ti
<li><a class="external text" href="//eo.wikipedia.org/wiki/"><span class="autonym" lang="eo" title="Espe
<li><a class="external text" href="//eu.wikipedia.org/wiki/"><span class="autonym" lang="eu" title="Basc
<li><a class="external text" href="//gl.wikipedia.org/wiki/"><span class="autonym" lang="gl" title="Gali
<li><a class="external text" href="//he.wikipedia.org/wiki/"><span class="autonym" lang="he" title="Hebr
<li><a class="external text" href="//hr.wikipedia.org/wiki/"><span class="autonym" lang="hr" title="Croa
<li><a class="external text" href="//lv.wikipedia.org/wiki/"><span class="autonym" lang="lv" title="Latv
<li><a class="external text" href="//lt.wikipedia.org/wiki/"><span class="autonym" lang="lt" title="Litl
<li><a class="external text" href="//nn.wikipedia.org/wiki/"><span class="autonym" lang="nn" title="Norw
<li><a class="external text" href="//sk.wikipedia.org/wiki/"><span class="autonym" lang="sk" title="Slov
<li><a class="external text" href="//sl.wikipedia.org/wiki/"><span class="autonym" lang="sl" title="Slov
<li><a class="external text" href="//th.wikipedia.org/wiki/"><span class="autonym" lang="th" title="Thai
</ul>
</div>
</li>
</ul>
<span class="autonym" lang="de" title="German (de:)" xml:lang="de">Deutsch</span>
```

```
<span class="autonym" lang="es" title="Spanish (es:)" xml:lang="es">Espaol</span>
<span class="autonym" lang="fr" title="French (fr:)" xml:lang="fr">Franais</span>
<span class="autonym" lang="it" title="Italian (it:)" xml:lang="it">Italiano</span>
<span class="autonym" lang="nl" title="Dutch (nl:)" xml:lang="nl">Nederlands</span>
<span class="autonym" lang="ja" title="Japanese (ja:)" xml:lang="ja"></span>
<span class="autonym" lang="pl" title="Polish (pl:)" xml:lang="pl">Polski</span>
<span class="autonym" lang="ru" title="Russian (ru:)" xml:lang="ru"></span>
<span class="autonym" lang="sv" title="Swedish (sv:)" xml:lang="sv">Svenska</span>
<span class="autonym" lang="vi" title="Vietnamese (vi:)" xml:lang="vi">Ting Vit</span>
<span class="autonym" lang="ar" title="Arabic (ar:)" xml:lang="ar"></span>
<span class="autonym" lang="id" title="Indonesian (id:)" xml:lang="id">Bahasa Indonesia</span>
<span class="autonym" lang="ms" title="Malay (ms:)" xml:lang="ms">Bahasa Melayu</span>
<span class="autonym" lang="ca" title="Catalan (ca:)" xml:lang="ca">Catal</span>
<span class="autonym" lang="cs" title="Czech (cs:)" xml:lang="cs">etina</span>
<span class="autonym" lang="fa" title="Persian (fa:)" xml:lang="fa"></span>
<span class="autonym" lang="ko" title="Korean (ko:)" xml:lang="ko"></span>
<span class="autonym" lang="hu" title="Hungarian (hu:)" xml:lang="hu">Magyar</span>
<span class="autonym" lang="no" title="Norwegian (no:)" xml:lang="no">Norsk bokml</span>
<span class="autonym" lang="pt" title="Portuguese (pt:)" xml:lang="pt">Portugus</span>
<span class="autonym" lang="ro" title="Romanian (ro:)" xml:lang="ro">Romn</span>
<span class="autonym" lang="sr" title="Serbian (sr:)" xml:lang="sr">Srpski / </span>
<span class="autonym" lang="sh" title="Serbo-Croatian (sh:)" xml:lang="sh">Srpskohrvatski / </span>
<span class="autonym" lang="fi" title="Finnish (fi:)" xml:lang="fi">Suomi</span>
<span class="autonym" lang="tr" title="Turkish (tr:)" xml:lang="tr">Trke</span>
<span class="autonym" lang="uk" title="Ukrainian (uk:)" xml:lang="uk"></span>
<span class="autonym" lang="zh" title="Chinese (zh:)" xml:lang="zh"></span>
<span class="autonym" lang="bs" title="Bosnian (bs:)" xml:lang="bs">Bosanski</span>
<span class="autonym" lang="bg" title="Bulgarian (bg:)" xml:lang="bg"></span>
<span class="autonym" lang="da" title="Danish (da:)" xml:lang="da">Dansk</span>
<span class="autonym" lang="et" title="Estonian (et:)" xml:lang="et">Eesti</span>
<span class="autonym" lang="el" title="Greek (el:)" xml:lang="el"></span>
<span class="autonym" lang="simple" title="Simple English (simple:)" xml:lang="simple">English (simple)</span>
<span class="autonym" lang="eo" title="Esperanto (eo:)" xml:lang="eo">Esperanto</span>
<span class="autonym" lang="eu" title="Basque (eu:)" xml:lang="eu">Euskara</span>
<span class="autonym" lang="gl" title="Galician (gl:)" xml:lang="gl">Galego</span>
<span class="autonym" lang="he" title="Hebrew (he:)" xml:lang="he"></span>
<span class="autonym" lang="hr" title="Croatian (hr:)" xml:lang="hr">Hrvatski</span>
<span class="autonym" lang="lv" title="Latvian (lv:)" xml:lang="lv">Latvieu</span>
<span class="autonym" lang="lt" title="Lithuanian (lt:)" xml:lang="lt">Lietuvi</span>
<span class="autonym" lang="nn" title="Norwegian Nynorsk (nn:)" xml:lang="nn">Norsk nynorsk</span>
<span class="autonym" lang="sk" title="Slovak (sk:)" xml:lang="sk">Slovenina</span>
<span class="autonym" lang="sl" title="Slovenian (sl:)" xml:lang="sl">Slovenina</span>
<span class="autonym" lang="th" title="Thai (th:)" xml:lang="th"></span>
<noscript><img alt="" height="1" src="//en.wikipedia.org/wiki/Special:CentralAutoLogin/start?type=1x1" 
<div class="catlinks catlinks-allhidden" data-mw="interface" id="catlinks"></div>
<li id="pt-anoncontribs"><a accesskey="y" href="/wiki/Special:MyContributions" title="A list of edits ma
<li id="pt-createaccount"><a href="/w/index.php?title=Special:UserLogin&amp;returnto=Main+Page&amp;type=
<li id="pt-login"><a accesskey="o" href="/w/index.php?title=Special:UserLogin&amp;returnto=Main+Page" t
<div id="left-navigation">
<div aria-labelledby="p-namespaces-label" class="vectorTabs" id="p-namespaces" role="navigation">
<h3 id="p-namespaces-label">Namespaces</h3>
<ul>
<li class="selected" id="ca-nstab-main"><span><a accesskey="c" href="/wiki/Main_Page" title="View the co
<li id="ca-talk"><span><a accesskey="t" href="/wiki/Talk:Main_Page" rel="discussion" title="Discussion a
```

```
</ul>
</div>
<div aria-labelledby="p-variants-label" class="vectorMenu emptyPortlet" id="p-variants" role="navigatio
<h3 id="p-variants-label">
<span>Variants</span><a href="#"></a>
</h3>
<div class="menu">
<ul>
</ul>
</div>
</div>
</div>
<li id="ca-talk"><span><a accesskey="t" href="/wiki/Talk:Main_Page" rel="discussion" title="Discussion a
<div aria-labelledby="p-variants-label" class="vectorMenu emptyPortlet" id="p-variants" role="navigatio
<h3 id="p-variants-label">
<span>Variants</span><a href="#"></a>
</h3>
<div class="menu">
<ul>
</ul>
</div>
</div>
<div class="menu">
<ul>
</ul>
</div>
<li id="ca-viewsource"><span><a accesskey="e" href="/w/index.php?title=Main_Page&amp;action=edit" title=
You can view its source [e]">View source</a></span></li>
<li class="collapsible" id="ca-history"><span><a accesskey="h" href="/w/index.php?title=Main_Page&amp;ac
<div aria-labelledby="p-cactions-label" class="vectorMenu emptyPortlet" id="p-cactions" role="navigatio
<h3 id="p-cactions-label"><span>More</span><a href="#"></a></h3>
<div class="menu">
<ul>
</ul>
</div>
</div>
<div class="menu">
<ul>
</ul>
</div>
<div aria-labelledby="p-navigation-label" class="portal" id="p-navigation" role="navigation">
<h3 id="p-navigation-label">Navigation</h3>
<div class="body">
<ul>
<li id="n-mainpage-description"><a accesskey="z" href="/wiki/Main_Page" title="Visit the main page [z]">
</div>
</div>
<li id="n-contents"><a href="/wiki/Portal:Contents" title="Guides to browsing Wikipedia">Contents</a></l
<li id="n-featuredcontent"><a href="/wiki/Portal:Featured_content" title="Featured content  the best of
<li id="n-currentevents"><a href="/wiki/Portal:Current_events" title="Find background information on cur
<li id="n-randompage"><a accesskey="x" href="/wiki/Special:Random" title="Load a random article [x]">Ran
<li id="n-sitesupport"><a href="https://donate.wikimedia.org/wiki/Special:FundraiserRedirector?utm_sourc
<li id="n-shoplink"><a href="//shop.wikimedia.org" title="Visit the Wikipedia store">Wikipedia store</a>
<div aria-labelledby="p-interaction-label" class="portal" id="p-interaction" role="navigation">
```

```html
<h3 id="p-interaction-label">Interaction</h3>
<div class="body">
<ul>
<li id="n-help"><a href="/wiki/Help:Contents" title="Guidance on how to use and edit Wikipedia">Help</a
</div>
</div>
<li id="n-aboutsite"><a href="/wiki/Wikipedia:About" title="Find out about Wikipedia">About Wikipedia</a
<li id="n-portal"><a href="/wiki/Wikipedia:Community_portal" title="About the project, what you can do,
<li id="n-recentchanges"><a accesskey="r" href="/wiki/Special:RecentChanges" title="A list of recent cha
<li id="n-contactpage"><a href="//en.wikipedia.org/wiki/Wikipedia:Contact_us" title="How to contact Wiki
<div aria-labelledby="p-tb-label" class="portal" id="p-tb" role="navigation">
<h3 id="p-tb-label">Tools</h3>
<div class="body">
<ul>
<li id="t-whatlinkshere"><a accesskey="j" href="/wiki/Special:WhatLinksHere/Main_Page" title="List of al
</div>
</div>
<li id="t-recentchangeslinked"><a accesskey="k" href="/wiki/Special:RecentChangesLinked/Main_Page" title
<li id="t-upload"><a accesskey="u" href="/wiki/Wikipedia:File_Upload_Wizard" title="Upload files [u]">Up
<li id="t-specialpages"><a accesskey="q" href="/wiki/Special:SpecialPages" title="A list of all special
<li id="t-permalink"><a href="/w/index.php?title=Main_Page&amp;oldid=696846920" title="Permanent link to
<li id="t-info"><a href="/w/index.php?title=Main_Page&amp;action=info" title="More information about thi
<li id="t-wikibase"><a accesskey="g" href="//www.wikidata.org/wiki/Q5296" title="Link to connected data
<li id="t-cite"><a href="/w/index.php?title=Special:CiteThisPage&amp;page=Main_Page&amp;id=696846920" ti
<div aria-labelledby="p-coll-print_export-label" class="portal" id="p-coll-print_export" role="navigatio
<h3 id="p-coll-print_export-label">Print/export</h3>
<div class="body">
<ul>
<li id="coll-create_a_book"><a href="/w/index.php?title=Special:Book&amp;bookcmd=book_creator&amp;refere
</div>
</div>
<li id="coll-download-as-rdf2latex"><a href="/w/index.php?title=Special:Book&amp;bookcmd=render_article&
<li id="t-print"><a accesskey="p" href="/w/index.php?title=Main_Page&amp;printable=yes" title="Printable
<div aria-labelledby="p-wikibase-otherprojects-label" class="portal" id="p-wikibase-otherprojects" role=
<h3 id="p-wikibase-otherprojects-label">In other projects</h3>
<div class="body">
<ul>
<li class="wb-otherproject-link wb-otherproject-commons"><a href="https://commons.wikimedia.org/wiki/Mai
</div>
</div>
<li class="wb-otherproject-link wb-otherproject-meta"><a href="https://meta.wikimedia.org/wiki/Main_Page
<li class="wb-otherproject-link wb-otherproject-species"><a href="https://species.wikimedia.org/wiki/Mai
<li class="wb-otherproject-link wb-otherproject-wikibooks"><a href="https://en.wikibooks.org/wiki/Main_F
<li class="wb-otherproject-link wb-otherproject-wikidata"><a href="https://www.wikidata.org/wiki/Wikidat
<li class="wb-otherproject-link wb-otherproject-wikinews"><a href="https://en.wikinews.org/wiki/Main_Pag
<li class="wb-otherproject-link wb-otherproject-wikiquote"><a href="https://en.wikiquote.org/wiki/Main_F
<li class="wb-otherproject-link wb-otherproject-wikisource"><a href="https://en.wikisource.org/wiki/Mair
<li class="wb-otherproject-link wb-otherproject-wikiversity"><a href="https://en.wikiversity.org/wiki/Wi
<li class="wb-otherproject-link wb-otherproject-wikivoyage"><a href="https://en.wikivoyage.org/wiki/Mair
<div aria-labelledby="p-lang-label" class="portal" id="p-lang" role="navigation">
<h3 id="p-lang-label">Languages</h3>
<div class="body">
<ul>
<li class="interlanguage-link interwiki-simple"><a href="//simple.wikipedia.org/wiki/" hreflang="simple
```

