ML_8_nltk

March 26, 2019

```
In [1]: """In this assignment students have to find the frequency of words in a webpage. User
        use urllib and BeautifulSoup to extract text from webpage.

        Hint:
        from bs4 import BeautifulSoup
        import urllib.request
        import nltk

        response = urllib.request.urlopen('http://php.net/')
        html = response.read()
        soup = BeautifulSoup(html,"html5lib")"""

        # Importing and reading the text from the webpage

        import numpy as np
        import pandas as pd
        from bs4 import BeautifulSoup
        import urllib.request
        import nltk

        response = urllib.request.urlopen('http://php.net/')
        html = response.read()
        soup = BeautifulSoup(html,"html5lib")
        text = soup.get_text(strip=True)
        print (text)
```

```
PHP: Hypertext PreprocessorDownloadsDocumentationGet InvolvedHelpGetting StartedIntroductionA
       7.1.27. This is a security release which also contains several bug fixes.All PHP 7.1 use
       Windows source and binaries can be found onwindows.php.net/download/.
       The list of changes is recorded in theChangeLog.07 Mar 2019PHP 7.2.16 ReleasedThe PHP de
     7.2.16. This is a security release which also contains several minor bug fixes.All PHP 7.
     Windows source and binaries can be found onwindows.php.net/download/.
     The list of changes is recorded in theChangeLog.07 Mar 2019PHP 7.3.3 ReleasedThe PHP devel
      7.3.3. This is a security release which also contains several bug fixes.All PHP 7.3 users
      Windows source and binaries can be found onwindows.php.net/download/.
      The list of changes is recorded in theChangeLog.07 Feb 2019PHP 7.2.15 ReleasedThe PHP dev
      This is a bugfix release.All PHP 7.2 users are encouraged to upgrade to this version.For
      Windows source and binaries can be found onwindows.php.net/download/.
```

The list of changes is recorded in theChangeLog.07 Feb 2019PHP 7.3.2 Release Announcement
7.3.2. This is a bugfix release, with several bug fixes included.All PHP 7.3 users are en
Windows source and binaries can be found onwindows.php.net/download/.
The list of changes is recorded in theChangeLog.10 Jan 2019PHP 5.6.40 ReleasedThe PHP dev
5.6.40. This is a security release. Several security bugs have been fixed
in this release.

All PHP 5.6 users are encouraged to upgrade to this version.For source downloads of PHP 5
Windows source and binaries can be found onwindows.php.net/download/.
The list of changes is recorded in theChangeLog.Please note that according to thePHP versi
support timelines,
PHP 5.6.40 is the last scheduled release of PHP 5.6 branch. There may be additional rel
discover
important security issues that warrant it, otherwise this release will be the final one
5.6 branch.
If your PHP installation is based on PHP 5.6, it may be a good time to start making the
the upgrade
to PHP 7.1, PHP 7.2 or PHP 7.3.22 Nov 2018PHP 7.3.0RC6 ReleasedThe PHP team is glad to
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source d
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test t
or theUPGRADINGfile for a complete list of upgrading notes. Internal changes are listed
These files can also be found in the release archive.The next release would be 7.3.0 (G
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source d
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test t
or theUPGRADINGfile for a complete list of upgrading notes. Internal changes are listed
These files can also be found in the release archive.The next release would be RC6, pla
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source d
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test t
or theUPGRADINGfile for a complete list of upgrading notes. Internal changes are listed
These files can also be found in the release archive.The next release would be RC5, pla
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source d
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test t
or theUPGRADINGfile for a complete list of upgrading notes. Internal changes are listed
These files can also be found in the release archive.The next release would be RC4, pla
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source d
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test t
or theUPGRADINGfile for a complete list of upgrading notes. Internal changes are listed
These files can also be found in the release archive.The next release would be RC3, pla
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source d
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test t
or theUPGRADINGfile for a complete list of upgrading notes. Internal changes are listed
These files can also be found in the release archive.The next release would be RC2, pla
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source d
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test t
or theUPGRADINGfile for a complete list of upgrading notes. Internal changes are listed
These files can also be found in the release archive.The next release would be RC1, pla
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source d
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test t

or theUPGRADINGfile for a complete list of upgrading notes. Internal changes are listed
These files can also be found in the release archive.The next release would be Beta 3,
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source d
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test t
or theUPGRADINGfile for a complete list of upgrading notes. These files can also be fou
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source d
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test t
or theUPGRADINGfile for a complete list of upgrading notes. These files can also be fou
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source dow
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test thi
or theUPGRADINGfile for a complete list of upgrading notes. These files can also be found
The rough outline of the PHP 7.3 release cycle is specified in thePHP Wiki.For source dow
Windows sources and binaries can be found onwindows.php.net/qa/.Please carefully test thi
or theUPGRADINGfile for a complete list of upgrading notes. These files can also be found
This starts the PHP 7.3 release cycle, the rough outline of which is specified in theP
or theUPGRADINGfile for a complete list of upgrading notes. These files can also b
7.2.2. This is a bugfix release, with several bug fixes included.All PHP 7.2 users are en
Windows source and binaries can be found onwindows.php.net/download/.
The list of changes is recorded in theChangeLog.12 Oct 2017PHP 7.2.0 Release Candidate 4
This release is the fourth Release Candidate for 7.2.0.
All users of PHP are encouraged to test this version carefully, and report any bugs
and incompatibilities in thebug tracking system.THIS IS A DEVELOPMENT PREVIEW - DO NOT USE
or theUPGRADINGfile for a complete list of upgrading notes. These files can also be found
Windows sources and binaries can be found atwindows.php.net/qa/.The next Release Candidate
You can also read the full list of planned releases onour wiki.Thank you for helping us ma
This release is the third Release Candidate for 7.2.0.
All users of PHP are encouraged to test this version carefully, and report any bugs
and incompatibilities in thebug tracking system.THIS IS A DEVELOPMENT PREVIEW - DO NOT USE
or theUPGRADINGfile for a complete list of upgrading notes. These files can also be found
Windows sources and binaries can be found atwindows.php.net/qa/.The next Release Candidate
You can also read the full list of planned releases onour wiki.Thank you for helping us ma
Candidate 1. This release is the first Release Candidate for 7.2.0.
All users of PHP are encouraged to test this version carefully, and report any bugs
and incompatibilities in thebug tracking system.THIS IS A DEVELOPMENT PREVIEW - DO NOT US
or theUPGRADINGfile for a complete list of upgrading notes. These files can also be found
Windows sources and binaries can be found atwindows.php.net/qa/.The second Release Candid
You can also read the full list of planned releases onour wiki.Thank you for helping us m
This release is the third and final beta for 7.2.0. All users of PHP are encouraged
to test this version carefully, and report any bugs and incompatibilities in thebug track
or theUPGRADINGfile for a complete list of upgrading notes. These files can also be found
Windows sources and binaries can be found atwindows.php.net/qa/.The first Release Candida
You can also read the full list of planned releases onour wiki.Thank you for helping us m
This release contains fixes and improvements relative to Alpha 2.