```html
</div>
</div>
<li class="interlanguage-link interwiki-ar"><a href="//ar.wikipedia.org/wiki/" hreflang="ar" lang="ar"
<li class="interlanguage-link interwiki-id"><a href="//id.wikipedia.org/wiki/" hreflang="id" lang="id"
<li class="interlanguage-link interwiki-ms"><a href="//ms.wikipedia.org/wiki/" hreflang="ms" lang="ms"
<li class="interlanguage-link interwiki-bs"><a href="//bs.wikipedia.org/wiki/" hreflang="bs" lang="bs"
<li class="interlanguage-link interwiki-bg"><a href="//bg.wikipedia.org/wiki/" hreflang="bg" lang="bg"
<li class="interlanguage-link interwiki-ca"><a href="//ca.wikipedia.org/wiki/" hreflang="ca" lang="ca"
<li class="interlanguage-link interwiki-cs"><a href="//cs.wikipedia.org/wiki/" hreflang="cs" lang="cs"
<li class="interlanguage-link interwiki-da"><a href="//da.wikipedia.org/wiki/" hreflang="da" lang="da"
<li class="interlanguage-link interwiki-de"><a href="//de.wikipedia.org/wiki/" hreflang="de" lang="de"
<li class="interlanguage-link interwiki-et"><a href="//et.wikipedia.org/wiki/" hreflang="et" lang="et"
<li class="interlanguage-link interwiki-el"><a href="//el.wikipedia.org/wiki/" hreflang="el" lang="el"
<li class="interlanguage-link interwiki-es"><a href="//es.wikipedia.org/wiki/" hreflang="es" lang="es"
<li class="interlanguage-link interwiki-eo"><a href="//eo.wikipedia.org/wiki/" hreflang="eo" lang="eo"
<li class="interlanguage-link interwiki-eu"><a href="//eu.wikipedia.org/wiki/" hreflang="eu" lang="eu"
<li class="interlanguage-link interwiki-fa"><a href="//fa.wikipedia.org/wiki/" hreflang="fa" lang="fa"
<li class="interlanguage-link interwiki-fr"><a href="//fr.wikipedia.org/wiki/" hreflang="fr" lang="fr"
<li class="interlanguage-link interwiki-gl"><a href="//gl.wikipedia.org/wiki/" hreflang="gl" lang="gl"
<li class="interlanguage-link interwiki-ko"><a href="//ko.wikipedia.org/wiki/" hreflang="ko" lang="ko"
<li class="interlanguage-link interwiki-he"><a href="//he.wikipedia.org/wiki/" hreflang="he" lang="he"
<li class="interlanguage-link interwiki-hr"><a href="//hr.wikipedia.org/wiki/" hreflang="hr" lang="hr"
<li class="interlanguage-link interwiki-it"><a href="//it.wikipedia.org/wiki/" hreflang="it" lang="it"
<li class="interlanguage-link interwiki-ka"><a href="//ka.wikipedia.org/wiki/" hreflang="ka" lang="ka"
<li class="interlanguage-link interwiki-lv"><a href="//lv.wikipedia.org/wiki/" hreflang="lv" lang="lv"
<li class="interlanguage-link interwiki-lt"><a href="//lt.wikipedia.org/wiki/" hreflang="lt" lang="lt"
<li class="interlanguage-link interwiki-hu"><a href="//hu.wikipedia.org/wiki/" hreflang="hu" lang="hu"
<li class="interlanguage-link interwiki-nl"><a href="//nl.wikipedia.org/wiki/" hreflang="nl" lang="nl"
<li class="interlanguage-link interwiki-ja"><a href="//ja.wikipedia.org/wiki/" hreflang="ja" lang="ja"
<li class="interlanguage-link interwiki-no"><a href="//no.wikipedia.org/wiki/" hreflang="no" lang="no"
<li class="interlanguage-link interwiki-nn"><a href="//nn.wikipedia.org/wiki/" hreflang="nn" lang="nn"
<li class="interlanguage-link interwiki-pl"><a href="//pl.wikipedia.org/wiki/" hreflang="pl" lang="pl"
<li class="interlanguage-link interwiki-pt"><a href="//pt.wikipedia.org/wiki/" hreflang="pt" lang="pt"
<li class="interlanguage-link interwiki-ro"><a href="//ro.wikipedia.org/wiki/" hreflang="ro" lang="ro"
<li class="interlanguage-link interwiki-ru"><a href="//ru.wikipedia.org/wiki/" hreflang="ru" lang="ru"
<li class="interlanguage-link interwiki-sk"><a href="//sk.wikipedia.org/wiki/" hreflang="sk" lang="sk"
<li class="interlanguage-link interwiki-sl"><a href="//sl.wikipedia.org/wiki/" hreflang="sl" lang="sl"
<li class="interlanguage-link interwiki-sr"><a href="//sr.wikipedia.org/wiki/" hreflang="sr" lang="sr"
<li class="interlanguage-link interwiki-sh"><a href="//sh.wikipedia.org/wiki/" hreflang="sh" lang="sh"
<li class="interlanguage-link interwiki-fi"><a href="//fi.wikipedia.org/wiki/" hreflang="fi" lang="fi"
<li class="interlanguage-link interwiki-sv"><a href="//sv.wikipedia.org/wiki/" hreflang="sv" lang="sv"
<li class="interlanguage-link interwiki-th"><a href="//th.wikipedia.org/wiki/" hreflang="th" lang="th"
<li class="interlanguage-link interwiki-vi"><a href="//vi.wikipedia.org/wiki/" hreflang="vi" lang="vi"
<li class="interlanguage-link interwiki-tr"><a href="//tr.wikipedia.org/wiki/" hreflang="tr" lang="tr"
<li class="interlanguage-link interwiki-uk"><a href="//uk.wikipedia.org/wiki/" hreflang="uk" lang="uk"
<li class="interlanguage-link interwiki-zh"><a href="//zh.wikipedia.org/wiki/" hreflang="zh" lang="zh"
<li class="uls-p-lang-dummy"><a href="#"></a></li>
<div id="footer" role="contentinfo">
<ul id="footer-info">
<li id="footer-info-lastmod"> This page was last modified on 26 December 2015, at 10:03.</li>
<li id="footer-info-copyright">Text is available under the <a href="//en.wikipedia.org/wiki/Wikipedia:T
additional terms may apply.  By using this site, you agree to the <a href="//wikimediafoundation.org/wi
</ul>
<ul id="footer-places">
```

```
<li id="footer-places-privacy"><a href="//wikimediafoundation.org/wiki/Privacy_policy" title="wmf:Privac
<li id="footer-places-about"><a href="/wiki/Wikipedia:About" title="Wikipedia:About">About Wikipedia</a
<li id="footer-places-disclaimer"><a href="/wiki/Wikipedia:General_disclaimer" title="Wikipedia:General
<li id="footer-places-contact"><a href="//en.wikipedia.org/wiki/Wikipedia:Contact_us">Contact Wikipedia<
<li id="footer-places-developers"><a href="https://www.mediawiki.org/wiki/Special:MyLanguage/How_to_cont
<li id="footer-places-cookiestatement"><a href="//wikimediafoundation.org/wiki/Cookie_statement">Cookie
<li id="footer-places-mobileview"><a class="noprint stopMobileRedirectToggle" href="//en.m.wikipedia.org
</ul>
<ul class="noprint" id="footer-icons">
<li id="footer-copyrightico">
<a href="//wikimediafoundation.org/"><img alt="Wikimedia Foundation" height="31" src="/static/images/wil
<li id="footer-poweredbyico">
<a href="//www.mediawiki.org/"><img alt="Powered by MediaWiki" height="31" src="/w/resources/assets/powe
</ul>
<div style="clear:both"></div>
</div>
<a href="//creativecommons.org/licenses/by-sa/3.0/" rel="license" style="display:none;"></a>
<a href="//wikimediafoundation.org/wiki/Terms_of_Use">Terms of Use</a>
<a href="//wikimediafoundation.org/wiki/Privacy_policy">Privacy Policy</a>
<a href="//www.wikimediafoundation.org/">Wikimedia Foundation, Inc.</a>
<ul id="footer-places">
<li id="footer-places-privacy"><a href="//wikimediafoundation.org/wiki/Privacy_policy" title="wmf:Privac
<li id="footer-places-about"><a href="/wiki/Wikipedia:About" title="Wikipedia:About">About Wikipedia</a
<li id="footer-places-disclaimer"><a href="/wiki/Wikipedia:General_disclaimer" title="Wikipedia:General
<li id="footer-places-contact"><a href="//en.wikipedia.org/wiki/Wikipedia:Contact_us">Contact Wikipedia<
<li id="footer-places-developers"><a href="https://www.mediawiki.org/wiki/Special:MyLanguage/How_to_cont
<li id="footer-places-cookiestatement"><a href="//wikimediafoundation.org/wiki/Cookie_statement">Cookie
<li id="footer-places-mobileview"><a class="noprint stopMobileRedirectToggle" href="//en.m.wikipedia.org
</ul>
<li id="footer-places-about"><a href="/wiki/Wikipedia:About" title="Wikipedia:About">About Wikipedia</a
<li id="footer-places-disclaimer"><a href="/wiki/Wikipedia:General_disclaimer" title="Wikipedia:General
<li id="footer-places-contact"><a href="//en.wikipedia.org/wiki/Wikipedia:Contact_us">Contact Wikipedia<
<li id="footer-places-developers"><a href="https://www.mediawiki.org/wiki/Special:MyLanguage/How_to_cont
<li id="footer-places-cookiestatement"><a href="//wikimediafoundation.org/wiki/Cookie_statement">Cookie
<li id="footer-places-mobileview"><a class="noprint stopMobileRedirectToggle" href="//en.m.wikipedia.org
<ul class="noprint" id="footer-icons">
<li id="footer-copyrightico">
<a href="//wikimediafoundation.org/"><img alt="Wikimedia Foundation" height="31" src="/static/images/wil
<li id="footer-poweredbyico">
<a href="//www.mediawiki.org/"><img alt="Powered by MediaWiki" height="31" src="/w/resources/assets/powe
</ul>
<img alt="Wikimedia Foundation" height="31" src="/static/images/wikimedia-button.png" srcset="/static/in
<img alt="Powered by MediaWiki" height="31" src="/w/resources/assets/poweredby_mediawiki_88x31.png" srcs
```

Time for a challenge! To make sure that everyone is on the same page (and to give you a little more practice dealing with HTML), let's partner up with the person next to you and try challenge A, on using html, in the challenges directory.

1.3.1 Extended example: scraping Wikipedia

Our first use case involves scraping some basic information about technology companies from Wikipedia. Say you are the chief innovation officer of a small city in the San Francisco Bay Area. A number of large-scale local investments in office space have taken place, with space opening up over the next few years. You wish to be part of the trend of technology companies moving out of San Francisco and Silicon Valley. You have

been networking and talking to companies at events and conferences, but would like a more systematic way of identifying companies to focus on.

You notice a list of 179 technology companies based in the San Francisco area on Wikipedia: https://en.wikipedia.org/wiki/Category:Technology_companies_based_in_the_San_Francisco_Bay_Area

Your goal is to scrape basic useful information about each company in a list, into which you can do some summary statistics to identify companies or even industries you are interested in focusing on.

** In particular, you want to know: **

1) what industry they are in

2) where the company is currently headquartered

3) the number of employees

4) website address of the company

This will allow you to know the current and budding tech hubs in the Bay area, get a better sense of your competition, and the number of jobs you can attract to your city. For convenience, you also collate the website addresses of the companies to pull into your list.

Typically, you'll first want to identify the element that you want to pull data from. Next, you'll need to figure out a strategy to locate and "crawl" through relevant pages. In a forum, for instance, a bot may be set up to click the "Next page" button once all posts on a single page have been visited and saved. More advanced strategies would include visiting all hyperlinks on every page visited, so that the bot continually updates the list of links to crawl through.

Inspecting the Index page   First, you will want to inspect the element associated with each link we want to visit on the index page.

Next, you will want to inspect the element with the data we would like to extract, corresponding to each link on the index page.

Inspecting each individual page   In our case, it looks like the format of the data in both the index and individual pages are regular enough for us to be able to parse them programatically. We next confirm this by interacting with both the index and individual pages in our console.

```
In [15]: page = BeautifulSoup(requests.get(
                'https://en.wikipedia.org/wiki/Category:Technology_companies_based_in_the_San_Francisco
             ).content)

/Users/dillon/anaconda/lib/python3.5/site-packages/bs4/__init__.py:166: UserWarning: No parser was explic

To get rid of this warning, change this:

 BeautifulSoup([your markup])

to this:

 BeautifulSoup([your markup], "lxml")

  markup_type=markup_type))
```

If you scroll over the div with id "mw-pages" on the Elements tab, you'll see that it corresponds to the entire "Technology companies based in the San Francisco Bay Area" pane.

Let's first try to select this, and confirm we've selected the right element by printing out the result. In the code, we are telling soup to find any elements with the "div" element tag, with id "mw-pages" that we saw in the inspect element pane.

```
In [16]: company_section = page.find_all("div", {"id": "mw-pages"})
         print(type(company_section))

<class 'bs4.element.ResultSet'>
```

As we navigate the result returned, we see that it is a "ResultSet", which suggests that it can be retrieved by index. You can also just try it out.

```
In [17]: print(company_section[0])

<div id="mw-pages">
<h2><span id="Pages_in_category"></span>Pages in category "Technology companies based in the San Francis
<p>The following 180 pages are in this category, out of 180 total. This list may not reflect recent chan
</p><div class="mw-content-ltr" dir="ltr" lang="en"><div class="mw-category"><div class="mw-category-gro
<ul><li><a href="/wiki/HP_3PAR" title="HP 3PAR">HP 3PAR</a></li></ul></div><div class="mw-category-group
<ul><li><a href="/wiki/4D_Inc" title="4D Inc">4D Inc</a></li></ul></div><div class="mw-category-group"><
<ul><li><a href="/wiki/Achronix" title="Achronix">Achronix</a></li>
<li><a href="/wiki/Advanced_Micro_Devices" title="Advanced Micro Devices">Advanced Micro Devices</a></li
<li><a href="/wiki/Aerohive_Networks" title="Aerohive Networks">Aerohive Networks</a></li>
<li><a href="/wiki/Affymetrix" title="Affymetrix">Affymetrix</a></li>
<li><a href="/wiki/Agami_Systems" title="Agami Systems">Agami Systems</a></li>
<li><a href="/wiki/Agilent_Technologies" title="Agilent Technologies">Agilent Technologies</a></li>
<li><a href="/wiki/AirTouch" title="AirTouch">AirTouch</a></li>
<li><a href="/wiki/AKM_Semiconductor,_Inc." title="AKM Semiconductor, Inc.">AKM Semiconductor, Inc.</a>
<li><a href="/wiki/All_Power_Labs" title="All Power Labs">All Power Labs</a></li>
<li><a href="/wiki/Alphabet_Energy" title="Alphabet Energy">Alphabet Energy</a></li>
<li><a href="/wiki/AlphaSense" title="AlphaSense">AlphaSense</a></li>
<li><a href="/wiki/Altamira_Software" title="Altamira Software">Altamira Software</a></li>
<li><a href="/wiki/AltaVista" title="AltaVista">AltaVista</a></li>
<li><a href="/wiki/Altos_Computer_Systems" title="Altos Computer Systems">Altos Computer Systems</a></li
<li><a href="/wiki/Alza" title="Alza">Alza</a></li>
<li><a href="/wiki/AMAX_Information_Technologies" title="AMAX Information Technologies">AMAX Information
<li><a href="/wiki/American_Logic_Machines" title="American Logic Machines">American Logic Machines</a><
<li><a href="/wiki/Amiga_Corporation" title="Amiga Corporation">Amiga Corporation</a></li>
<li><a href="/wiki/Antec" title="Antec">Antec</a></li>
<li><a href="/wiki/Anthera_Pharmaceuticals" title="Anthera Pharmaceuticals">Anthera Pharmaceuticals</a>
<li><a href="/wiki/Apple_Inc." title="Apple Inc.">Apple Inc.</a></li>
<li><a href="/wiki/ASSIA_(company)" title="ASSIA (company)">ASSIA (company)</a></li>
<li><a href="/wiki/Audience_(company)" title="Audience (company)">Audience (company)</a></li>
<li><a href="/wiki/AuraOne_Systems" title="AuraOne Systems">AuraOne Systems</a></li>
<li><a href="/wiki/Aureal_Semiconductor" title="Aureal Semiconductor">Aureal Semiconductor</a></li></ul>
<ul><li><a href="/wiki/BioMarin_Pharmaceutical" title="BioMarin Pharmaceutical">BioMarin Pharmaceutical<
<li><a href="/wiki/BioPharm_(US_company)" title="BioPharm (US company)">BioPharm (US company)</a></li>
<li><a href="/wiki/Biosearch_Technologies" title="Biosearch Technologies">Biosearch Technologies</a></li
<li><a href="/wiki/BrightSource_Energy" title="BrightSource Energy">BrightSource Energy</a></li>
<li><a href="/wiki/Br%C3%B8derbund" title="Brderbund">Brderbund</a></li>
<li><span class="redirect-in-category"><a class="mw-redirect" href="/wiki/Byte_Shop" title="Byte Shop">B
<ul><li><a href="/wiki/Cask_(company)" title="Cask (company)">Cask (company)</a></li>
<li><a href="/wiki/Cerego" title="Cerego">Cerego</a></li>
<li><a href="/wiki/Cetus_Corporation" title="Cetus Corporation">Cetus Corporation</a></li>
<li><a href="/wiki/Cisco_Systems" title="Cisco Systems">Cisco Systems</a></li>
<li><a href="/wiki/Clean_Edge" title="Clean Edge">Clean Edge</a></li>
<li><a href="/wiki/Complete_Genomics" title="Complete Genomics">Complete Genomics</a></li>
<li><a href="/wiki/Corsair_Components" title="Corsair Components">Corsair Components</a></li>
<li><a href="/wiki/Covad" title="Covad">Covad</a></li>
<li><a href="/wiki/Crossbar_(computer_hardware_manufacturer)" title="Crossbar (computer hardware manufac
<li><a href="/wiki/Cutter_Laboratories" title="Cutter Laboratories">Cutter Laboratories</a></li>
<li><a href="/wiki/Cypress_Semiconductor" title="Cypress Semiconductor">Cypress Semiconductor</a></li></
<ul><li><a href="/wiki/Deeplearning4j" title="Deeplearning4j">Deeplearning4j</a></li>
<li><a href="/wiki/DiscoveRx" title="DiscoveRx">DiscoveRx</a></li>
```

```
<li><a href="/wiki/DNA2.0" title="DNA2.0">DNA2.0</a></li>
<li><a href="/wiki/DOER_Marine" title="DOER Marine">DOER Marine</a></li>
<li><a href="/wiki/Dolby_Laboratories" title="Dolby Laboratories">Dolby Laboratories</a></li>
<li><a href="/wiki/Double_Robotics" title="Double Robotics">Double Robotics</a></li>
<li><a href="/wiki/Dust_Networks" title="Dust Networks">Dust Networks</a></li></ul></div><div class="mw-
<ul><li><a href="/wiki/Electronics_for_Imaging" title="Electronics for Imaging">Electronics for Imaging<
<li><a href="/wiki/Energy_Recovery_Inc." title="Energy Recovery Inc.">Energy Recovery Inc.</a></li>
<li><a href="/wiki/Enphase_Energy" title="Enphase Energy">Enphase Energy</a></li>
<li><a href="/wiki/Envivio" title="Envivio">Envivio</a></li>
<li><a href="/wiki/Etec_Systems,_Inc." title="Etec Systems, Inc.">Etec Systems, Inc.</a></li>
<li><a href="/wiki/Everex" title="Everex">Everex</a></li>
<li><a href="/wiki/Exidy" title="Exidy">Exidy</a></li></ul></div><div class="mw-category-group"><h3>F</h
<ul><li><a href="/wiki/Fairchild_Semiconductor" title="Fairchild Semiconductor">Fairchild Semiconductor<
<li><a href="/wiki/Fortify_Software" title="Fortify Software">Fortify Software</a></li>
<li><a href="/wiki/Four-Phase_Systems" title="Four-Phase Systems">Four-Phase Systems</a></li>
<li><a href="/wiki/Foveon" title="Foveon">Foveon</a></li></ul></div><div class="mw-category-group"><h3>
<ul><li><a href="/wiki/Genentech" title="Genentech">Genentech</a></li>
<li><a href="/wiki/General_Magic" title="General Magic">General Magic</a></li>
<li><a href="/wiki/Genesys_(company)" title="Genesys (company)">Genesys (company)</a></li>
<li><a href="/wiki/GlassPoint_Solar" title="GlassPoint Solar">GlassPoint Solar</a></li>
<li><a href="/wiki/GlobalFoundries" title="GlobalFoundries">GlobalFoundries</a></li>
<li><a href="/wiki/Green_Charge_Networks" title="Green Charge Networks">Green Charge Networks</a></li>
<li><a href="/wiki/Gusto_(software)" title="Gusto (software)">Gusto (software)</a></li></ul></div><div c
<ul><li><a href="/wiki/Handspring_(company)" title="Handspring (company)">Handspring (company)</a></li>
<li><a href="/wiki/Hercules_Computer_Technology" title="Hercules Computer Technology">Hercules Computer 
<li><a href="/wiki/Hewlett_Packard_Enterprise" title="Hewlett Packard Enterprise">Hewlett Packard Enterp
<li><a href="/wiki/Hewlett-Packard" title="Hewlett-Packard">Hewlett-Packard</a></li>
<li><a href="/wiki/Hoopla_Software" title="Hoopla Software">Hoopla Software</a></li>
<li><a href="/wiki/HP_Inc." title="HP Inc.">HP Inc.</a></li>
<li><a href="/wiki/Human_Engineered_Software" title="Human Engineered Software">Human Engineered Softwar
<li><a href="/wiki/Hurricane_Electric" title="Hurricane Electric">Hurricane Electric</a></li></ul></div>
<ul><li><a href="/wiki/Ikanos_Communications" title="Ikanos Communications">Ikanos Communications</a></l
<li><a href="/wiki/Impax_Laboratories" title="Impax Laboratories">Impax Laboratories</a></li>
<li><a href="/wiki/Intel" title="Intel">Intel</a></li>
<li><a href="/wiki/InvenSense" title="InvenSense">InvenSense</a></li>
<li><a href="/wiki/InVision_Technologies" title="InVision Technologies">InVision Technologies</a></li></
<ul><li><a href="/wiki/Jennerex" title="Jennerex">Jennerex</a></li>
<li><a href="/wiki/Juniper_Networks" title="Juniper Networks">Juniper Networks</a></li></ul></div><div c
<ul><li><a href="/wiki/KleenSpeed_Technologies" title="KleenSpeed Technologies">KleenSpeed Technologies<
<li><a href="/wiki/Kosan_Biosciences" title="Kosan Biosciences">Kosan Biosciences</a></li></ul></div><di
<ul><li><a href="/wiki/Lam_Research" title="Lam Research">Lam Research</a></li>
<li><a href="/wiki/LeapFrog_Enterprises" title="LeapFrog Enterprises">LeapFrog Enterprises</a></li>
<li><a href="/wiki/Lexar" title="Lexar">Lexar</a></li>
<li><a href="/wiki/Lightwave_Electronics_Corporation" title="Lightwave Electronics Corporation">Lightwav
<li><a href="/wiki/Linear_Technology" title="Linear Technology">Linear Technology</a></li></ul></div><di
<ul><li><a href="/wiki/Made_In_Space,_Inc." title="Made In Space, Inc.">Made In Space, Inc.</a></li>
<li><a href="/wiki/Makani_Power" title="Makani Power">Makani Power</a></li>
<li><a href="/wiki/Maxim_Integrated" title="Maxim Integrated">Maxim Integrated</a></li>
<li><a href="/wiki/Intel_Security" title="Intel Security">Intel Security</a></li>
<li><a href="/wiki/McCune_Audio/Video/Lighting" title="McCune Audio/Video/Lighting">McCune Audio/Video/L
<li><a href="/wiki/Media_Vision" title="Media Vision">Media Vision</a></li>
<li><a href="/wiki/Medivation" title="Medivation">Medivation</a></li>
<li><a href="/wiki/Meka_Robotics" title="Meka Robotics">Meka Robotics</a></li>
<li><a href="/wiki/Mendel_Biotechnology,_Inc." title="Mendel Biotechnology, Inc.">Mendel Biotechnology, 
```

```html
<li><a href="/wiki/Meru_Networks" title="Meru Networks">Meru Networks</a></li>
<li><a href="/wiki/MeWe" title="MeWe">MeWe</a></li>
<li><a href="/wiki/Meyer_Sound_Laboratories" title="Meyer Sound Laboratories">Meyer Sound Laboratories</
<li><a href="/wiki/MLab" title="MLab">MLab</a></li>
<li><a href="/wiki/Mozilla_Corporation" title="Mozilla Corporation">Mozilla Corporation</a></li></ul></d
<ul><li><a href="/wiki/Nanosolar" title="Nanosolar">Nanosolar</a></li>
<li><a href="/wiki/National_Semiconductor" title="National Semiconductor">National Semiconductor</a></li
<li><a href="/wiki/Navigenics" title="Navigenics">Navigenics</a></li>
<li><a href="/wiki/Netscape" title="Netscape">Netscape</a></li>
<li><span class="redirect-in-category"><a class="mw-redirect" href="/wiki/Netscape_Communications_Corpor
<li><a href="/wiki/NeXT" title="NeXT">NeXT</a></li>
<li><a href="/wiki/Next_Thing_Co." title="Next Thing Co.">Next Thing Co.</a></li>
<li><a href="/wiki/NovaBay_Pharmaceuticals" title="NovaBay Pharmaceuticals">NovaBay Pharmaceuticals</a><
<li><a href="/wiki/Numenta" title="Numenta">Numenta</a></li></ul></div><div class="mw-category-group"><
<ul><li><a href="/wiki/Oberheim_Electronics" title="Oberheim Electronics">Oberheim Electronics</a></li>
<li><a href="/wiki/OLogic" title="OLogic">OLogic</a></li>
<li><a href="/wiki/OpenAI" title="OpenAI">OpenAI</a></li>
<li><a href="/wiki/Oracle_Corporation" title="Oracle Corporation">Oracle Corporation</a></li>
<li><a href="/wiki/Osborne_Computer_Corporation" title="Osborne Computer Corporation">Osborne Computer C
<li><a href="/wiki/OSIsoft" title="OSIsoft">OSIsoft</a></li></ul></div><div class="mw-category-group"><
<ul><li><a href="/wiki/P.A._Semi" title="P.A. Semi">P.A. Semi</a></li>
<li><span class="redirect-in-category"><a class="mw-redirect" href="/wiki/PA_Semiconductor" title="PA Se
<li><a href="/wiki/Palm,_Inc." title="Palm, Inc.">Palm, Inc.</a></li>
<li><a href="/wiki/Palo_Alto_Networks" title="Palo Alto Networks">Palo Alto Networks</a></li>
<li><a href="/wiki/Peninsula_Engineering_Group,_Inc." title="Peninsula Engineering Group, Inc.">Peninsul
<li><a href="/wiki/Peribit_Networks" title="Peribit Networks">Peribit Networks</a></li>
<li><a href="/wiki/Pixo" title="Pixo">Pixo</a></li>
<li><a href="/wiki/Primus_Power" title="Primus Power">Primus Power</a></li>
<li><a href="/wiki/Processor_Technology" title="Processor Technology">Processor Technology</a></li>
<li><a href="/wiki/Prosetta" title="Prosetta">Prosetta</a></li>
<li><a href="/wiki/Pyramid_Technology" title="Pyramid Technology">Pyramid Technology</a></li></ul></div>
<ul><li><a href="/wiki/Qualcomm_Atheros" title="Qualcomm Atheros">Qualcomm Atheros</a></li>
<li><a href="/wiki/Quantum_Effect_Devices" title="Quantum Effect Devices">Quantum Effect Devices</a></li
<li><a href="/wiki/Quark_Pharmaceuticals" title="Quark Pharmaceuticals">Quark Pharmaceuticals</a></li>
<li><a href="/wiki/Qume" title="Qume">Qume</a></li></ul></div><div class="mw-category-group"><h3>R</h3>
<ul><li><a href="/wiki/Rambus" title="Rambus">Rambus</a></li>
<li><a href="/wiki/Recommind" title="Recommind">Recommind</a></li>
<li><a href="/wiki/Redwood_Robotics" title="Redwood Robotics">Redwood Robotics</a></li></ul></div><div c
<ul><li><a href="/wiki/Sangfor_Technologies" title="Sangfor Technologies">Sangfor Technologies</a></li>
<li><a href="/wiki/Seagate_Technology" title="Seagate Technology">Seagate Technology</a></li>
<li><a href="/wiki/Sensory,_Inc." title="Sensory, Inc.">Sensory, Inc.</a></li>
<li><a href="/wiki/Shugart_Associates" title="Shugart Associates">Shugart Associates</a></li>
<li><a href="/wiki/Sidecar_(company)" title="Sidecar (company)">Sidecar (company)</a></li>
<li><a href="/wiki/Silego_Technology_Inc." title="Silego Technology Inc.">Silego Technology Inc.</a></li
<li><a href="/wiki/Silicon_Graphics" title="Silicon Graphics">Silicon Graphics</a></li>
<li><a href="/wiki/Siluria_Technologies" title="Siluria Technologies">Siluria Technologies</a></li>
<li><a href="/wiki/Skymind" title="Skymind">Skymind</a></li>
<li><a href="/wiki/Sling_Media" title="Sling Media">Sling Media</a></li>
<li><a href="/wiki/SmugMug" title="SmugMug">SmugMug</a></li>
<li><a href="/wiki/SolarCity" title="SolarCity">SolarCity</a></li>
<li><a href="/wiki/Solido_Design_Automation" title="Solido Design Automation">Solido Design Automation</
<li><a href="/wiki/SoloPower" title="SoloPower">SoloPower</a></li>
<li><a href="/wiki/Solyndra" title="Solyndra">Solyndra</a></li>
<li><a href="/wiki/Stem_Cell_Theranostics" title="Stem Cell Theranostics">Stem Cell Theranostics</a></li
```

```
<li><a href="/wiki/SunPower" title="SunPower">SunPower</a></li>
<li><a href="/wiki/Sunrun" title="Sunrun">Sunrun</a></li>
<li><a href="/wiki/Supertek_Computers" title="Supertek Computers">Supertek Computers</a></li>
<li><a href="/wiki/Sybase" title="Sybase">Sybase</a></li>
<li><a href="/wiki/Synaptics" title="Synaptics">Synaptics</a></li></ul></div><div class="mw-category-gr
<ul><li><a href="/wiki/Tabula_(company)" title="Tabula (company)">Tabula (company)</a></li>
<li><a href="/wiki/Talari_Networks" title="Talari Networks">Talari Networks</a></li>
<li><a href="/wiki/Tandem_Computers" title="Tandem Computers">Tandem Computers</a></li>
<li><a href="/wiki/Tengen_(company)" title="Tengen (company)">Tengen (company)</a></li>
<li><a href="/wiki/Theranos" title="Theranos">Theranos</a></li>
<li><a href="/wiki/Thoratec" title="Thoratec">Thoratec</a></li>
<li><a href="/wiki/Tout_(company)" title="Tout (company)">Tout (company)</a></li></ul></div><div class='
<ul><li><a href="/wiki/Ubiquitous_Energy" title="Ubiquitous Energy">Ubiquitous Energy</a></li>
<li><a href="/wiki/Umtech" title="Umtech">Umtech</a></li>
<li><a href="/wiki/UTStarcom" title="UTStarcom">UTStarcom</a></li></ul></div><div class="mw-category-gro
<ul><li><a href="/wiki/Vivante_Corporation" title="Vivante Corporation">Vivante Corporation</a></li>
<li><a href="/wiki/Volterra_Semiconductor" title="Volterra Semiconductor">Volterra Semiconductor</a></li
<li><a href="/wiki/VPL_Research" title="VPL Research">VPL Research</a></li>
<li><a href="/wiki/VSee" title="VSee">VSee</a></li>
<li><a href="/wiki/VW_Electronics_Research_Laboratory" title="VW Electronics Research Laboratory">VW Ele
<ul><li><a href="/wiki/@WalmartLabs" title="@WalmartLabs">@WalmartLabs</a></li>
<li><a href="/wiki/Willow_Garage" title="Willow Garage">Willow Garage</a></li>
<li><a href="/wiki/Wind_River_Systems" title="Wind River Systems">Wind River Systems</a></li></ul></div>
<ul><li><a href="/wiki/Xilinx" title="Xilinx">Xilinx</a></li></ul></div><div class="mw-category-group">
<ul><li><a href="/wiki/Zenefits" title="Zenefits">Zenefits</a></li>
<li><a href="/wiki/Zscaler" title="Zscaler">Zscaler</a></li></ul></div></div></div>
</div>
```

You can see at the start of the element retrieved that it is indeed a division with id "mw-pages" - we can confirm by browsing the text that we've selected the correct element. Next, let's retrieve each section (corresponding to each alphabet), now searching the company section with class type "mw-category-group".

```
In [18]: each_alphabet = company_section[0].find_all("div", {"class":"mw-category-group"})
         print(len(each_alphabet))
         print(each_alphabet[0])

27
<div class="mw-category-group"><h3>3</h3>
<ul><li><a href="/wiki/HP_3PAR" title="HP 3PAR">HP 3PAR</a></li></ul></div>
```

Finally, within each section, we want to pull out the individual hyperlinks corresponding to each company. Let's use the second element in the index (the letter "A" instead of the category group for "3") as it has more than one company.

```
In [19]: alphabet_a = each_alphabet[1]
         print(alphabet_a)

<div class="mw-category-group"><h3>4</h3>
<ul><li><a href="/wiki/4D_Inc" title="4D Inc">4D Inc</a></li></ul></div>
```

We next want to select all elements with the "li" tag, and print them out to make sure they correspond to what we expect to see on the page.

```
In [20]: company_list = alphabet_a.find_all("li")
         for i in company_list:
             print("")
             print(i)
```

```
<li><a href="/wiki/4D_Inc" title="4D Inc">4D Inc</a></li>
```

If we select one company and print it out, we can see we're pretty close.

```
In [21]: one_company = company_list[0]
         print(one_company)
```

```
<li><a href="/wiki/4D_Inc" title="4D Inc">4D Inc</a></li>
```

We can also select the next child element by doing the following:

```
In [22]: one_company.a
```

```
Out[22]: <a href="/wiki/4D_Inc" title="4D Inc">4D Inc</a>
```

And finally, get the attributes associated with the "a" hyperlink tag, which returns a Python dictionary.

```
In [23]: one_company.a.attrs
```

```
Out[23]: {'href': '/wiki/4D_Inc', 'title': '4D Inc'}
```

Now that we've received the element containing the element we want, we can also print out its parents to view the position within the html "tree."

```
In [24]: print(type(one_company.a.parents))
         for i in one_company.a.parents:
             print(i.name)
```

```
<class 'generator'>
li
ul
div
div
div
div
div
div
div
div
body
html
[document]
```

This lets you see the different nested elements you'll need to traverse to get to the element you need. In many cases, you'll use explicit selection of the element together with the find_all command to isolate the element you need.

Finally, let's write a loop to store all of our desired hyperlink dictionaries in a single Python list.