All users of PHP are encouraged to test this version carefully,
and report any bugs and incompatibilities in thebug tracking system.THIS IS A DEVELOPMENT
or theUPGRADINGfile
for a complete list of upgrading notes. These files can also be found in the release archi
Windows sources and binaries can be found onwindows.php.net/qa/.The first beta will be rel

```
In [2]: # Tokenization

        tokens = [t for t in text.split()]
        print (tokens)
```

```
['PHP:', 'Hypertext', 'PreprocessorDownloadsDocumentationGet', 'InvolvedHelpGetting', 'StartedI
```

```
In [3]: # Distribution of frequency from nltk.FreqDist()

        freq = nltk.FreqDist(tokens)
        for key,val in freq.items():
            print (str(key) + ':' + str(val))
```

```
PHP::1
Hypertext:1
PreprocessorDownloadsDocumentationGet:1
InvolvedHelpGetting:1
StartedIntroductionA:1
simple:1
tutorialLanguage:1
ReferenceBasic:1
syntaxTypesVariablesConstantsExpressionsOperatorsControl:1
StructuresFunctionsClasses:1
and:77
ObjectsNamespacesErrorsExceptionsGeneratorsReferences:1
ExplainedPredefined:1
VariablesPredefined:1
ExceptionsPredefined:1
Interfaces:1
ClassesContext:1
options:1
parametersSupported:1
Protocols:1
WrappersSecurityIntroductionGeneral:1
considerationsInstalled:1
as:2
CGI:1
binaryInstalled:1
an:2
Apache:1
moduleSession:1
SecurityFilesystem:1
SecurityDatabase:1
SecurityError:1
ReportingUsing:1
Register:1
```

```
GlobalsUser:1
Submitted:1
DataMagic:1
QuotesHiding:1
PHPKeeping:1
CurrentFeaturesHTTP:1
authentication:1
with:4
PHPCookiesSessionsDealing:1
XFormsHandling:1
file:1
uploadsUsing:1
remote:1
filesConnection:1
handlingPersistent:1
Database:1
ConnectionsSafe:1
ModeCommand:1
line:1
usageGarbage:1
CollectionDTrace:1
Dynamic:1
TracingFunction:1
ReferenceAffecting:1
PHP's:1
BehaviourAudio:1
Formats:1
ManipulationAuthentication:1
ServicesCommand:1
Line:1
Specific:2
ExtensionsCompression:1
Archive:1
ExtensionsCredit:1
Card:1
ProcessingCryptography:1
ExtensionsDatabase:1
ExtensionsDate:1
Time:1
Related:4
ExtensionsFile:1
System:1
ExtensionsHuman:1
Language:1
Character:1
Encoding:1
SupportImage:1
Processing:1
```

```
GenerationMail:1
ExtensionsMathematical:1
ExtensionsNon-Text:1
MIME:1
OutputProcess:1
Control:1
ExtensionsOther:2
Basic:1
ServicesSearch:1
Engine:1
ExtensionsServer:1
ExtensionsSession:1
ExtensionsText:1
ProcessingVariable:1
Type:1