```
In [25]: link_list = []
         for each_section in company_section:
             company_list = each_section.find_all("li")
             for each_company in company_list:
                 new_dict = each_company.a
                 link_list.append(new_dict)
         print(len(link_list))
```

```
180
```

We now have a list of 175 hyperlinks to loop through for our next section.
Now using the list, let's load the first page and locate the text elements we want

```
In [26]: example_site = link_list[0]
         print(example_site)

         company_page = requests.get("http://wikipedia.org" + example_site['href'])
```

```
<a href="/wiki/HP_3PAR" title="HP 3PAR">HP 3PAR</a>
```

```
In [27]: print(company_page.content[0:200])
```

```
b'<!DOCTYPE html>\n<html lang="en" dir="ltr" class="client-nojs">\n<head>\n<meta charset="UTF-8" />\n<ti
```

In your browser, you should be using inspect element to confirm the position of the desired element in the html tree. We can see the element is a table with class name "infobox vcard". Let's try to select this next. First, we need the html document into soup as we did before. (We convert to string just to allow us to print the first 500 charactes of the text here.)

```
In [28]: page = BeautifulSoup(company_page.content)
         info_box = page.find("table", {"class": "infobox vcard"})
         print(str(info_box)[0:500])
```

```
<table class="infobox vcard" style="width:22em">
<caption class="fn org">HP 3PAR</caption>
<tr>
<th scope="row" style="padding-right:0.5em;">
<div style="padding:0.1em 0;line-height:1.2em;"><a href="/wiki/Types_of_business_entity" title="Types of
</th>
<td class="category" style="line-height:1.35em;">Subsidiary</td>
</tr>
<tr>
<th scope="row" style="padding-right:0.5em;">Industry</th>
<td class="category" style="line-height:1.35em;"><a href="/wiki/Data_storage_dev
```

```
/Users/dillon/anaconda/lib/python3.5/site-packages/bs4/__init__.py:166: UserWarning: No parser was explic

To get rid of this warning, change this:

 BeautifulSoup([your markup])

to this:

 BeautifulSoup([your markup], "lxml")

  markup_type=markup_type))
```

Now, using the various tools we've had before, we can drill down to the specific element containing the data we need. As before, we select and print a single row to help guide the process.

```
In [29]: table_elements = info_box.find_all("tr")
         one_row = table_elements[0]
         print(one_row)
```

```
<tr>
<th scope="row" style="padding-right:0.5em;">
```

```
<div style="padding:0.1em 0;line-height:1.2em;"><a href="/wiki/Types_of_business_entity" title="Types of
</th>
<td class="category" style="line-height:1.35em;">Subsidiary</td>
</tr>
```

First, let's try to select the element containing the variable name "Type".

```
In [30]: print(one_row.th)
         print("")
         print(one_row.th.div)
         print("")
         print(one_row.th.div.text)
```

```
<th scope="row" style="padding-right:0.5em;">
<div style="padding:0.1em 0;line-height:1.2em;"><a href="/wiki/Types_of_business_entity" title="Types of
</th>
```

```
<div style="padding:0.1em 0;line-height:1.2em;"><a href="/wiki/Types_of_business_entity" title="Types of
```

```
Type
```

Next, let's select the element containing the variable value, in this case "Subsidiary".

```
In [31]: print(one_row.td)
         print("")
         print(one_row.td.text)
```

```
<td class="category" style="line-height:1.35em;">Subsidiary</td>
```

```
Subsidiary
```

Now, let's loop through all rows to get all data that's available on the company. Depending on how well-structured the data is, this can be something of a trial and error process.

```
In [32]: for one_row in table_elements:
             print(one_row.th.div.text + ": " + one_row.td.text)
```

```
Type: Subsidiary
```

```
        ---------------------------------------------------------------------------

        AttributeError                            Traceback (most recent call last)

        <ipython-input-32-5678f42391a1> in <module>()
          1 for one_row in table_elements:
    ----> 2     print(one_row.th.div.text + ": " + one_row.td.text)


        AttributeError: 'NoneType' object has no attribute 'text'
```

We get an AttributeError for the "NoneType" object due to some of the "th" elements being empty. If we do some simple Exception capturing, we can get the loop to run through.

```
In [33]: for one_row in table_elements:
             try:
                 print(one_row.th.div.text + ": " + one_row.td.text)
             except Exception:
                 continue

Type: Subsidiary
Area served: Worldwide
Key people: David C. Scott
(President), (CEO) & (Director)
Operating income: US$ -3.33 million (FY10)
Net income: US$ -3.18 million (FY10)
Number of employees: 657 (FY10)
```

Let's try a challenge! Error handling - or having a computer program anticipate and respond to errors created by other functions - is a big part of programming. To give you a little more practice with this, we're going to have you team up with person sitting next to you and try challenge B in the challenges directory.

Now that we have the data we need, let's store it in a Python dictionary.

```
In [34]: new_dict = {}
         for one_row in table_elements:
             try:
                 print(one_row.th.div.text + ": " + one_row.td.text)
                 new_dict[one_row.th.div.text] = one_row.td.text
             except Exception:
                 continue

Type: Subsidiary
Area served: Worldwide
Key people: David C. Scott
(President), (CEO) & (Director)
Operating income: US$ -3.33 million (FY10)
Net income: US$ -3.18 million (FY10)
Number of employees: 657 (FY10)
```

We can browse the dictionary to make sure it is capturing the data correctly.

```
In [35]: print(new_dict.keys())
         print(new_dict)

dict_keys(['Operating income', 'Net income', 'Area served', 'Type', 'Number of employees', 'Key people']
{'Operating income': 'US$ -3.33 million (FY10)', 'Net income': 'US$ -3.18 million (FY10)', 'Area served
```

Let's quickly rehash what we did. We first extracted a list of hyperlinks from our index page, storing in our link_list variable. Next, we visited one of the pages, pulled its html into soup, and extracted the data from the element with class name "infobox vcard" into our new_dict variable.

The next step is to write an overall loop so that we can collect the "infobox vcard" data for all elements in our list. info_box = soup.find("table", {"class": "infobox vcard"})

```
In [36]: list_of_dicts = []

         for each_link in link_list[0:3]:
             print("")
             print(each_link)
             print("")
             company_page = requests.get("http://wikipedia.org" + each_link['href'])
```

```
        soup = BeautifulSoup(company_page.content)
        info_box = soup.find("table", {"class": "infobox vcard"})
        table_elements = info_box.find_all("tr")
        new_dict = {}
        new_dict['Company_name'] = each_link['title']
        for one_row in table_elements:
            try:
                print(one_row.th.div.text + ": " + one_row.td.text)
                # we convert to string as a precaution to make sure more complex elements are stor
                new_dict[one_row.th.div.text] = str(one_row.td.text)
            except Exception:
                continue
        # add the dictionary after we've added all variable names and values to each dictionary
        list_of_dicts.append(new_dict)

    print("")
    print(len(list_of_dicts))

<a href="/wiki/HP_3PAR" title="HP 3PAR">HP 3PAR</a>

Type: Subsidiary
Area served: Worldwide
Key people: David C. Scott
(President), (CEO) & (Director)
Operating income: US$ -3.33 million (FY10)
Net income: US$ -3.18 million (FY10)
Number of employees: 657 (FY10)

<a href="/wiki/4D_Inc" title="4D Inc">4D Inc</a>

Formerly called: ACI US
Type: Private
Areas served: USA & English-speaking Canada

<a href="/wiki/Achronix" title="Achronix">Achronix</a>

Type: Private
Key people: Robert Blake (CEO), John Lofton Holt (Chairman), Rahul Nimaiyyar (VP HW Engineering), Kamal
Number of employees: <200

3

/Users/dillon/anaconda/lib/python3.5/site-packages/bs4/__init__.py:166: UserWarning: No parser was explic

To get rid of this warning, change this:

 BeautifulSoup([your markup])

to this:

 BeautifulSoup([your markup], "lxml")

  markup_type=markup_type))
```

Let's browse our list_of_dicts object to make sure it contains the data we need.

```
In [37]: print(list_of_dicts)
```

```
[{'Operating income': 'US$ -3.33 million (FY10)', 'Net income': 'US$ -3.18 million (FY10)', 'Company_nam
```

   Typically, a "friendly" bot would try to space out the number of requests (in addition to not scraping the pages robots.txt disallows and obeying the general terms of service) so that the server can manage its traffic. One simple way to do this is to add a time.sleep command into your loop.
   As the entire class will be sharing the same IP, it's recommended that you add a longer wait time and limit the number of companies from link_list you scrape while in class.

```python
In [38]: import time

         list_of_dicts = []

         for each_link in link_list[0:3]:
             print("")
             wait_time = 1
             print('waiting ' + str(wait_time) + "s...")
             time.sleep(1)
             print("")
             print(each_link)
             print("")
             company_page = requests.get("http://wikipedia.org" + each_link['href'])
             soup = BeautifulSoup(company_page.content)
             info_box = soup.find("table", {"class": "infobox vcard"})
             table_elements = info_box.find_all("tr")
             new_dict = {}
             new_dict['Company_name'] = each_link['title']
             for one_row in table_elements:
                 try:
                     print(one_row.th.div.text + ": " + one_row.td.text)
                     # we convert to string as a precaution to make sure more complex elements are stor
                     new_dict[one_row.th.div.text] = str(one_row.td.text)
                 except Exception:
                     continue
             # add the dictionary after we've added all variable names and values to each dictionary
             list_of_dicts.append(new_dict)
```

```
waiting 1s...

<a href="/wiki/HP_3PAR" title="HP 3PAR">HP 3PAR</a>

Type: Subsidiary
Area served: Worldwide
Key people: David C. Scott
(President), (CEO) & (Director)
Operating income: US$ -3.33 million (FY10)
Net income: US$ -3.18 million (FY10)
Number of employees: 657 (FY10)

waiting 1s...

<a href="/wiki/4D_Inc" title="4D Inc">4D Inc</a>

Formerly called: ACI US
```

```
Type: Private
Areas served: USA & English-speaking Canada

waiting 1s...

<a href="/wiki/Achronix" title="Achronix">Achronix</a>

Type: Private
Key people: Robert Blake (CEO), John Lofton Holt (Chairman), Rahul Nimaiyyar (VP HW Engineering), Kamal
Number of employees: <200

/Users/dillon/anaconda/lib/python3.5/site-packages/bs4/__init__.py:166: UserWarning: No parser was explic

To get rid of this warning, change this:

 BeautifulSoup([your markup])

to this:

 BeautifulSoup([your markup], "lxml")

  markup_type=markup_type))
```

Now we have our list of dictionaries containing the data we want for each company. If you browse a few of the company pages, you'll notice that the number of variables each "infobox vcard" has is not always the same. Some dictionaries will have fields that others don't.

To store this data flexibly, we can iterate through all our dictionaries and collect all keys from them.

```
In [39]: key_list = []
         for each_dict in list_of_dicts:
             for key in each_dict.keys():
                 key_list.append(key)
         print(key_list)
```

```
['Operating income', 'Net income', 'Company name', 'Area served', 'Type', 'Number of employees', 'Key pe
```

Next, we convert the list to a set to remove all repeat keys. This then contains all unique keys across our dictionaries.

```
In [40]: key_set = set(key_list)
         print(key_set)
```

```
{'Area served', 'Formerly called', 'Areas served', 'Number of employees', 'Net income', 'Company name',
```

We convert the set back to a list (and sort it) as csv.DictWriter takes a list for its fieldnames parameter.

```
In [41]: final_key_list = sorted(list(key_set))
         print(final_key_list)
```

```
['Area served', 'Areas served', 'Company name', 'Formerly called', 'Key people', 'Net income', 'Number c
```

With our complete key list, we can now write our dictionary into a csv file.

```
In [42]: import csv

         outpath= "../data/02_companies.csv"
         outfile = open(outpath, 'w')
```

```
        writer = csv.DictWriter(outfile, fieldnames=final_key_list, dialect='excel')

        writer.writeheader()
        for row in list_of_dicts:
            writer.writerow(row)
            print(row)
        outfile.close()
        print("done")
```

{'Operating income': 'US$ -3.33 million (FY10)', 'Net income': 'US$ -3.18 million (FY10)', 'Company name
{'Formerly called': 'ACI US', 'Areas served': 'USA & English-speaking Canada', 'Type': 'Private', 'Compa
{'Type': 'Private', 'Number of employees': '<200', 'Key people': 'Robert Blake (CEO), John Lofton Holt
done

You can open the file in Excel to view the data. Congrats, you just scraped valuable data for your project off the web!

2 Creating data with web APIs

Most people who think they want to do web scraping actually want to pull data down from site-supplied APIs. Using an API is better in almost every way, and really the only reason to scrape data is if:

1. The website was constructed in the 90s and does not have an API; or,
2. You are doing something illegal

If LiveJournal has an API, the website you are interested in probably does too.

2.1 What is an API?

API is shorthand for **A**pplication **P**rogramming **I**nterface, which is in turn computer-ese for a middleman.

Think about it this way. You have a bunch of things on your computer that you want other people to be able to look at. Some of them are static documents, some of them call programs in real time, and some of them are programs themselves.

Solution 1 You publish login credentials on the internet, and let anyone log into your computer
Problems:

1. People will need to know how each document and program works to be able to access their data

2. You don't want the world looking at your browser history

Solution 2 You paste everything into HTML and publish it on the internet
Problems:

1. This can be information overload

2. Making things dynamic can be tricky

Solution 3 You create a set of methods to act as an intermediary between the people you want to help and the things you want them to have access to.
Why this is the best solution:

1. People only access what you want them to have, in the way that you want them to have it

2. People use one language to get the things they want

Why this is still not Panglossian:

1. You will have to explain to people how to use your middleman

2.2 Twitter's API

Twitter has an API - mostly written for third-party apps - that is comparatively straightforward and gives you access to *nearly* all of the information that Twitter has about its users, including:

1. User histories

2. User (and tweet) location

3. User language

4. Tweet popularity

5. Tweet spread

6. Conversation chains

Also, Twitter returns data to you in json, or **J**ava **S**cript **O**bject **N**otation. This is a very common format for passing data around http connections for browsers and servers, so many APIs return it as a datatype as well (instead of using something like xml or plain text).

Luckily, json converts into native Python data structures. Specifically, every json object you get from Twitter will be a combination of nested `dicts` and `lists`, which you learned about yesterday. This makes Twitter a lot easier to manipulate in Python than html objects, for example.

Here's what a tweet looks like:

```
In [43]: import json

         with open('../data/02_tweet.json','r') as f:
             a_tweet = json.loads(f.read())
```

We can take a quick look at the structure by pretty printing it:

```
In [44]: from pprint import pprint

         pprint(a_tweet)
```

```
{'contributors': None,
 'coordinates': None,
 'created_at': 'Thu Apr 02 06:09:39 +0000 2015',
 'entities': {'hashtags': [], 'symbols': [], 'urls': [], 'user_mentions': []},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 583511591334719488,
 'id_str': '583511591334719488',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'lang': 'ht',
 'place': None,
 'retweet_count': 0,
 'retweeted': False,
 'source': '<a href="http://twitter.com" rel="nofollow">Twitter Web Client</a>',
 'text': '.IPA rettiwT eht tuoba nraeL',
 'truncated': False,
 'user': {'contributors_enabled': False,
```

```
'created_at': 'Thu Apr 02 05:54:25 +0000 2015',
'default_profile': True,
'default_profile_image': False,
'description': '',
'entities': {'description': {'urls': []}},
'favourites_count': 0,
'follow_request_sent': False,
'followers_count': 0,
'following': False,
'friends_count': 0,
'geo_enabled': False,
'id': 3129088320,
'id_str': '3129088320',
'is_translation_enabled': False,
'is_translator': False,
'lang': 'en',
'listed_count': 0,
'location': '',
'name': 'Yelekreb Bald',
'notifications': False,
'profile_background_color': 'C0DEED',
'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png',
'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png',
'profile_background_tile': False,
'profile_image_url': 'http://pbs.twimg.com/profile_images/583509317476712449/mkd8KGeu_normal.jp
'profile_image_url_https': 'https://pbs.twimg.com/profile_images/583509317476712449/mkd8KGeu_no
'profile_link_color': '0084B4',
'profile_location': None,
'profile_sidebar_border_color': 'C0DEED',
'profile_sidebar_fill_color': 'DDEEF6',
'profile_text_color': '333333',
'profile_use_background_image': True,
'protected': False,
'screen_name': 'tob_pohskrow',
'statuses_count': 1,
'time_zone': None,
'url': None,
'utc_offset': None,
'verified': False}}
```

Time for a challenge!   Let's see how much you remember about lists and dicts from yesterday. Go into the challenges directory and try your hand at 02_scraping/C_json.py.

2.3   Authentication

Twitter controls access to their servers via a process of authentication and authorization. Authentication is how you let Twitter know who you are, in a way that is very hard to fake. Authorization is how the account owner (which will usually be yourself unless you are writing a Twitter app) controls what you are allowed to do in Twitter using their account. In Twitter, different levels of authorization require different levels of authentication.

Because we want to be able to interact with everything, we'll need the highest level of authorization and the strictest level of authentication. In Twitter, this means that we need two sets of ID's (called keys or tokens) and passwords (called secrets):

- consumer_key

- consumer_secret
- access_token_key
- access_token_secret

We'll provide some for you to use, but if you want to get your own you need to create an account on Twitter with a verified phone number. Then, while signed in to your Twitter account, go to: https://apps.twitter.com/. Follow the prompts to generate your keys and access tokens. Note that getting the second ID/password pair requires that you manually set the authorization level of your app.

We've stored our credentials in a separate file, which is smart. However, we have uploaded it to Github so that you have them too, which is not smart.

You should NEVER NEVER NEVER do this in real life.

We've stored it in YAML format, because it is more human-readible than JSON is. However, once it's inside Python, these data structures behave the same way.

```
In [45]: import yaml

         with open('../etc/creds.yml', 'r') as f:
             creds = yaml.load(f)
```

We're going to load these credentials into a requests module specifically designed for handling the flavor of authentication management that Twitter uses.

```
In [46]: from requests_oauthlib import OAuth1Session

         twitter = OAuth1Session(**creds)
```

That ** syntax we just used is called a "double splat" and is a python convenience function for converting the key-value pairs of a dictionary into keyword-argument pairs to pass to a function.

2.4 Accessing the API

Access to Twitter's API is organized through URLs called "endpoints". An endpoint is the location at which you can submit a request for Twitter to do something for you.

For example, the "endpoint" to search for specific kinds of tweets is at:

```
https://api.twitter.com/1.1/search/tweets.json
```

whereas posting new tweets is at:

```
https://api.twitter.com/1.1/statuses/update.json
```

For more information on the REST APIs, end points, and terms, check out: https://dev.twitter.com/rest/public. For the Streaming APIs: https://dev.twitter.com/streaming/overview.

All APIs on Twitter are "rate-limited" - this means that you are only allowed to ask a set number of questions per unit time (to keep their servers from being overloaded). This rate varies by endpoint and authorization, so be sure to check their developer site for the action you are trying to take.

For example, at the lowest level of authorization (Twitter calls this `application only`), you are allowed to make 450 search requests per 15 minute window, or about one every two seconds. At the highest level of authorization (Twitter calls this `user`) you can submit 180 requests every 15 minutes, or only about once every five seconds.

> side note - Google search is the worst rate-limiting I've ever seen, with an allowance of one hundred requests per day, or about once every *nine hundred seconds*

Let's try a couple of simple API queries. We're going to specify query parameters with `param`.

```
In [47]: search = "https://api.twitter.com/1.1/search/tweets.json"

         r = twitter.get(search, params={'q' : 'technology'})
```

This has returned an http response object, which contains data like whether or not the request succeeded:

```
In [48]: r.ok

Out[48]: True
```

You can also get the http response code, and the reason why Twitter sent you that code (these are all super important for controlling the flow of your program).

```
In [49]: r.status_code, r.reason

Out[49]: (200, 'OK')
```

The data that we asked Twitter to send us in r.content

```
In [50]: r.content
```

b'{"statuses":[{"metadata":{"iso_language_code":"en","result_type":"recent"},"created_at":"Tue Mar 15 22:25:06 +0000 2016","id":709868051307454465,"id_str":"709868051307454465"...

But that's not helpful. We can extract it in python's representation of json with the json method:

```
In [51]: r.json()

Out[51]: {'search_metadata': {'completed_in': 0.04,
           'count': 15,
           'max_id': 709868051307454465,
           'max_id_str': '709868051307454465',
           'next_results': '?max_id=709868034266157056&q=technology&include_entities=1',
           'query': 'technology',
           'refresh_url': '?since_id=709868051307454465&q=technology&include_entities=1',
           'since_id': 0,
           'since_id_str': '0'},
          'statuses': [{'contributors': None,
            'coordinates': None,
            'created_at': 'Tue Mar 15 22:25:06 +0000 2016',
            'entities': {'hashtags': [{'indices': [98, 109], 'text': 'Technology'}],
             'symbols': [],
             'urls': [{'display_url': 'huff.to/1U5Aurt',
               'expanded_url': 'http://huff.to/1U5Aurt',
               'indices': [74, 97],
               'url': 'https://t.co/E8yVunsXmX'}],
             'user_mentions': []},
            'favorite_count': 0,
            'favorited': False,
            'geo': None,
            'id': 709868051307454465,
            'id_str': '709868051307454465',
            'in_reply_to_screen_name': None,
            'in_reply_to_status_id': None,
            'in_reply_to_status_id_str': None,
            'in_reply_to_user_id': None,
            'in_reply_to_user_id_str': None,
            'is_quote_status': False,
```

'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': 'twitterfeed',
'text': "Inside Africa's Role in China's Ambitious 'One Belt, One Road' Initiative https://
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Thu Apr 16 21:20:37 +0000 2009',
 'default_profile': False,
 'default_profile_image': False,
 'description': "Everyone is SomeOne, but who's HOO and HOO are YOU? Promote and link yours
 'entities': {'description': {'urls': [{'display_url': 'somehoo.com',
     'expanded_url': 'http://somehoo.com',
     'indices': [86, 108],
     'url': 'http://t.co/5bkWs3BW7X'}]},
  'url': {'urls': [{'display_url': 'somehoo.com',
     'expanded_url': 'http://somehoo.com/',
     'indices': [0, 22],
     'url': 'http://t.co/y3ictymCcI'}]}},
 'favourites_count': 0,
 'follow_request_sent': False,
 'followers_count': 5054,
 'following': False,
 'friends_count': 5206,
 'geo_enabled': False,
 'has_extended_profile': False,
 'id': 32157561,
 'id_str': '32157561',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 147,
 'location': 'Scottsdale, AZ',
 'name': 'Social Media Who',
 'notifications': False,
 'profile_background_color': '9AE4E8',
 'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/38268241/Som
 'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/38268
 'profile_background_tile': True,
 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/32157561/1361067994',
 'profile_image_url': 'http://pbs.twimg.com/profile_images/224077644/SomeHooMini240_normal.p
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/224077644/SomeHooMini240_nc
 'profile_link_color': '0084B4',
 'profile_sidebar_border_color': 'BDDCAD',
 'profile_sidebar_fill_color': 'DDFFCC',
 'profile_text_color': '333333',
 'profile_use_background_image': True,
 'protected': False,
 'screen_name': 'SomeHoo',
 'statuses_count': 34116,
 'time_zone': 'Arizona',

      'url': 'http://t.co/y3ictymCcI',
      'utc_offset': -25200,
      'verified': False}},
{'contributors': None,
 'coordinates': None,
 'created_at': 'Tue Mar 15 22:25:06 +0000 2016',
 'entities': {'hashtags': [{'indices': [104, 114], 'text': 'MYTK4LIFE'}],
  'symbols': [],
  'urls': [{'display_url': 'tinyurl.com/zxjful6',
    'expanded_url': 'http://tinyurl.com/zxjful6',
    'indices': [80, 103],
    'url': 'https://t.co/WNRz66gCUZ'}],
  'user_mentions': []},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 709868051055792128,
 'id_str': '709868051055792128',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,
 'lang': 'en',
 'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
 'place': None,
 'possibly_sensitive': False,
 'retweet_count': 0,
 'retweeted': False,
 'source': 'twitterfeed',
 'text': 'George Takei, Sandra Oh Call Out Oscars For Racist Jokes, Academy Shrugs It Off ht
 'truncated': False,
 'user': {'contributors_enabled': False,
  'created_at': 'Mon Oct 10 20:55:40 +0000 2011',
  'default_profile': False,
  'default_profile_image': False,
  'description': 'To the young and the old, the blind and the dumb! To the hype and the meek
  'entities': {'description': {'urls': []},
   'url': {'urls': [{'display_url': 'about.me/mfm4urmind',
      'expanded_url': 'http://about.me/mfm4urmind',
      'indices': [0, 22],
      'url': 'http://t.co/xC5NM9wVLx'}]}},
  'favourites_count': 5548,
  'follow_request_sent': False,
  'followers_count': 2949,
  'following': False,
  'friends_count': 2438,
  'geo_enabled': False,
  'has_extended_profile': False,
  'id': 388495274,
  'id_str': '388495274',
  'is_translation_enabled': False,
  'is_translator': False,

     'lang': 'en',
     'listed_count': 116,
     'location': 'Concrete Jungle, Global',
     'name': 'Mohd Faidz Massari',
     'notifications': False,
     'profile_background_color': 'FFFFFF',
     'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/378800000114
     'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/37880
     'profile_background_tile': True,
     'profile_banner_url': 'https://pbs.twimg.com/profile_banners/388495274/1373385013',
     'profile_image_url': 'http://pbs.twimg.com/profile_images/3637317448/8e06c564529fc94a308a34
     'profile_image_url_https': 'https://pbs.twimg.com/profile_images/3637317448/8e06c564529fc94a
     'profile_link_color': '0084B4',
     'profile_sidebar_border_color': '000000',
     'profile_sidebar_fill_color': 'EFEFEF',
     'profile_text_color': '333333',
     'profile_use_background_image': True,
     'protected': False,
     'screen_name': 'MFM4URMIND',
     'statuses_count': 421722,
     'time_zone': 'Kuala Lumpur',
     'url': 'http://t.co/xC5NM9wVLx',
     'utc_offset': 28800,
     'verified': False}},
 {'contributors': None,
  'coordinates': None,
  'created_at': 'Tue Mar 15 22:25:06 +0000 2016',
  'entities': {'hashtags': [{'indices': [72, 79], 'text': 'staged'}],
   'media': [{'display_url': 'pic.twitter.com/Mw1qeIqSqs',
     'expanded_url': 'http://twitter.com/dinostylian/status/709868049923489796/photo/1',
     'id': 709868049118183424,
     'id_str': '709868049118183424',
     'indices': [80, 103],
     'media_url': 'http://pbs.twimg.com/media/Cdn1OskWwAAkJyL.jpg',
     'media_url_https': 'https://pbs.twimg.com/media/Cdn1OskWwAAkJyL.jpg',
     'sizes': {'large': {'h': 360, 'resize': 'fit', 'w': 480},
      'medium': {'h': 360, 'resize': 'fit', 'w': 480},
      'small': {'h': 255, 'resize': 'fit', 'w': 340},
      'thumb': {'h': 150, 'resize': 'crop', 'w': 150}},
     'type': 'photo',
     'url': 'https://t.co/Mw1qeIqSqs'}],
   'symbols': [],
   'urls': [{'display_url': 'vid.staged.com/GZjs',
     'expanded_url': 'http://vid.staged.com/GZjs',
     'indices': [48, 71],
     'url': 'https://t.co/PXeU2hN7vu'}],
   'user_mentions': []},
  'favorite_count': 0,
  'favorited': False,
  'geo': None,
  'id': 709868049923489796,
  'id_str': '709868049923489796',
  'in_reply_to_screen_name': None,
  'in_reply_to_status_id': None,

'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': 'Staged Marketing Toolsuite',
'text': 'The Bitcoin and Blockchain Technology Explained https://t.co/PXeU2hN7vu #staged ht
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Thu Jul 30 10:09:48 +0000 2015',
 'default_profile': True,
 'default_profile_image': False,
 'description': 'Want to earn #Bitcoin every day in your sleep? Click the link to get a #FRI
 'entities': {'description': {'urls': [{'display_url': 'bitclub.io/eagledriver',
     'expanded_url': 'http://bitclub.io/eagledriver',
     'indices': [88, 110],
     'url': 'http://t.co/7s6cIfL3tM'}]},
  'url': {'urls': [{'display_url': 'facebook.com/eagledriver8',
     'expanded_url': 'http://facebook.com/eagledriver8',
     'indices': [0, 22],
     'url': 'http://t.co/IM5PhR5jvp'}]}},
 'favourites_count': 35,
 'follow_request_sent': False,
 'followers_count': 2532,
 'following': False,
 'friends_count': 3421,
 'geo_enabled': False,
 'has_extended_profile': False,
 'id': 3395273235,
 'id_str': '3395273235',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 46,
 'location': 'worldwide',
 'name': 'Dino Stylianopoulos',
 'notifications': False,
 'profile_background_color': 'C0DEED',
 'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png',
 'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png',
 'profile_background_tile': False,
 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/3395273235/1438253069',
 'profile_image_url': 'http://pbs.twimg.com/profile_images/626696447249879040/cbHKJ3ZR_normal
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/626696447249879040/cbHKJ3ZR
 'profile_link_color': '0084B4',
 'profile_sidebar_border_color': 'C0DEED',
 'profile_sidebar_fill_color': 'DDEEF6',
 'profile_text_color': '333333',
 'profile_use_background_image': True,

```
     'protected': False,
     'screen_name': 'dinostylian',
     'statuses_count': 1777,
     'time_zone': None,
     'url': 'http://t.co/IM5PhR5jvp',
     'utc_offset': None,
     'verified': False}},
   {'contributors': None,
    'coordinates': None,
    'created_at': 'Tue Mar 15 22:25:06 +0000 2016',
    'entities': {'hashtags': [],
     'symbols': [],
     'urls': [{'display_url': 'huff.to/1U5Aurt',
       'expanded_url': 'http://huff.to/1U5Aurt',
       'indices': [114, 137],
       'url': 'https://t.co/usO8XqhXyZ'}],
     'user_mentions': []},
    'favorite_count': 0,
    'favorited': False,
    'geo': None,
    'id': 709868049516482562,
    'id_str': '709868049516482562',
    'in_reply_to_screen_name': None,
    'in_reply_to_status_id': None,
    'in_reply_to_status_id_str': None,
    'in_reply_to_user_id': None,
    'in_reply_to_user_id_str': None,
    'is_quote_status': False,
    'lang': 'en',
    'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
    'place': None,
    'possibly_sensitive': False,
    'retweet_count': 0,
    'retweeted': False,
    'source': '<a href="http://twitterfeed.com" rel="nofollow">twitterfeed</a>',
    'text': "Inside Africa's Role in China's Ambitious 'One Belt, One Road' Initiative:  China's
    'truncated': False,
    'user': {'contributors_enabled': False,
     'created_at': 'Thu Mar 04 18:14:14 +0000 2010',
     'default_profile': False,
     'default_profile_image': False,
     'description': 'Mr. KooLoDo : Journaliste | Blogueur | Webmaster | Freelancer| Community Ma
     'entities': {'description': {'urls': []},
      'url': {'urls': [{'display_url': 'ricardovintris.info',
        'expanded_url': 'http://ricardovintris.info/',
        'indices': [0, 22],
        'url': 'http://t.co/pLzfZjpoMP'}]}},
     'favourites_count': 16,
     'follow_request_sent': False,
     'followers_count': 4001,
     'following': False,
     'friends_count': 3364,
     'geo_enabled': False,
     'has_extended_profile': False,
```