ExtensionsWeb:1
ServicesWindows:1
Only:1
ExtensionsXML:1
ManipulationGUI:1
ExtensionsKeyboard:1
Shortcuts?This:1
helpjNext:1
menu:2
itemkPrevious:1
itemg:1
pPrevious:1
man:2
pageg:1
nNext:1
pageGScroll:1
to:40
bottomg:1
gScroll:1
topg:1
hGoto:1
homepageg:1
sGoto:1
search(current:1
page)/Focus:1
search:1
boxPHP:1
is:49
a:27
popular:2
general-purpose:1
scripting:1
language:1
```

```
that:3
especially:1
suited:1
web:1
development.Fast,:1
flexible:1
pragmatic,:1
PHP:141
powers:1
everything:1
from:1
your:2
blog:1
the:116
most:1
websites:1
in:68
world.Download7.1.27ůRelease:1
NotesůUpgrading7.2.16ůRelease:1
NotesůUpgrading7.3.3ůRelease:1
NotesůUpgrading07:1
Mar:3
2019PHP:7
7.1.27:2
ReleasedThe:23
development:12
team:25
announces:12
immediate:12
availability:12
of:99
7.1.27.:1
This:13
security:6
release:76
which:4
also:26
contains:4
several:5
bug:5
fixes.All:3
7.1:1
users:12
are:20
encouraged:12
upgrade:8
this:28
version.For:8
```

```
source:32
downloads:25
please:25
visit:25
ourdownloads:7
page,:7
Windows:24
binaries:24
can:78
be:76
found:68
onwindows.php.net/download/.:7
The:20
list:30
changes:15
recorded:7
theChangeLog.07:4
7.2.16:2
7.2.16.:1
minor:1
7.2:4
7.3.3:2
7.3.3.:1
7.3:15
Feb:3
7.2.15:2
7.2.15.:1
bugfix:3
release.All:1
7.3.2:2
Release:15
AnnouncementThe:1
7.3.2.:1
release,:2
fixes:3
included.All:2
theChangeLog.10:1
Jan:1
5.6.40:3
5.6.40.:1
release.:2
Several:1
bugs:6
have:1
been:1
fixed:1
All:6
5.6:3
```

```
theChangeLog.Please:1
note:1
according:1
thePHP:14
version:20
support:1
timelines,:1
last:2
scheduled:1
branch.:2
There:1
may:2
additional:1
if:1
we:1
discover:1
important:1
issues:14
warrant:1
it,:1
otherwise:1
will:6
final:2
one:1
If:1
installation:1
based:1
on:25
5.6,:1
it:2
good:1
time:1
start:1
making:1
plans:1
for:68
7.1,:1
or:19
7.3.22:1
Nov:2
2018PHP:14
7.3.0RC6:2
glad:13
announce:13
presumably:1
7.3.0:23
pre-release,:6
7.3.0RC6.:1
```

```
rough:13
outline:13
cycle:12
specified:13
Wiki.For:13
thedownload:13
page.:12
sources:17
onwindows.php.net/qa/.Please:12
carefully:13
test:19
report:18
any:18
thebug:18
reporting:13
system.THIS:17
IS:17
A:17
DEVELOPMENT:17
PREVIEW:17
-:17
DO:18
NOT:18
USE:17
IT:17
IN:17
PRODUCTION!For:17
more:17
information:18
new:18
features:18
other:18
changes,:18
you:36
read:23
theNEWSfile,:18
theUPGRADINGfile:18
complete:18
upgrading:18
notes.:18
Internal:8
listed:8
theUPGRADING.INTERNALSfile.:8
These:18
files:18
archive.The:13
next:20
would:13
```

```
(GA),:1
planned:18
December:1
6th.The:1
signatures:13
inthe:13
manifestor:13
onthe:13
QA:13
site.Thank:13
helping:18
us:18
make:18
better.08:1
7.3.0RC5:2
7.3.0RC5.:1
RC6,:1
November:2
22nd.The:1
better.25:1
Oct:3
7.3.0RC4:2
7.3.0RC4.:1
RC5,:1
8th.The:1
better.11:1
7.3.0RC3:2
7.3.0RC3.:1
RC4,:1
October:2
25th.The:1
better.28:2
Sep:3
7.3.0RC2:2
7.3.0RC2.:1
RC3,:1
11th.The:1
better.13:1
7.3.0RC1:2
7.3.0RC1.:1
RC2,:1
September:2
27th.The:1
better.30:1
Aug:5
7.3.0.beta3:1
seventh:1
version,:7
```

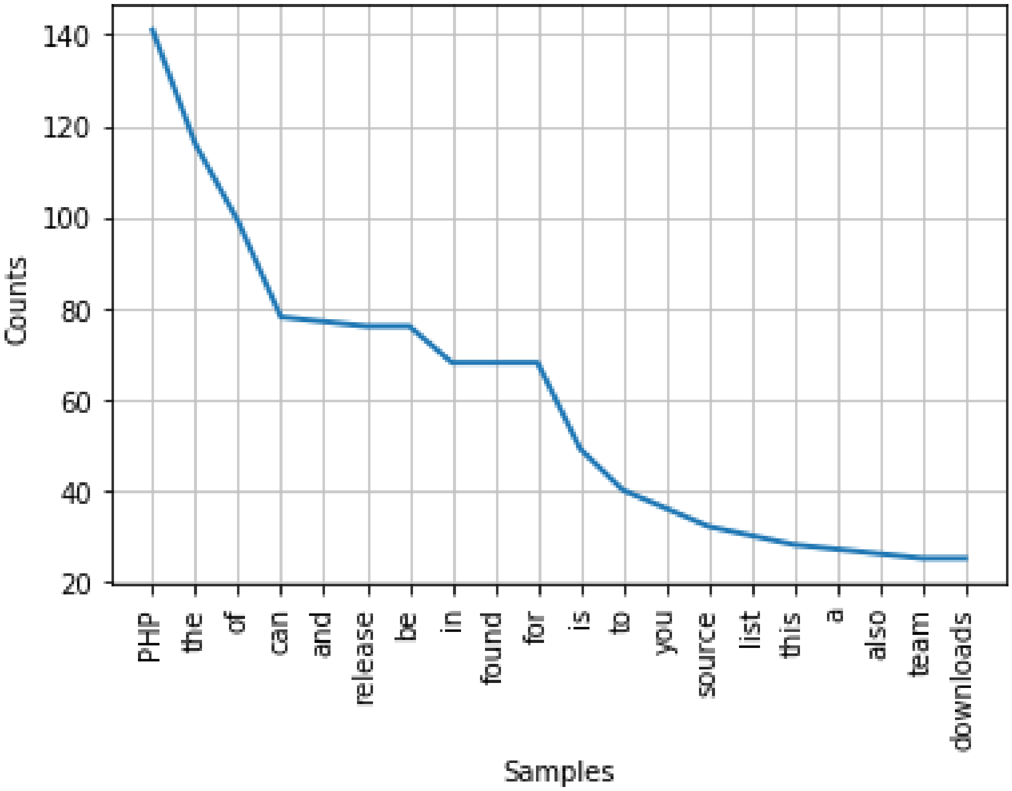
```
7.3.0beta3.:1
7.3.0beta3:1
RC1,:1
13th.The:1
better.16:1
7.3.0.beta2:1
sixth:1
7.3.0beta2.:1
7.3.0beta2:1
Beta:7
3,:2
August:3
30th.The:1
better.02:1
7.3.0.beta1:1
fifth:1
7.3.0beta1.:1
7.3.0beta1:1
2,:2
16th.The:1
better.19:1
Jul:3
7.3.0alpha4:2
fourth:2
7.3.0alpha4.:1
1,:2
2nd.The:1
better.05:1
alpha:3
3:8
third:3
Alpha:12
3.:3
July:2
19th.The:1
better.21:1
Jun:2
2:2
second:2
2.:2
5.The:1
better.07:1
1:4
ReleasedPHP:1
first:4
1.:2
starts:1
cycle,:1
```

```
page.Please:1
system.Please:1
use:1
production,:1
early:1
June:1
21.The:1
better.01:1
7.2.2:2
7.2.2.:1
theChangeLog.12:1
2017PHP:5
7.2.0:15
Candidate:14
4:2
RC4.:1
7.2.0.:4
carefully,:5
incompatibilities:5
tracking:5
archive.For:5
thedownloadpage,:5
atwindows.php.net/qa/.The:4
announced:2
26th:1
October.:2
You:5
full:5
releases:5
onour:4
wiki.Thank:4
RC3.:1
12th:1
better.31:1
released:3
14th:1
September.:1
better.17:1
beta:2
31th:1
August.:1
better.06:1
improvements:1
relative:1
onwindows.php.net/qa/.The:1
20th:1
July.:1
ourwiki.Thank:1
```

```
better.Older:1
News:1
EntriesConferences:1
calling:1
papersMid-Atlantic:1
Developer:1
ConferenceBulgaria:1
Conference:1
2019ScotlandPHP:1
2019Upcoming:1
conferencesPHPerKaigi:1
Russia:1
2019Web:1
Summer:1
Camp:1
2019Laracon:1
EU:1
2019:1
MadridUser:1
Group:1
EventsSpecial:1
ThanksSocial:1
media@official_phpCopyright:1
I’:1
2001-2019:1
GroupMy:1
PHP.netContactOther:1
PHP.net:1
sitesMirror:1
sitesPrivacy:1
policy:1
```

```
In [5]: # VIsualization of frequency distribution

        import matplotlib.pyplot as plt
        freq.plot(20, cumulative=False)
        plt.show()
```


Counts
140
120
100
80
60
40
20
PHP
the
of
can
and
release
be
in
found
for
is
to
you
source
list
this
a
also
team
downloads
Samples