```
        'id': 119807651,
        'id_str': '119807651',
        'is_translation_enabled': False,
        'is_translator': False,
        'lang': 'fr',
        'listed_count': 207,
        'location': 'Haiti',
        'name': 'Ricardo VINTRIS',
        'notifications': False,
        'profile_background_color': '3B94D9',
        'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/633318548169
        'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/63331
        'profile_background_tile': True,
        'profile_banner_url': 'https://pbs.twimg.com/profile_banners/119807651/1439829575',
        'profile_image_url': 'http://pbs.twimg.com/profile_images/2131762866/koolodo_normal.jpg',
        'profile_image_url_https': 'https://pbs.twimg.com/profile_images/2131762866/koolodo_normal.j
        'profile_link_color': '3B94D9',
        'profile_sidebar_border_color': '000000',
        'profile_sidebar_fill_color': '000000',
        'profile_text_color': '000000',
        'profile_use_background_image': True,
        'protected': False,
        'screen_name': 'mrkoolodo',
        'statuses_count': 234187,
        'time_zone': 'Central Time (US & Canada)',
        'url': 'http://t.co/pLzfZjpoMP',
        'utc_offset': -18000,
        'verified': False}},
 {'contributors': None,
  'coordinates': None,
  'created_at': 'Tue Mar 15 22:25:06 +0000 2016',
  'entities': {'hashtags': [],
   'media': [{'display_url': 'pic.twitter.com/l2cXtZDULY',
     'expanded_url': 'http://twitter.com/PetiotEric/status/709868048702967810/photo/1',
     'id': 709868048237273088,
     'id_str': '709868048237273088',
     'indices': [116, 139],
     'media_url': 'http://pbs.twimg.com/media/Cdn1OpSVIAANXvN.jpg',
     'media_url_https': 'https://pbs.twimg.com/media/Cdn1OpSVIAANXvN.jpg',
     'sizes': {'large': {'h': 388, 'resize': 'fit', 'w': 620},
      'medium': {'h': 375, 'resize': 'fit', 'w': 600},
      'small': {'h': 213, 'resize': 'fit', 'w': 340},
      'thumb': {'h': 150, 'resize': 'crop', 'w': 150}},
     'type': 'photo',
     'url': 'https://t.co/l2cXtZDULY'}],
   'symbols': [],
   'urls': [{'display_url': 'creativity-online.com/work/ibm-jacks',
     'expanded_url': 'http://creativity-online.com/work/ibm-jackson-an-ibm-data-story/45944',
     'indices': [92, 115],
     'url': 'https://t.co/N9wVv9QUpD'}],
   'user_mentions': []},
  'favorite_count': 0,
  'favorited': False,
  'geo': None,
```

'id': 709868048702967810,
'id_str': '709868048702967810',
'in_reply_to_screen_name': None,
'in_reply_to_status_id': None,
'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': 'Hootsuite',
'text': 'See How a Guide Dog Is Made - Thru a Guide Dogs Eyes\nEmotive Film on IBMs Data Te
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Thu May 31 11:14:09 +0000 2012',
 'default_profile': True,
 'default_profile_image': False,
 'description': '#Cloud Services Leader. #Digital Transformation - #Hybrid #Cognitive #Wats
 'entities': {'description': {'urls': []}},
 'favourites_count': 518,
 'follow_request_sent': False,
 'followers_count': 2344,
 'following': False,
 'friends_count': 2104,
 'geo_enabled': False,
 'has_extended_profile': False,
 'id': 595479894,
 'id_str': '595479894',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'fr',
 'listed_count': 453,
 'location': 'Paris France',
 'name': 'Eric Petiot',
 'notifications': False,
 'profile_background_color': 'C0DEED',
 'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png',
 'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png',
 'profile_background_tile': False,
 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/595479894/1448333572',
 'profile_image_url': 'http://pbs.twimg.com/profile_images/668977366807486464/jCJOnGUT_normal
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/668977366807486464/jCJOnGU
 'profile_link_color': '0084B4',
 'profile_sidebar_border_color': 'C0DEED',
 'profile_sidebar_fill_color': 'DDEEF6',
 'profile_text_color': '333333',
 'profile_use_background_image': True,
 'protected': False,
 'screen_name': 'PetiotEric',
 'statuses_count': 4662,

     'time_zone': 'Paris',
     'url': None,
     'utc_offset': 3600,
     'verified': False}},
  {'contributors': None,
   'coordinates': None,
   'created_at': 'Tue Mar 15 22:25:05 +0000 2016',
   'entities': {'hashtags': [{'indices': [0, 8], 'text': 'ITbyKKG'}],
    'symbols': [],
    'urls': [{'display_url': 'huff.to/1U5Atn7',
      'expanded_url': 'http://huff.to/1U5Atn7',
      'indices': [114, 137],
      'url': 'https://t.co/x8G75fCf1h'}],
    'user_mentions': []},
   'favorite_count': 0,
   'favorited': False,
   'geo': None,
   'id': 709868047746502656,
   'id_str': '709868047746502656',
   'in_reply_to_screen_name': None,
   'in_reply_to_status_id': None,
   'in_reply_to_status_id_str': None,
   'in_reply_to_user_id': None,
   'in_reply_to_user_id_str': None,
   'is_quote_status': False,
   'lang': 'en',
   'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
   'place': None,
   'possibly_sensitive': False,
   'retweet_count': 0,
   'retweeted': False,
   'source': 'twitterfeed',
   'text': "#ITbyKKG Inside Africa's Role in China's Ambitious 'One Belt, One Road' Initiative
   'truncated': False,
   'user': {'contributors_enabled': False,
    'created_at': 'Tue Aug 02 16:35:11 +0000 2011',
    'default_profile': False,
    'default_profile_image': False,
    'description': 'Official KKG Twitter page. Follow us for the latest tech news and company a
    'entities': {'description': {'urls': []},
     'url': {'urls': [{'display_url': 'mykkg.com',
        'expanded_url': 'http://www.mykkg.com',
        'indices': [0, 22],
        'url': 'http://t.co/9O01LLam4p'}]}},
    'favourites_count': 0,
    'follow_request_sent': False,
    'followers_count': 351,
    'following': False,
    'friends_count': 248,
    'geo_enabled': False,
    'has_extended_profile': False,
    'id': 347324889,
    'id_str': '347324889',
    'is_translation_enabled': False,

      'is_translator': False,
      'lang': 'en',
      'listed_count': 34,
      'location': 'Chicago, Illinois',
      'name': 'myKKG',
      'notifications': False,
      'profile_background_color': '131516',
      'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme14/bg.gif',
      'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme14/bg.gif',
      'profile_background_tile': True,
      'profile_banner_url': 'https://pbs.twimg.com/profile_banners/347324889/1397998513',
      'profile_image_url': 'http://pbs.twimg.com/profile_images/1486968716/KKG_Logo_Only_normal.jp
      'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1486968716/KKG_Logo_Only_no:
      'profile_link_color': '009999',
      'profile_sidebar_border_color': 'EEEEEE',
      'profile_sidebar_fill_color': 'EFEFEF',
      'profile_text_color': '333333',
      'profile_use_background_image': True,
      'protected': False,
      'screen_name': 'myKKGconsulting',
      'statuses_count': 41096,
      'time_zone': 'Central Time (US & Canada)',
      'url': 'http://t.co/9O01LLam4p',
      'utc_offset': -18000,
      'verified': False}},
 {'contributors': None,
  'coordinates': None,
  'created_at': 'Tue Mar 15 22:25:05 +0000 2016',
  'entities': {'hashtags': [{'indices': [58, 68], 'text': 'MYTK4LIFE'}],
   'symbols': [],
   'urls': [{'display_url': 'tinyurl.com/gpec8dl',
     'expanded_url': 'http://tinyurl.com/gpec8dl',
     'indices': [34, 57],
     'url': 'https://t.co/kaxci016vf'}],
   'user_mentions': []},
  'favorite_count': 0,
  'favorited': False,
  'geo': None,
  'id': 709868047704653824,
  'id_str': '709868047704653824',
  'in_reply_to_screen_name': None,
  'in_reply_to_status_id': None,
  'in_reply_to_status_id_str': None,
  'in_reply_to_user_id': None,
  'in_reply_to_user_id_str': None,
  'is_quote_status': False,
  'lang': 'en',
  'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
  'place': None,
  'possibly_sensitive': False,
  'retweet_count': 0,
  'retweeted': False,
  'source': 'twitterfeed',
  'text': 'Spring Cleaning Your Digital Life https://t.co/kaxci016vf #MYTK4LIFE',

    'truncated': False,
   'user': {'contributors_enabled': False,
    'created_at': 'Mon Oct 10 20:55:40 +0000 2011',
    'default_profile': False,
    'default_profile_image': False,
    'description': 'To the young and the old, the blind and the dumb! To the hype and the meek
    'entities': {'description': {'urls': []},
     'url': {'urls': [{'display_url': 'about.me/mfm4urmind',
        'expanded_url': 'http://about.me/mfm4urmind',
        'indices': [0, 22],
        'url': 'http://t.co/xC5NM9wVLx'}]}},
    'favourites_count': 5548,
    'follow_request_sent': False,
    'followers_count': 2949,
    'following': False,
    'friends_count': 2438,
    'geo_enabled': False,
    'has_extended_profile': False,
    'id': 388495274,
    'id_str': '388495274',
    'is_translation_enabled': False,
    'is_translator': False,
    'lang': 'en',
    'listed_count': 116,
    'location': 'Concrete Jungle, Global',
    'name': 'Mohd Faidz Massari',
    'notifications': False,
    'profile_background_color': 'FFFFFF',
    'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/378800000114
    'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/37880
    'profile_background_tile': True,
    'profile_banner_url': 'https://pbs.twimg.com/profile_banners/388495274/1373385013',
    'profile_image_url': 'http://pbs.twimg.com/profile_images/3637317448/8e06c564529fc94a308a34
    'profile_image_url_https': 'https://pbs.twimg.com/profile_images/3637317448/8e06c564529fc94a
    'profile_link_color': '0084B4',
    'profile_sidebar_border_color': '000000',
    'profile_sidebar_fill_color': 'EFEFEF',
    'profile_text_color': '333333',
    'profile_use_background_image': True,
    'protected': False,
    'screen_name': 'MFM4URMIND',
    'statuses_count': 421722,
    'time_zone': 'Kuala Lumpur',
    'url': 'http://t.co/xC5NM9wVLx',
    'utc_offset': 28800,
    'verified': False}},
  {'contributors': None,
   'coordinates': None,
   'created_at': 'Tue Mar 15 22:25:05 +0000 2016',
   'entities': {'hashtags': [{'indices': [74, 85], 'text': 'Technology'},
     {'indices': [86, 95], 'text': 'Facebook'}],
    'symbols': [],
    'urls': [],
    'user_mentions': [{'id': 40148479,

         'id_str': '40148479',
         'indices': [7, 16],
         'name': 'WhatsApp Inc.',
         'screen_name': 'WhatsApp'},
      {'id': 487312107,
         'id_str': '487312107',
         'indices': [36, 46],
         'name': 'WeChat',
         'screen_name': 'WeChatApp'},
      {'id': 487312107,
         'id_str': '487312107',
         'indices': [62, 72],
         'name': 'WeChat',
         'screen_name': 'WeChatApp'}]},
'favorite_count': 0,
'favorited': False,
'geo': None,
'id': 709868046882439172,
'id_str': '709868046882439172',
'in_reply_to_screen_name': None,
'in_reply_to_status_id': None,
'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': {'attributes': {},
  'bounding_box': {'coordinates': [[[144.593741856, -38.433859306],
      [145.512528832, -38.433859306],
      [145.512528832, -37.5112737225],
      [144.593741856, -37.5112737225]]],
    'type': 'Polygon'},
  'contained_within': [],
  'country': 'Australia',
  'country_code': 'AU',
  'full_name': 'Melbourne, Victoria',
  'id': '01864a8a64df9dc4',
  'name': 'Melbourne',
  'place_type': 'city',
  'url': 'https://api.twitter.com/1.1/geo/id/01864a8a64df9dc4.json'},
'retweet_count': 0,
'retweeted': False,
'source': '<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</
'text': 'I hope @WhatsApp have Stickers like @WeChatApp but now I Love @WeChatApp #Technolo
'truncated': False,
'user': {'contributors_enabled': False,
  'created_at': 'Sat Aug 08 17:37:01 +0000 2009',
  'default_profile': False,
  'default_profile_image': False,
  'description': '#Digital #Business #Analyst, @MonashUni, @WMU #Tourist, #Photographer,#Foo
  'entities': {'description': {'urls': []}},
  'favourites_count': 94,
  'follow_request_sent': False,

'followers_count': 213,
 'following': False,
 'friends_count': 417,
 'geo_enabled': True,
 'has_extended_profile': True,
 'id': 64006146,
 'id_str': '64006146',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 34,
 'location': 'Melbourne',
 'name': '@AlindTiwari',
 'notifications': False,
 'profile_background_color': 'FFF04D',
 'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme19/bg.gif',
 'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme19/bg.gif',
 'profile_background_tile': False,
 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/64006146/1427513907',
 'profile_image_url': 'http://pbs.twimg.com/profile_images/622946215559237633/9yWKrckf_normal
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/622946215559237633/9yWKrckf
 'profile_link_color': '0099CC',
 'profile_sidebar_border_color': 'FFF8AD',
 'profile_sidebar_fill_color': 'F6FFD1',
 'profile_text_color': '333333',
 'profile_use_background_image': True,
 'protected': False,
 'screen_name': 'alindtiwari',
 'statuses_count': 1549,
 'time_zone': 'Eastern Time (US & Canada)',
 'url': None,
 'utc_offset': -14400,
 'verified': False}},
{'contributors': None,
 'coordinates': None,
 'created_at': 'Tue Mar 15 22:25:05 +0000 2016',
 'entities': {'hashtags': [{'indices': [0, 8], 'text': 'ITbyKKG'}],
 'symbols': [],
 'urls': [{'display_url': 'huff.to/1U5Atn6',
   'expanded_url': 'http://huff.to/1U5Atn6',
   'indices': [114, 137],
   'url': 'https://t.co/upTPrewaV5'}],
 'user_mentions': []},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 709868046878318593,
 'id_str': '709868046878318593',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,

'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': 'twitterfeed',
'text': '#ITbyKKG Instagram May Change Your Feed, Personalizing It With An Algorithm: For y
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Tue Aug 02 16:35:11 +0000 2011',
 'default_profile': False,
 'default_profile_image': False,
 'description': 'Official KKG Twitter page. Follow us for the latest tech news and company a
 'entities': {'description': {'urls': []},
  'url': {'urls': [{'display_url': 'mykkg.com',
     'expanded_url': 'http://www.mykkg.com',
     'indices': [0, 22],
     'url': 'http://t.co/9Oo1LLam4p'}]}},
 'favourites_count': 0,
 'follow_request_sent': False,
 'followers_count': 351,
 'following': False,
 'friends_count': 248,
 'geo_enabled': False,
 'has_extended_profile': False,
 'id': 347324889,
 'id_str': '347324889',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 34,
 'location': 'Chicago, Illinois',
 'name': 'myKKG',
 'notifications': False,
 'profile_background_color': '131516',
 'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme14/bg.gif',
 'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme14/bg.gif',
 'profile_background_tile': True,
 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/347324889/1397998513',
 'profile_image_url': 'http://pbs.twimg.com/profile_images/1486968716/KKG_Logo_Only_normal.jp
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1486968716/KKG_Logo_Only_no
 'profile_link_color': '009999',
 'profile_sidebar_border_color': 'EEEEEE',
 'profile_sidebar_fill_color': 'EFEFEF',
 'profile_text_color': '333333',
 'profile_use_background_image': True,
 'protected': False,
 'screen_name': 'myKKGconsulting',
 'statuses_count': 41096,
 'time_zone': 'Central Time (US & Canada)',
 'url': 'http://t.co/9Oo1LLam4p',
 'utc_offset': -18000,
 'verified': False}},

{'contributors': None,
 'coordinates': None,
 'created_at': 'Tue Mar 15 22:25:05 +0000 2016',
 'entities': {'hashtags': [{'indices': [0, 8], 'text': 'ITbyKKG'}],
  'symbols': [],
  'urls': [{'display_url': 'huff.to/1YXynF4',
    'expanded_url': 'http://huff.to/1YXynF4',
    'indices': [114, 137],
    'url': 'https://t.co/SjddZ4U8Fe'}],
  'user_mentions': []},
 'favorite_count': 0,
 'favorited': False,
 'geo': None,
 'id': 709868045968084993,
 'id_str': '709868045968084993',
 'in_reply_to_screen_name': None,
 'in_reply_to_status_id': None,
 'in_reply_to_status_id_str': None,
 'in_reply_to_user_id': None,
 'in_reply_to_user_id_str': None,
 'is_quote_status': False,
 'lang': 'en',
 'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
 'place': None,
 'possibly_sensitive': False,
 'retweet_count': 0,
 'retweeted': False,
 'source': 'twitterfeed',
 'text': "#ITbyKKG The Genius of Go Grandmaster Lee Se-dol Made AlphaGo's Performance That Mu
 'truncated': False,
 'user': {'contributors_enabled': False,
  'created_at': 'Tue Aug 02 16:35:11 +0000 2011',
  'default_profile': False,
  'default_profile_image': False,
  'description': 'Official KKG Twitter page. Follow us for the latest tech news and company a
  'entities': {'description': {'urls': []},
   'url': {'urls': [{'display_url': 'mykkg.com',
     'expanded_url': 'http://www.mykkg.com',
     'indices': [0, 22],
     'url': 'http://t.co/9O01LLam4p'}]}},
  'favourites_count': 0,
  'follow_request_sent': False,
  'followers_count': 351,
  'following': False,
  'friends_count': 248,
  'geo_enabled': False,
  'has_extended_profile': False,
  'id': 347324889,
  'id_str': '347324889',
  'is_translation_enabled': False,
  'is_translator': False,
  'lang': 'en',
  'listed_count': 34,
  'location': 'Chicago, Illinois',

    'name': 'myKKG',
    'notifications': False,
    'profile_background_color': '131516',
    'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme14/bg.gif',
    'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme14/bg.gif',
    'profile_background_tile': True,
    'profile_banner_url': 'https://pbs.twimg.com/profile_banners/347324889/1397998513',
    'profile_image_url': 'http://pbs.twimg.com/profile_images/1486968716/KKG_Logo_Only_normal.jp
    'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1486968716/KKG_Logo_Only_no
    'profile_link_color': '009999',
    'profile_sidebar_border_color': 'EEEEEE',
    'profile_sidebar_fill_color': 'EFEFEF',
    'profile_text_color': '333333',
    'profile_use_background_image': True,
    'protected': False,
    'screen_name': 'myKKGconsulting',
    'statuses_count': 41096,
    'time_zone': 'Central Time (US & Canada)',
    'url': 'http://t.co/9O01LLam4p',
    'utc_offset': -18000,
    'verified': False}},
 {'contributors': None,
  'coordinates': None,
  'created_at': 'Tue Mar 15 22:25:04 +0000 2016',
  'entities': {'hashtags': [{'indices': [76, 87], 'text': 'technology'}],
   'symbols': [],
   'urls': [{'display_url': 'dlvr.it/KnpZr2',
     'expanded_url': 'http://dlvr.it/KnpZr2',
     'indices': [51, 74],
     'url': 'https://t.co/AEoa5IrEl3'}],
   'user_mentions': []},
  'favorite_count': 0,
  'favorited': False,
  'geo': None,
  'id': 709868042759442432,
  'id_str': '709868042759442432',
  'in_reply_to_screen_name': None,
  'in_reply_to_status_id': None,
  'in_reply_to_status_id_str': None,
  'in_reply_to_user_id': None,
  'in_reply_to_user_id_str': None,
  'is_quote_status': False,
  'lang': 'en',
  'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
  'place': None,
  'possibly_sensitive': False,
  'retweet_count': 0,
  'retweeted': False,
  'source': 'dlvr.it',
  'text': 'Why the most popular gadget at SXSW was a pencil https://t.co/AEoa5IrEl3 #technol
  'truncated': False,
  'user': {'contributors_enabled': False,
   'created_at': 'Sat Feb 20 17:41:20 +0000 2016',
   'default_profile': True,

      'default_profile_image': False,
      'description': 'Everything happens for a reason. ~Unknown',
      'entities': {'description': {'urls': []},
       'url': {'urls': [{'display_url': 'RebelMouse.com/ViralRangers/',
          'expanded_url': 'https://www.RebelMouse.com/ViralRangers/',
          'indices': [0, 23],
          'url': 'https://t.co/NQ6xMpio9u'}]}},
      'favourites_count': 14,
      'follow_request_sent': False,
      'followers_count': 672,
      'following': False,
      'friends_count': 683,
      'geo_enabled': False,
      'has_extended_profile': False,
      'id': 701099327360364544,
      'id_str': '701099327360364544',
      'is_translation_enabled': False,
      'is_translator': False,
      'lang': 'en',
      'listed_count': 2,
      'location': 'Saban Castil',
      'name': 'urho hentinen',
      'notifications': False,
      'profile_background_color': 'F5F8FA',
      'profile_background_image_url': None,
      'profile_background_image_url_https': None,
      'profile_background_tile': False,
      'profile_banner_url': 'https://pbs.twimg.com/profile_banners/701099327360364544/1457299779'
      'profile_image_url': 'http://pbs.twimg.com/profile_images/704773905957781504/xkRoKXwp_normal
      'profile_image_url_https': 'https://pbs.twimg.com/profile_images/704773905957781504/xkRoKXwp
      'profile_link_color': '2B7BB9',
      'profile_sidebar_border_color': 'C0DEED',
      'profile_sidebar_fill_color': 'DDEEF6',
      'profile_text_color': '333333',
      'profile_use_background_image': True,
      'protected': False,
      'screen_name': 'viralrangers',
      'statuses_count': 273,
      'time_zone': None,
      'url': 'https://t.co/NQ6xMpio9u',
      'utc_offset': None,
      'verified': False}},
 {'contributors': None,
  'coordinates': None,
  'created_at': 'Tue Mar 15 22:25:04 +0000 2016',
  'entities': {'hashtags': [],
   'symbols': [],
   'urls': [{'display_url': 'ift.tt/1YXxZpX',
     'expanded_url': 'http://ift.tt/1YXxZpX',
     'indices': [111, 134],
     'url': 'https://t.co/MrMYKbXtoQ'}],
   'user_mentions': []},
  'favorite_count': 0,
  'favorited': False,

'geo': None,
'id': 709868040935116801,
'id_str': '709868040935116801',
'in_reply_to_screen_name': None,
'in_reply_to_status_id': None,
'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': True,
'retweet_count': 0,
'retweeted': False,
'source': 'IFTTT',
'text': '"Apple Files Last Volley Against F.B.I. Before a Hearing Next Week" by KATIE BENNE
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Fri Feb 20 15:03:06 +0000 2009',
 'default_profile': False,
 'default_profile_image': False,
 'description': 'Civilization & Discontent',
 'entities': {'description': {'urls': []}},
 'favourites_count': 1071,
 'follow_request_sent': False,
 'followers_count': 1836,
 'following': False,
 'friends_count': 1858,
 'geo_enabled': False,
 'has_extended_profile': False,
 'id': 21402821,
 'id_str': '21402821',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'de',
 'listed_count': 54,
 'location': 'Hamburg Germany',
 'name': 'Frank Walle',
 'notifications': False,
 'profile_background_color': '099C5A',
 'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/789658228/84
 'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/78965
 'profile_background_tile': False,
 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/21402821/1446070262',
 'profile_image_url': 'http://pbs.twimg.com/profile_images/659493294011977730/V7cNqJ7E_normal
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/659493294011977730/V7cNqJ7E
 'profile_link_color': 'ABB8C2',
 'profile_sidebar_border_color': '000000',
 'profile_sidebar_fill_color': 'FFFFFF',
 'profile_text_color': '1D1B1B',
 'profile_use_background_image': True,
 'protected': False,
 'screen_name': 'eWebjojo',

```
                           'statuses_count': 48572,
                           'time_zone': 'Berlin',
                           'url': None,
                           'utc_offset': 3600,
                           'verified': False}},
                        {'contributors': None,
                         'coordinates': None,
                         'created_at': 'Tue Mar 15 22:25:04 +0000 2016',
                         'entities': {'hashtags': [],
                          'symbols': [],
                          'urls': [{'display_url': 'huff.to/1U5Aurt',
                            'expanded_url': 'http://huff.to/1U5Aurt',
                            'indices': [74, 97],
                            'url': 'https://t.co/QKiFwAQFck'}],
                          'user_mentions': []},
                         'favorite_count': 0,
                         'favorited': False,
                         'geo': None,
                         'id': 709868040477773824,
                         'id_str': '709868040477773824',
                         'in_reply_to_screen_name': None,
                         'in_reply_to_status_id': None,
                         'in_reply_to_status_id_str': None,
                         'in_reply_to_user_id': None,
                         'in_reply_to_user_id_str': None,
                         'is_quote_status': False,
                         'lang': 'en',
                         'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
                         'place': None,
                         'possibly_sensitive': False,
                         'retweet_count': 0,
                         'retweeted': False,
                         'source': '<a href="http://twitterfeed.com" rel="nofollow">twitterfeed</a>',
                         'text': "Inside Africa's Role in China's Ambitious 'One Belt, One Road' Initiative https://
                         'truncated': False,
                         'user': {'contributors_enabled': False,
                          'created_at': 'Thu Jan 24 08:07:20 +0000 2008',
                          'default_profile': False,
                          'default_profile_image': False,
                          'description': 'msj blogger/healer/hoper of hope',
                          'entities': {'description': {'urls': []},
                           'url': {'urls': [{'display_url': 'themomstreetjournal.com',
                              'expanded_url': 'http://www.themomstreetjournal.com',
                              'indices': [0, 22],
                              'url': 'http://t.co/xhkdzsTlaj'}]}},
                          'favourites_count': 86,
                          'follow_request_sent': False,
                          'followers_count': 1206,
                          'following': False,
                          'friends_count': 1107,
                          'geo_enabled': False,
                          'has_extended_profile': False,
                          'id': 12630972,
                          'id_str': '12630972',
```

     'is_translation_enabled': False,
     'is_translator': False,
     'lang': 'en',
     'listed_count': 59,
     'location': 'omniverse',
     'name': 'Lori Martin Gregory',
     'notifications': False,
     'profile_background_color': 'C0DEED',
     'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/579767389790
     'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/57976
     'profile_background_tile': False,
     'profile_banner_url': 'https://pbs.twimg.com/profile_banners/12630972/1445710982',
     'profile_image_url': 'http://pbs.twimg.com/profile_images/656839511687585792/5wjUbtEO_normal
     'profile_image_url_https': 'https://pbs.twimg.com/profile_images/656839511687585792/5wjUbtEO
     'profile_link_color': 'DD2E44',
     'profile_sidebar_border_color': 'FFFFFF',
     'profile_sidebar_fill_color': 'DDEEF6',
     'profile_text_color': '333333',
     'profile_use_background_image': True,
     'protected': False,
     'screen_name': 'LoriGregory',
     'statuses_count': 61847,
     'time_zone': 'Pacific Time (US & Canada)',
     'url': 'http://t.co/xhkdzsTlaj',
     'utc_offset': -25200,
     'verified': False}},
   {'contributors': None,
    'coordinates': None,
    'created_at': 'Tue Mar 15 22:25:03 +0000 2016',
    'entities': {'hashtags': [{'indices': [62, 65], 'text': 'in'},
      {'indices': [66, 71], 'text': 'tech'}],
     'symbols': [],
     'urls': [{'display_url': 'bbc.in/1nMuyVr',
       'expanded_url': 'http://bbc.in/1nMuyVr',
       'indices': [38, 61],
       'url': 'https://t.co/ntaam63ycm'}],
     'user_mentions': []},
    'favorite_count': 0,
    'favorited': False,
    'geo': None,
    'id': 709868039236362240,
    'id_str': '709868039236362240',
    'in_reply_to_screen_name': None,
    'in_reply_to_status_id': None,
    'in_reply_to_status_id_str': None,
    'in_reply_to_user_id': None,
    'in_reply_to_user_id_str': None,
    'is_quote_status': False,
    'lang': 'en',
    'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
    'place': None,
    'possibly_sensitive': False,
    'retweet_count': 0,
    'retweeted': False,

'source': 'IFTTT',
'text': 'PlayStation VR is cheaper than rivals https://t.co/ntaam63ycm #in #tech',
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Mon Mar 10 12:56:41 +0000 2008',
 'default_profile': False,
 'default_profile_image': False,
 'description': '#HouseDJ | Web Architect @UnitedUtilities, Agile UX Consultant @TheTomTayl
 'entities': {'description': {'urls': [{'display_url': 'tommytaylor.co.uk/#terms',
     'expanded_url': 'http://tommytaylor.co.uk/#terms',
     'indices': [124, 147],
     'url': 'https://t.co/GOOoYCHtg6'}]},
  'url': {'urls': [{'display_url': 'tommytaylor.co.uk',
     'expanded_url': 'http://tommytaylor.co.uk',
     'indices': [0, 23],
     'url': 'https://t.co/Mjzcv6w02b'}]}},
 'favourites_count': 453,
 'follow_request_sent': False,
 'followers_count': 1242,
 'following': False,
 'friends_count': 1069,
 'geo_enabled': True,
 'has_extended_profile': False,
 'id': 14113270,
 'id_str': '14113270',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 82,
 'location': 'Warrington, UK',
 'name': 'Tom Taylor',
 'notifications': False,
 'profile_background_color': '022330',
 'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/378800000010
 'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/37880
 'profile_background_tile': False,
 'profile_banner_url': 'https://pbs.twimg.com/profile_banners/14113270/1372454498',
 'profile_image_url': 'http://pbs.twimg.com/profile_images/378800000699229454/03975af39acb2d
 'profile_image_url_https': 'https://pbs.twimg.com/profile_images/378800000699229454/03975af3
 'profile_link_color': '0084B4',
 'profile_sidebar_border_color': 'FFFFFF',
 'profile_sidebar_fill_color': 'C0DFEC',
 'profile_text_color': '333333',
 'profile_use_background_image': True,
 'protected': False,
 'screen_name': 'tom_taylor',
 'statuses_count': 24941,
 'time_zone': 'London',
 'url': 'https://t.co/Mjzcv6w02b',
 'utc_offset': 0,
 'verified': False}},
{'contributors': None,
 'coordinates': None,
 'created_at': 'Tue Mar 15 22:25:02 +0000 2016',

'entities': {'hashtags': [{'indices': [78, 90], 'text': 'attemployee'}],
 'media': [{'display_url': 'pic.twitter.com/LTjn9COuph',
   'expanded_url': 'http://twitter.com/qtip55/status/709868034266157057/photo/1',
   'id': 709868034110922752,
   'id_str': '709868034110922752',
   'indices': [91, 114],
   'media_url': 'http://pbs.twimg.com/media/Cdn1N0qWIAAcQD8.jpg',
   'media_url_https': 'https://pbs.twimg.com/media/Cdn1N0qWIAAcQD8.jpg',
   'sizes': {'large': {'h': 683, 'resize': 'fit', 'w': 1024},
    'medium': {'h': 400, 'resize': 'fit', 'w': 600},
    'small': {'h': 227, 'resize': 'fit', 'w': 340},
    'thumb': {'h': 150, 'resize': 'crop', 'w': 150}},
   'type': 'photo',
   'url': 'https://t.co/LTjn9COuph'}],
 'symbols': [],
 'urls': [{'display_url': 'soc.att.com/1WoKkSy',
   'expanded_url': 'http://soc.att.com/1WoKkSy',
   'indices': [54, 77],
   'url': 'https://t.co/OZMXTD13yO'}],
 'user_mentions': []},
'favorite_count': 0,
'favorited': False,
'geo': None,
'id': 709868034266157057,
'id_str': '709868034266157057',
'in_reply_to_screen_name': None,
'in_reply_to_status_id': None,
'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': 'SocialChorus',
'text': '4 Ways to Advance your Career Using Mobile Technology https://t.co/OZMXTD13yO #atte
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Thu Apr 22 23:34:00 +0000 2010',
 'default_profile': False,
 'default_profile_image': False,
 'description': 'Love Star Trek; Live for Vacations; Never leave home without my Kindle; AT&
 'entities': {'description': {'urls': []}},
 'favourites_count': 3,
 'follow_request_sent': False,
 'followers_count': 94,
 'following': False,
 'friends_count': 58,
 'geo_enabled': False,
 'has_extended_profile': False,
 'id': 136069549,

```
                   'id_str': '136069549',
                   'is_translation_enabled': False,
                   'is_translator': False,
                   'lang': 'en',
                   'listed_count': 6,
                   'location': 'Texas',
                   'name': 'eliska paratore',
                   'notifications': False,
                   'profile_background_color': '0099B9',
                   'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme4/bg.gif',
                   'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme4/bg.gif',
                   'profile_background_tile': False,
                   'profile_image_url': 'http://pbs.twimg.com/profile_images/3659479729/9f22e4732fdb76d8552ddf
                   'profile_image_url_https': 'https://pbs.twimg.com/profile_images/3659479729/9f22e4732fdb76d8
                   'profile_link_color': '0099B9',
                   'profile_sidebar_border_color': '5ED4DC',
                   'profile_sidebar_fill_color': '95E8EC',
                   'profile_text_color': '3C3940',
                   'profile_use_background_image': True,
                   'protected': False,
                   'screen_name': 'qtip55',
                   'statuses_count': 379,
                   'time_zone': 'Central Time (US & Canada)',
                   'url': None,
                   'utc_offset': -18000,
                   'verified': False}}]}
```

This has some helpful metadata about our request, like a url where we can get the next batch of results from Twitter for the same query:

```
In [52]: data = r.json()
         data['search_metadata']

Out[52]: {'completed_in': 0.04,
          'count': 15,
          'max_id': 709868051307454465,
          'max_id_str': '709868051307454465',
          'next_results': '?max_id=709868034266157056&q=technology&include_entities=1',
          'query': 'technology',
          'refresh_url': '?since_id=709868051307454465&q=technology&include_entities=1',
          'since_id': 0,
          'since_id_str': '0'}
```

The tweets that we want are under the key "statuses"

```
In [53]: statuses = data['statuses']
         statuses[0]

Out[53]: {'contributors': None,
          'coordinates': None,
          'created_at': 'Tue Mar 15 22:25:06 +0000 2016',
          'entities': {'hashtags': [{'indices': [98, 109], 'text': 'Technology'}],
           'symbols': [],
           'urls': [{'display_url': 'huff.to/1U5Aurt',
             'expanded_url': 'http://huff.to/1U5Aurt',
```

        'indices': [74, 97],
        'url': 'https://t.co/E8yVunsXmX'}],
 'user_mentions': []},
'favorite_count': 0,
'favorited': False,
'geo': None,
'id': 7098680051307454465,
'id_str': '7098680051307454465',
'in_reply_to_screen_name': None,
'in_reply_to_status_id': None,
'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': None,
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': 'twitterfeed',
'text': "Inside Africa's Role in China's Ambitious 'One Belt, One Road' Initiative https://t.c
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Thu Apr 16 21:20:37 +0000 2009',
 'default_profile': False,
 'default_profile_image': False,
 'description': "Everyone is SomeOne, but who's HOO and HOO are YOU? Promote and link yoursel
 'entities': {'description': {'urls': [{'display_url': 'somehoo.com',
      'expanded_url': 'http://somehoo.com',
      'indices': [86, 108],
      'url': 'http://t.co/5bkWs3BW7X'}]},
  'url': {'urls': [{'display_url': 'somehoo.com',
      'expanded_url': 'http://somehoo.com/',
      'indices': [0, 22],
      'url': 'http://t.co/y3ictymCcI'}]}},
 'favourites_count': 0,
 'follow_request_sent': False,
 'followers_count': 5054,
 'following': False,
 'friends_count': 5206,
 'geo_enabled': False,
 'has_extended_profile': False,
 'id': 32157561,
 'id_str': '32157561',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 147,
 'location': 'Scottsdale, AZ',
 'name': 'Social Media Who',
 'notifications': False,
 'profile_background_color': '9AE4E8',
 'profile_background_image_url': 'http://pbs.twimg.com/profile_background_images/38268241/SomeH

```
'profile_background_image_url_https': 'https://pbs.twimg.com/profile_background_images/3826824
'profile_background_tile': True,
'profile_banner_url': 'https://pbs.twimg.com/profile_banners/32157561/1361067994',
'profile_image_url': 'http://pbs.twimg.com/profile_images/224077644/SomeHooMini240_normal.png
'profile_image_url_https': 'https://pbs.twimg.com/profile_images/224077644/SomeHooMini240_norm
'profile_link_color': '0084B4',
'profile_sidebar_border_color': 'BDDCAD',
'profile_sidebar_fill_color': 'DDFFCC',
'profile_text_color': '333333',
'profile_use_background_image': True,
'protected': False,
'screen_name': 'SomeHoo',
'statuses_count': 34116,
'time_zone': 'Arizona',
'url': 'http://t.co/y3ictymCcI',
'utc_offset': -25200,
'verified': False}}
```

This is one tweet.

> Depending on which tweet this is, you may or may not see that Twitter automatically pulls out links and mentions and gives you their index location in the raw tweet string

Twitter gives you a whole lot of information about their users, including geographical coordinates, the device they are tweeting from, and links to their photographs.

Twitter supports what it calls query operators, which modify the search behavior. For example, if you want to search for tweets where a particular user is mentioned, include the at-sign, `@`, followed by the username. To search for tweets sent to a particular user, use `to:username`. For tweets from a particular user, `from:username`. For hashtags, use `#hashtag`.

For a complete set of options: https://dev.twitter.com/rest/public/search.

Let's try a more complicated search:

```
In [54]: r = twitter.get(search, params={
                'q' : 'happy',
                'geocode' : '37.8734855,-122.2597169,10mi'
            })
         r.ok

Out[54]: True

In [55]: statuses = r.json()['statuses']
         statuses[0]

Out[55]: {'contributors': None,
          'coordinates': {'coordinates': [-122.4303894, 37.7853546], 'type': 'Point'},
          'created_at': 'Tue Mar 15 22:02:42 +0000 2016',
          'entities': {'hashtags': [],
           'symbols': [],
           'urls': [{'display_url': 'instagram.com/p/BC_V6V4St_Y/',
             'expanded_url': 'https://www.instagram.com/p/BC_V6V4St_Y/',
             'indices': [53, 76],
             'url': 'https://t.co/DmIu08U1XK'}],
           'user_mentions': []},
          'favorite_count': 0,
          'favorited': False,
          'geo': {'coordinates': [37.7853546, -122.4303894], 'type': 'Point'},
```

'id': 709862414163058690,
'id_str': '709862414163058690',
'in_reply_to_screen_name': None,
'in_reply_to_status_id': None,
'in_reply_to_status_id_str': None,
'in_reply_to_user_id': None,
'in_reply_to_user_id_str': None,
'is_quote_status': False,
'lang': 'en',
'metadata': {'iso_language_code': 'en', 'result_type': 'recent'},
'place': {'attributes': {},
 'bounding_box': {'coordinates': [[[-122.514926, 37.708075],
    [-122.357031, 37.708075],
    [-122.357031, 37.833238],
    [-122.514926, 37.833238]]],
  'type': 'Polygon'},
 'contained_within': [],
 'country': 'United States',
 'country_code': 'US',
 'full_name': 'San Francisco, CA',
 'id': '5a110d312052166f',
 'name': 'San Francisco',
 'place_type': 'city',
 'url': 'https://api.twitter.com/1.1/geo/id/5a110d312052166f.json'},
'possibly_sensitive': False,
'retweet_count': 0,
'retweeted': False,
'source': 'Instagram',
'text': 'Cute happy_face_crepe @ Belly Good Cafe & Crepes https://t.co/DmIu08U1XK',
'truncated': False,
'user': {'contributors_enabled': False,
 'created_at': 'Mon Aug 10 05:53:30 +0000 2009',
 'default_profile': True,
 'default_profile_image': False,
 'description': "Windmills do not work that way!\r\n\r\nI like food, music and travel. I work
 'entities': {'description': {'urls': []}},
 'favourites_count': 568,
 'follow_request_sent': False,
 'followers_count': 387,
 'following': False,
 'friends_count': 611,
 'geo_enabled': True,
 'has_extended_profile': False,
 'id': 64345702,
 'id_str': '64345702',
 'is_translation_enabled': False,
 'is_translator': False,
 'lang': 'en',
 'listed_count': 20,
 'location': 'San Francisco',
 'name': 'Laurence Berland',
 'notifications': False,
 'profile_background_color': 'C0DEED',
 'profile_background_image_url': 'http://abs.twimg.com/images/themes/theme1/bg.png',

```
                    'profile_background_image_url_https': 'https://abs.twimg.com/images/themes/theme1/bg.png',
                    'profile_background_tile': False,
                    'profile_banner_url': 'https://pbs.twimg.com/profile_banners/64345702/1357553872',
                    'profile_image_url': 'http://pbs.twimg.com/profile_images/1346029743/macchiato_normal.jpg',
                    'profile_image_url_https': 'https://pbs.twimg.com/profile_images/1346029743/macchiato_normal.j
                    'profile_link_color': '0084B4',
                    'profile_sidebar_border_color': 'C0DEED',
                    'profile_sidebar_fill_color': 'DDEEF6',
                    'profile_text_color': '333333',
                    'profile_use_background_image': True,
                    'protected': False,
                    'screen_name': 'laurenceb',
                    'statuses_count': 7603,
                    'time_zone': 'Pacific Time (US & Canada)',
                    'url': None,
                    'utc_offset': -25200,
                    'verified': False}}
```

If we want to store this data somewhere, we can output it as json using the json library from above. However, if you're doing a lot of these, you'll probaby want to use a database to handle everything.

```
In [56]: with open('my_tweets.json', 'w') as f:
             json.dump(statuses, f)
```

To post tweets, we need to use a different endpoint:

```
In [57]: post = "https://api.twitter.com/1.1/statuses/update.json"
```

And now we can pass a new tweet (remember, Twitter calls these 'statuses') as a parameter to our post request.

```
In [58]: r = twitter.post(post, params={
                 'status' : "I stole Juan's Twitter credentials"
             })
         r.ok
```

```
Out[58]: False
```

Other (optional) parameters include things like location, and replies.

2.5   Scheduling

The real beauty of bots is that they are designed to work without interaction or oversight. Imagine a situation where you want to automatically retweet everything coming out of the D-Lab's twitter account, "@DLabAtBerkeley". You could:

1. spend the rest of your life glued to D-Lab's twitter page and hitting refresh; or,
2. write a function

We're going to import a module called `time` that will pause our code, so that we don't hit Twitter's rate limit

```
In [59]: import time

         def retweet():
             r = twitter.get(search, {'q':'DLabAtBerkeley'})
             if r.ok:
```

```
            statuses = r.json()['statuses']
            for update in statuses:
                username = item['user']['screen_name']
                parameters = {'status':'HOORAY! @' + username}
                r = twitter.post(post, parameters)
                print(r.status_code, r.reason)
                time.sleep(5)
```

But you are a human that needs to eat, sleep, and be social with other humans. Luckily, Linux systems have a time-based daemon called **cron** that will run scripts like this *for you*.

People on windows and macs will not be able to run this. That's okay.

The way that **cron** works is it reads in files where each line has a time followed by a job (these are called cronjobs). You can edit your crontab by typing **crontab -e** into a terminal.
They looks like this:

```
In [60]: with open('../etc/crontab_example', 'r') as f:
             print(f.read())

# In a user's crontab, jobs run under that user
# Time is specified as <min> <hour> <day> <month> <wday>
# To specify any time, use '*'
# For unknown reasons, cronjobs fail unless the tab ends with a newline

00 08 * * 1 echo "It is 8am on Monday" >> /var/dumblog
```

This is telling **cron** to print that statement to a file called "dumblog" at 8am every Monday.

It's generally frowned upon to enter jobs through crontabs because they are hard to modify without breaking them. The better solution is to put your timed command into a file and copy the file into **/etc/cron.d/**. These files look like this:

```
In [61]: with open('../etc/crond_example', 'r') as f:
             print(f.read())

#!/bin/bash
# First, make sure you specify all of the paths that you might need to run
# your task. If you aren't sure, copy the entire $PATH variable

PATH=/home/dillon/.conda/envs/py27/bin:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin

# Then, specify when you want the task to occur; the user account to run it;
# and the job

@hourly dillon cd ~/scripts; python simple.py
```

At this point, you might be a little upset that you can't do this on your laptop, but the truth is you don't really want to run daemons and cronjobs on your laptop, which goes to sleep and runs out of batteries. This is what servers are for (like AWS).

2.6 Now it is time for you to make your own twitter bot!

To get you started, we've put a template in the **scripts** folder. Try it out, but be generous with your **time.sleep()** calls as the whole class is sharing this account.

If you have tried to run this, or some of the earlier code in this notebook, you have probably encountered some of Twitter's error codes. Here are the most common, and why you are triggering them.

1. `400 = bad request` - This means the API (middleman) doesn't like how you formatted your request. Check the API documentation to make sure you are doing things correctly.

2. `401 = unauthorized` - This either means you entered your auth codes incorrectly, or those auth codes don't have permission to do what you're trying to do. It takes Twitter a while to assign posting rights to your auth tokens after you've given them your phone number. If you have just done this, wait five minutes, then try again.

3. `403 = forbidden` - Twitter won't let you post what you are trying to post, most likely because you are trying to post the same tweet twice in a row within a few minutes of each other. Try changing your status update. If that doesn't fix it, then you are either:

   A. Hitting Twitter's daily posting limit. They don't say what this is.

   B. Trying to follow too many people, rapidly following and unfollowing the same person, or are otherwise making Twitter think you are a spambot

4. `429 = too many requests` - This means that you have exceeded Twitter's rate limit for whatever it is you are trying to do. Increase your `time.sleep()` value.

2.7   Considerate robots and legality

Typically, in starting a new web scraping project, you'll want to follow these steps:
1) Find the websites' robots.txt and do not access those pages through your bot
2) Make sure your bot does not make too many requests in a specific period (etc. by using Python's sleep.wait function)
3) Look up the website's term of use or terms of service.
   We'll discuss each of these briefly.

2.7.1   What data owners care about

Data owners are concerned with:
1) Keeping their website up
2) Protecting the commercial value of their data
   Their policies and responses differ with respect to these two areas. You'll need to do some research to determine what is appropriate with regards to your research.

1) Keeping their website up   Most commercial websites have strategies to throttle or block IPs that make too many requests within a fixed amount of time. Because a bot can make a large number of requests in a small amount of time (etc. entering 100 different terms into Google in one second), servers are able to determine if traffic is coming from a bot or a person (among many other methods). For companies that rely on advertising, like Google or Twitter, these requests do not represent "human eyeballs" and need to be filtered out from their bill to advertisers.
   In order to keep their site up and running, companies may block your IP temporarily or permanently if they detect too many requests coming from your IP, or other signs that requests are being made by a bot instead of a person. If you systematically down a site (such as sending millions of requests to an official government site), there is the small chance your actions may be interpreted maliciously (and regarded as hacking), with risk of prosecution.

2) Protecting the commercial value of their data   Companies are also typically very protective of their data, especially data that ties directly into how they make money. A listings site (like Craigslist), for instance, would lose traffic if listings on its site were poached and transfered to a competitor, or if a rival company used scraping tools to derive lists of users to contact. For this reason, companies' term of use agreements are typically very restrictive of what you can do with their data.
   Different companies may have a range of responses to your scraping, depending on what you do with the data. Typically, repurposing the data for a rival application or business will trigger a strong response from

the company (i.e. legal attention). Publishing any analysis or results, either in a formal academic journal or on a blog or webpage, may be of less concern, though legal attention is still possible.

2.7.2   robots.txt: internet convention

The robots.txt file is typically located in the root folder of the site, with instructions to various services (User-agents) on what they are not allowed to scrape.

Typically, the robots.txt file is more geared towards search engines (and their crawlers) more than anything else.

However, companies and agencies typically will not want you to scrape any pages that they disallow search engines from accessing. Scraping these pages makes it more likely for your IP to be detected and blocked (along with other possible actions.)

Below is an example of reddit's robots.txt file: https://www.reddit.com/robots.txt

3   80legs

User-agent: 008 Disallow: /
   User-Agent: bender Disallow: /my_shiny_metal_ass
   User-Agent: Gort Disallow: /earth
   User-Agent: *
Disallow: /*.json
Disallow: /*.json-compact
Disallow: /*.json-html
Disallow: /*.xml
Disallow: /*.rss
Disallow: /*.i
Disallow: /*.embed
Disallow: /*/comments/*?sort=
Disallow: /r/*/comments/*/*/c
Disallow: /comments/*/*c*
Disallow: /r/*/submit
Disallow: /message/compose
Disallow: /api
Disallow: /post
Disallow: /submit
Disallow: /goto
Disallow: /*after=
Disallow: /*before=
Disallow: /domain/*t=
Disallow: /login
Disallow: /reddits/search
Disallow: /search
Disallow: /r/*/search
Allow: /
   User blahblahblah provides a concise description of how to read the robots.txt file: https://www.reddit.com/r/learnprogramming/comments/3l1lcq/how_do_you_find_out_if_a_website_is_scrapable/

- The bot that calls itself 008 (apparently from 80legs) isn't allowed to access anything
- bender is not allowed to visit my_shiny_metal_ass (it's a Futurama joke, the page doesn't actually exist)
- Gort isn't allowed to visit Earth (another joke, from The Day the Earth Stood Still)
- Other scrapers should avoid checking the API methods or "compose message" or 'search" or the "over 18?" page (because those aren't something you really want showing up in Google), but they're allowed to visit anything else.

In general, your bot will fall into the * wildcard category of what the site generally do not want bots to access. You should make sure your scraper does not access any of those pages, etc. www.reddit.com/login etc.

In [62]: