

Activity 1: Introduction to NLP basic techniques

Instructions:

- Please download the provided IPython Notebook (ipynb) file and open it in Google Colab. Once opened, enter your code in the same file directly beneath the relevant question's code block.
- Purpose of this activity is to practice and get hands on experience (Ungraded Activity)

Fundamentals of NLP: Tokenization

Natural language processing (NLP) has made substantial advances in the past few years due to the success of [modern techniques](#) that are based on [deep learning](#). With the rise of the popularity of NLP and the availability of different forms of large-scale data, it is now even more imperative to understand the inner workings of NLP techniques and concepts, from first principles, as they find their way into real-world usage and applications that affect society at large. Building intuitions and having a solid grasp of concepts are both important for coming up with innovative techniques, improving research, and building safe, human-centered AI and NLP technologies.

In this first chapter, which is part of a series called **Fundamentals of NLP**, we will learn about some of the most important **basic concepts** that power NLP techniques used for research and building real-world applications. Some of these techniques include *lemmatization*, *stemming*, *tokenization*, and *sentence segmentation*. These are all important techniques to train efficient and effective NLP models. Along the way, we will also cover best practices and common mistakes to avoid when training and building NLP models. We also provide some exercises for you to keep practicing and exploring some ideas.

In every chapter, we will introduce the theoretical aspect and motivation of each concept covered. Then we will obtain hands-on experience by using bootstrap methods, industry-standard tools, and other open-source libraries to implement the different techniques. Along the way, we will also cover best practices, share important references, point out common mistakes to avoid when training and building NLP models, and discuss what lies ahead.

Tokenization

Below I am showing you an example of a simple tokenizer without any following any standards. All it does is extract tokens based on a white space separator.

Try to running the following code blocks.

In [1]:

```
doc = "I love coding and writing"
for i, w in enumerate(doc.split(" ")): #enumerate() is used to iterate over a sequence while
    keeping track of both the index and corresponding value
    print("Token " + str(i) + ": " + w)
```

```
Token 0: I
Token 1: love
Token 2: coding
Token 3: and
Token 4: writing
```

All the code does is separate the sentence into individual tokens. The above simple block of code works well on the text I have provided. But typically, text is a lot noisier and complex than the example I used. For instance, if I used the word "so-called" is that one word or two words? For such scenarios, you may need more advanced approaches for tokenization. You can consider stripping away the "-" and splitting into two tokens or just

combining into one token but this all depends on the problem and domain you are working on.

Another problem with our simple algorithm is that it cannot deal with extra whitespaces in the text. In addition, how do we deal with cities like "New York" and "San Francisco"?

Question 1: Copy the code from above and add extra whitespaces to the string value assigned to the `doc` variable and identify the issue with the code. Then try to fix the issue. Hint: Use `text.strip()` to fix the problem.

In [2]:

```
### ENTER CODE HERE
doc = "  I love  coding  and  writing  "
for i, w in enumerate(doc.split(" ")): # enumerate() is used to iterate over a sequence
    while keeping track of both the index and corresponding value
        print("Token " + str(i) + ": " + w)
```

```
Token 0:
Token 1:
Token 2:
Token 3: I
Token 4: love
Token 5:
Token 6:
Token 7: coding
Token 8:
Token 9: and
Token 10:
Token 11: writing
Token 12:
Token 13:
Token 14:
Token 15:
```

spaCy: spaCy is an open-source Python library for natural language processing (NLP). It provides tools to efficiently process and analyze text data, offering features like tokenization, part-of-speech tagging, named entity recognition, and more. Known for its speed and accuracy, spaCy is widely used in NLP tasks for extracting insights and information from text.

In [3]:

```
## required libraries that need to be installed
%%capture
!pip install -U spacy
#!pip install -U spacy-lookups-data
!python -m spacy download en_core_web_sm
```

In [4]:

```
## import the libraries
import spacy #opensource lib used for tokenization, lemmatization and provides pre trained
models
## load the pre trained language model
nlp = spacy.load("en_core_web_sm") # "en" indicates English, "core" suggests that it's a c
ore language model, "web" refers to the source of the training data, and "sm" stands for
"small"

## tokenization
doc = nlp("This is the so-called lemmatization")
for token in doc:
    print(token.text)
```

This
.

```
is
the
so
-
called
lemmatization
```

NLTK (Natural Language Toolkit) is a Python library that provides tools and resources for working with human language data, enabling tasks like text processing, analysis, and understanding. It's widely used for various natural language processing tasks, offering functions for tasks like tokenization, part-of-speech tagging, and more

In [5]:

```
import nltk
nltk.download('punkt') #Punkt is a pre-trained sentence tokenizer provided by NLTK that can intelligently split text into individual sentences
from nltk.tokenize import word_tokenize #word_tokenize function is used to split a piece of text into individual words or tokens
sentence = """At eight o'clock on Thursday morning Arthur didn't feel very good."""
tokens = nltk.word_tokenize(sentence)
tokens
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
```

Out[5]:

```
['At',
 'eight',
 'o'clock',
 'on',
 'Thursday',
 'morning',
 'Arthur',
 'did',
 'n't',
 'feel',
 'very',
 'good',
 '.']
```

Question 2: Copy the code from above and extend the code to print tokens from the given sentence that have a length greater than 5?

In [6]:

```
##CODE HERE
sentence = """At eight o'clock on Thursday morning Arthur didn't feel very good."""
tokens = nltk.word_tokenize(sentence)

# Print tokens with length greater than 5
long_tokens = [token for token in tokens if len(token) > 5]
print("Tokens with length greater than 5:", long_tokens)
```

```
Tokens with length greater than 5: ['o'clock', 'Thursday', 'morning', 'Arthur']
```

Gensim is a Python library designed for topic modeling and document similarity analysis. It specializes in handling large text corpora efficiently. Gensim provides tools for tasks like topic modeling, word embedding, and document similarity calculation. It's widely used for uncovering hidden patterns in text data and generating insights from unstructured text.

In summary, Gensim is a powerful library in the field of natural language processing (NLP) that focuses on tasks related to text analysis and modeling, such as extracting topics from documents and creating word embeddings.

In [7]:

```
#gensim work tokenization
!pip install gensim
from gensim.utils import tokenize
text = """Founded in 2002, SpaceX's mission is to enable humans to become a spacefaring c
ivilization and a multi-planet
species by building a self-sustaining city on Mars. In 2008, SpaceX's Falcon 1 became the
first privately developed
liquid-fuel launch vehicle to orbit the Earth."""
list(tokenize(text))
```

Requirement already satisfied: gensim in /usr/local/lib/python3.10/dist-packages (4.3.3)
Requirement already satisfied: numpy<2.0,>=1.18.5 in /usr/local/lib/python3.10/dist-packa
ges (from gensim) (1.26.4)
Requirement already satisfied: scipy<1.14.0,>=1.7.0 in /usr/local/lib/python3.10/dist-pac
kages (from gensim) (1.13.1)
Requirement already satisfied: smart-open>=1.8.1 in /usr/local/lib/python3.10/dist-packag
es (from gensim) (7.0.4)
Requirement already satisfied: wrapt in /usr/local/lib/python3.10/dist-packages (from sma
rt-open>=1.8.1->gensim) (1.16.0)

Out[7]:

```
['Founded',  
'in',  
'SpaceX',  
's',  
'mission',  
'is',  
'to',  
'enable',  
'humans',  
'to',  
'become',  
'a',  
'spacefaring',  
'civilization',  
'and',  
'a',  
'multi',  
'planet',  
'species',  
'by',  
'building',  
'a',  
'self',  
'sustaining',  
'city',  
'on',  
'Mars',  
'In',  
'SpaceX',  
's',  
'Falcon',  
'became',  
'the',  
'first',  
'privately',  
'developed',  
'liquid',  
'fuel',  
'launch',  
'vehicle',  
'to',  
'orbit',  
'the',  
'Earth']
```

Q 3. Now based on above word tokenizations perform a sentence tokenization using spacy or nltk?

Using NLTK

In [8]:

```
##CODE HERE
text = """Founded in 2002, SpaceX's mission is to enable humans to become a spacefaring c
ivilization and a multi-planet
species by building a self-sustaining city on Mars. In 2008, SpaceX's Falcon 1 became the
first privately developed
liquid-fuel launch vehicle to orbit the Earth."""

# Tokenize the text into sentences
sentences = nltk.sent_tokenize(text)

# Print the tokenized sentences
print("Sentences:", sentences)
```

Sentences: ['Founded in 2002, SpaceX's mission is to enable humans to become a spacefaring c
ivilization and a multi-planet\nspecies by building a self-sustaining city on Mars.',
'In 2008, SpaceX's Falcon 1 became the first privately developed\nliquid-fuel launch vehi
cle to orbit the Earth.']

Using Spacy

In [9]:

```
import spacy

# Load the SpaCy model
nlp = spacy.load('en_core_web_sm')

text = """Founded in 2002, SpaceX's mission is to enable humans to become a spacefaring c
ivilization and a multi-planet
species by building a self-sustaining city on Mars. In 2008, SpaceX's Falcon 1 became the
first privately developed
liquid-fuel launch vehicle to orbit the Earth."""

# Tokenize the text into sentences
doc = nlp(text)
sentences = [sent.text for sent in doc.sents]

# Print the tokenized sentences
print("Sentences:", sentences)
```

Sentences: ['Founded in 2002, SpaceX's mission is to enable humans to become a spacefaring c
ivilization and a multi-planet\nspecies by building a self-sustaining city on Mars.',
'In 2008, SpaceX's Falcon 1 became the first privately developed\nliquid-fuel launch vehi
cle to orbit the Earth.']

BeautifulSoup: BeautifulSoup is a Python library used for web scraping. It helps extract data from HTML and XML documents by providing methods to parse and navigate their content. BeautifulSoup simplifies the process of extracting information from web pages, making it accessible for various data collection and analysis tasks.

In [10]:

```
#code block 1
#import the required libraries

from bs4 import BeautifulSoup
import urllib.request
import nltk
from nltk.corpus import stopwords
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
```

Out[10]:

True

First, we will grab a web page content then we will analyze the text to see what the page is about. We will use the urllib module to crawl the web page:

In [11]:

```
#code block 2
```

```
import urllib.request
```

```
#open and read the spacex data
```

```
response = urllib.request.urlopen('http://php.net/')
```

```
html = response.read()
```

```
print(html)
```

```
b'<!DOCTYPE html>\n<html xmlns="http://www.w3.org/1999/xhtml" lang="en">\n<head>\n\n  <meta charset="utf-8">\n  <meta name="viewport" content="width=device-width, initial-scale=1.0">\n\n  <title>PHP: Hypertext Preprocessor</title>\n\n  <link rel="icon" type="image/svg+xml" sizes="any" href="https://www.php.net/favicon.svg?v=2">\n  <link rel="icon" type="image/png" sizes="196x196" href="https://www.php.net/favicon-196x196.png?v=2">\n  <link rel="icon" type="image/png" sizes="32x32" href="https://www.php.net/favicon-32x32.png?v=2">\n  <link rel="icon" type="image/png" sizes="16x16" href="https://www.php.net/favicon-16x16.png?v=2">\n  <link rel="shortcut icon" href="https://www.php.net/favicon.ico?v=2">\n\n  <link rel="search" type="application/opensearchdescription+xml" href="http://php.net/phpnetimprovedsearch.src" title="Add PHP.net search">\n  <link rel="alternate" type="application/atom+xml" href="https://www.php.net/releases/feed.php" title="PHP Release feed">\n  <link rel="alternate" type="application/atom+xml" href="https://www.php.net/feed.atom" title="PHP: Hypertext Preprocessor">\n\n  <link rel="canonical" href="https://www.php.net/index.php">\n\n  <link rel="shorturl" href="https://www.php.net/index">\n  <link rel="alternate" href="https://www.php.net/index" hreflang="x-default">\n\n  <link rel="stylesheet" type="text/css" href="/cached.php?t=1707321815&f=/fonts/Fira/fira.css" media="screen">\n  <link rel="stylesheet" type="text/css" href="/cached.php?t=1707321815&f=/fonts/Font-Awesome/css/fontello.css" media="screen">\n  <link rel="stylesheet" type="text/css" href="/cached.php?t=1723177202&f=/styles/theme-base.css" media="screen">\n  <link rel="stylesheet" type="text/css" href="/cached.php?t=1720213803&f=/styles/theme-medium.css" media="screen">\n  <link rel="stylesheet" type="text/css" href="/cached.php?t=1707321815&f=/styles/home.css" media="screen">\n\n  <base href="https://www.php.net/index.php">\n\n  <meta name="Description" content="PHP is a popular general-purpose scripting language that powers everything from your blog to the most popular websites in the world." />\n  <meta name="twitter:card" content="summary_large_image" />\n  <meta name="twitter:site" content="@official_php" />\n  <meta name="twitter:title" content="PHP: Hypertext Preprocessor" />\n  <meta name="twitter:description" content="PHP is a popular general-purpose scripting language that powers everything from your blog to the most popular websites in the world." />\n  <meta name="twitter:creator" content="@official_php" />\n  <meta name="twitter:image:src" content="https://www.php.net/images/meta-image.png" />\n  <meta itemprop="name" content="PHP: Hypertext Preprocessor" />\n  <meta itemprop="description" content="PHP is a popular general-purpose scripting language that powers everything from your blog to the most popular websites in the world." />\n  <meta itemprop="image" content="https://www.php.net/images/meta-image.png" />\n  <meta property="og:image" content="https://www.php.net/images/meta-image.png" />\n  <meta property="og:description" content="PHP is a popular general-purpose scripting language that powers everything from your blog to the most popular websites in the world." />\n\n  <link href="https://fosstodon.org/@php" rel="me" />\n</head>\n<body class="home">\n\n  <nav id="head-nav" class="navbar navbar-fixed-top">\n    <div class="navbar-inner clearfix">\n      <a href="/" class="brand"></a>\n      <div id="mainmenu-toggle-overlay"></div>\n      <input type="checkbox" id="mainmenu-toggle">\n      <ul class="nav">\n        <li class=""><a href="/downloads">Downloads</a></li>\n        <li class=""><a href="/docs.php">Documentation</a></li>\n        <li class=""><a href="/get-involved">Get Involved</a></li>\n        <li class=""><a href="/support">Help</a></li>\n        <li class="">\n          <a href="/releases/8.3/index.php">\n            \n          </a>\n        </li>\n      </ul>\n      <form class="navbar-search" id="topsearch" action="/search.php">\n        <input type="hidden" name="show" value="quickref">\n        <input type="search" name="pattern" class="search-query" placeholder="Search" accesskey="s">\n      </form>\n    </div>\n    <div id="flash-message"></div>\n  </nav>\n  <nav id="trick"><div><dl>\n    <dt><a href="/manual/en/getting-started.php">Getting Started</a></dt>\n    <dd><a href="/manual/en/introduction.php">Introduction</a></dd>\n    <dd><a href="/manual/en/tutorial.php">A simple tutorial</a></dd>\n    <dd><a href="/manual/en/langref.php">Language Reference</a></dd>\n    <dd><a href="/manual/en/language.basic-syntax.php">Basic syntax</a></dd>\n    <dd><a href="/manual/en/language.types.php">Types</a></dd>\n    <dd><a href="/manual/en/language.variables.php">Variables</a></dd>\n    <dd><a href="/manual/en/language.constants.php">Constants</a></dd>\n    <dd><a href="/manual/en/language.expressions.php">Expressions
```

[Operators](/manual/en/language.operators.php)

[Control Structures](/manual/en/language.control-structures.php)

[Functions](/manual/en/language.functions.php)

[Classes and Objects](/manual/en/language.oop5.php)

[Namespaces](/manual/en/language.namespaces.php)

[Enumerations](/manual/en/language.enumerations.php)

[Errors](/manual/en/language.errors.php)

[Exceptions](/manual/en/language.exceptions.php)

[Fibers](/manual/en/language.fibers.php)

[Generators](/manual/en/language.generators.php)

[Attributes](/manual/en/language.attributes.php)

[References Explained](/manual/en/language.references.php)

[Predefined Variables](/manual/en/reserved.variables.php)

[Predefined Exceptions](/manual/en/reserved.exceptions.php)

[Predefined Interfaces and Classes](/manual/en/reserved.interfaces.php)

[Predefined Attributes](/manual/en/reserved.attributes.php)

[Context options and parameters](/manual/en/context.php)

[Supported Protocols and Wrappers](/manual/en/wrappers.php)

[Security](/manual/en/security.php)

[Introduction](/manual/en/security.intro.php)

[General considerations](/manual/en/security.general.php)

[Installed as CGI binary](/manual/en/security.cgi-bin.php)

[Installed as an Apache module](/manual/en/security.apache.php)

[Session Security](/manual/en/security.sessions.php)

[Filesystem Security](/manual/en/security.filesystem.php)

[Database Security](/manual/en/security.database.php)

[Error Reporting](/manual/en/security.errors.php)

[User Submitted Data](/manual/en/security.variables.php)

[Hiding PHP](/manual/en/security.hiding.php)

[Keeping Current](/manual/en/security.current.php)

[Features](/manual/en/features.php)

[HTTP authentication with PHP](/manual/en/features.http-auth.php)

[Cookies](/manual/en/features.cookies.php)

[Sessions](/manual/en/features.sessions.php)

[Handling file uploads](/manual/en/features.file-upload.php)

[Using remote files](/manual/en/features.remote-files.php)

[Connection handling](/manual/en/features.connection-handling.php)

[Persistent Database Connections](/manual/en/features.persistent-connections.php)

[Command line usage](/manual/en/features.commandline.php)

[Garbage Collection](/manual/en/features.gc.php)

[DTrace Dynamic Tracing](/manual/en/features.dtrace.php)

[Function Reference](/manual/en/funcref.php)

[Affecting PHP's Behaviour](/manual/en/refs.basics.php)

[Audio Formats Manipulation](/manual/en/refs.utilspec.audio.php)

[Authentication Services](/manual/en/refs.remote.auth.php)

[Command Line Specific Extensions](/manual/en/refs.utilspec.cmdline.php)

[Compression and Archive Extensions](/manual/en/refs.compression.php)

[Cryptography Extensions](/manual/en/refs.crypto.php)

[Database Extensions](/manual/en/refs.database.php)

[Date and Time Related Extensions](/manual/en/refs.calendar.php)

[File System Related Extensions](/manual/en/refs.fileprocess.file.php)

[Human Language and Character Encoding Support](/manual/en/refs.international.php)

[Image Processing and Generation](/manual/en/refs.utilspec.image.php)

[Mail Related Extensions](/manual/en/refs.remote.mail.php)

[Mathematical Extensions](/manual/en/refs.math.php)

[Non-Text MIME Output](/manual/en/refs.utilspec.nontext.php)

[Process Control Extensions](/manual/en/refs.fileprocess.process.php)

[Other Basic Extensions](/manual/en/refs.basic.other.php)

[Other Services](/manual/en/refs.remote.other.php)

[Search Engine Extensions](/manual/en/refs.search.php)

[Server Specific Extensions](/manual/en/refs.utilspec.server.php)

[Session Extensions](/manual/en/refs.basic.session.php)

[Text Processing](/manual/en/refs.basic.text.php)

[Variable and Type Related Extensions](/manual/en/refs.basic.vartype.php)

[Web Services](/manual/en/refs.webservice.php)

[Windows Only Extensions](/manual/en/refs.utilspec.windows.php)

[XML Manipulation](/manual/en/refs.xml.php)

[GUI Extensions](/manual/en/refs.ui.php)

[Keyboard Shortcuts](#)

[This help](#)

[Next menu item](#)

[Previous menu item](#)

[Previous man page](#)

[Next man page](#)

[Scroll to bottom](#)

[Scroll to top](#)

[Goto homepage](#)

[Goto search](#) (current page)

[Focus search box](#)

goto

search

text

results

intro

clearfix

container

hero



A popular

[illegible]

fix release.

All PHP 8.3 users are encouraged to upgrade to this version.

For source downloads of PHP 8.3.10 please visit our [downloads page](https://www.php.net/downloads.php), Windows source and binaries can be found on windows.php.net/download/. The list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.3.10).

18 Jul 2024

PHP 8.4.0 Alpha 2 available for testing

The PHP team is pleased to announce the second testing release of PHP 8.4.0, Alpha 2. This continues the PHP 8.4 release cycle, the rough outline of which is specified in the [PHP Wiki](https://wiki.php.net/todo/php84).

For source downloads of PHP 8.4.0 Alpha 2 please visit the [download page](https://downloads.php.net/~calvinb/).

Please carefully test this version and report any issues found in the [bug reporting system](https://github.com/php/php-src/issues).

Please DO NOT use this version in production, it is an early test version.

For more information on the new features and other changes, you can read the [NEWS](https://github.com/php/php-src/blob/php-8.4.0alpha2/NEWS) file, or the [UPGRADING](https://github.com/php/php-src/blob/php-8.4.0alpha2/UPGRADING) file for a complete list of upgrading notes. These files can also be found in the release archive.

The next release will be Alpha 3, planned for 1 Aug 2024.

The signatures for the release can be found in [the manifest](https://gist.github.com/NattyNarwhal/74b55b8fcc6411ff8db05f87a6f41864) or on [the QA site](https://qa.php.net/).

Thank you for helping us make PHP better.

05 Jul 2024

PHP 8.4.0 Alpha 1 available for testing

The PHP team is pleased to announce the first testing release of PHP 8.4.0, Alpha 1. This starts the PHP 8.4 release cycle, the rough outline of which is specified in the [PHP Wiki](https://wiki.php.net/todo/php84).

For source downloads of PHP 8.4.0 Alpha 1 please visit the [download page](https://downloads.php.net/~saki/).

Please carefully test this version and report any issues found using [the bug tracking system](https://github.com/php/php-src/issues).

Please DO NOT use this version in production, it is an early test version.

For more information on the new features and other changes, you can read the [NEWS](https://github.com/php/php-src/blob/php-8.4.0alpha1/NEWS) file, or the [UPGRADING](https://github.com/php/php-src/blob/php-8.4.0alpha1/UPGRADING) file for a complete list of upgrading notes. These files can also be found in the release archive.

The next release will be Alpha 2, planned for 18 Jul 2024.

The signatures for the release can be found in the [manifest](https://gist.github.com/SakiTakamachi/d3e32b780590319bee72d8402593a3ca) or on the [QA site](https://qa.php.net/).

Thank you for helping us make PHP better.

04 Jul 2024

PHP 8.3.9 Released!

The PHP development team announces the immediate availability of PHP 8.3.9. This is a bug fix release.

All PHP 8.3 users are encouraged to upgrade to this version.

For source downloads of PHP 8.3.9 please visit our [downloads page](https://www.php.net/downloads.php), Windows source and binaries can be found on windows.php.net/download/. The list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.3.9).

04 Jul 2024

PHP 8.2.21 Released!

The PHP development team announces the immediate availability of PHP 8.2.21. This is a bug fix release.

All PHP 8.2 users are encouraged to upgrade to this version.

For source downloads of PHP 8.2.21 please visit our [downloads page](https://www.php.net/download.s.php), Windows source and binaries can be found on windows.php.net/download/. The list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.2.21).

06 Jun 2024

PHP 8.2.20 Released!

The PHP development team announces the immediate availability of PHP 8.2.20. This is a security release.

All PHP 8.2 users are encouraged to upgrade to this version.

For source downloads of PHP 8.2.20 please

visit our [downloads page](https://www.php.net/downloads.php),\nWindows source and binaries can be found on windows.php.net/download/.\nThe list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.2.20).\n

06 Jun 2024

PHP 8.3.8 Released!

The PHP development team announces the immediate availability of PHP 8.3.8. This is a security release.\nAll PHP 8.3 users are encouraged to upgrade to this version.\nFor source downloads of PHP 8.3.8 please visit our [downloads page](https://www.php.net/downloads.php),\nWindows source and binaries can be found on windows.php.net/download/.\nThe list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.3.8).\n

06 Jun 2024

PHP 8.1.29 Released!

The PHP development team announces the immediate availability of PHP 8.1.29. This is a security release.\nAll PHP 8.1 users are encouraged to upgrade to this version.\nFor source downloads of PHP 8.1.29 please visit our [downloads page](https://www.php.net/downloads.php),\nWindows source and binaries can be found on windows.php.net/download/.\nThe list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.1.29).\n

09 May 2024

PHP 8.3.7 Released!

The PHP development team announces the immediate availability of PHP 8.3.7. This is a bug fix release.\nAll PHP 8.3 users are encouraged to upgrade to this version.\nFor source downloads of PHP 8.3.7 please visit our [downloads page](https://www.php.net/downloads.php),\nWindows source and binaries can be found on windows.php.net/download/.\nThe list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.3.7).\n

09 May 2024

PHP 8.2.19 Released!

The PHP development team announces the immediate availability of PHP 8.2.19. This is a bug fix release.\nAll PHP 8.2 users are encouraged to upgrade to this version.\nFor source downloads of PHP 8.2.19 please visit our [downloads page](https://www.php.net/downloads.php),\nWindows source and binaries can be found on windows.php.net/download/.\nThe list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.2.19).\n

24 Apr 2024

Statement on glibc/iconv Vulnerability

EDIT 2024-04-25: Clarified when a PHP application is vulnerable to this bug.\nRecently, a bug in **glibc** version 2.39 and older ([CVE-2024-2961](https://nvd.nist.gov/vuln/detail/CVE-2024-2961)) was uncovered where a buffer overflow in character set conversions **to** the ISO-2022-CN-EXT character set can result in remote code execution.\nThis specific buffer overflow in glibc is exploitable through PHP, which exposes the iconv functionality of glibc to do character set conversions via the [iconv extension](https://www.php.net/manual/en/function.iconv.php). Although the bug is exploitable in the context of the PHP Engine, the bug is not in PHP. It is also not directly exploitable remotely.\n**if and only if**, the PHP application calls iconv [functions](https://www.php.net/manual/en/ref.iconv.php) or [filters](https://www.php.net/manual/en/filters.convert.php#filters.convert.iconv) with user-supplied character sets.\nApplications are **not** vulnerable if:\n

- Glibc security updates from the distribution have been installed.
- Or the iconv extension is not loaded.
- Or the vulnerable character set has been removed from gconv-modules-extra.conf.
- Or the application passes only specifically allowed character sets to iconv.

\nMoreover, when using a user-supplied character set, it is good practice for applications to accept only specific charsets that have been explicitly allowed by the application. One example of how this can be done is by using an allow-list and the [array_search\(\)](https://www.php.net/manual/en/function.array-search) function to check the encoding before passing it to iconv.\nFor example: `array_search($charset, $al`

lowed_list, true)</code>\n </p>\n\n\t <p>There are numerous reports online with title s like "Mitigating the\n\t iconv Vulnerability for PHP (CVE-2024-2961)" or "PHP Under Att ack". These\n\t titles are misleading as this is not a bug in PHP itself .</p>\n\n\t <p>\n\t If your PHP application is vulnerable, we first recommend to check if your Linux distribution\n\t has already published patched variants of glibc.\n\t Debian,\n\t Ce ntOS, and others, have already done so, and please upgrade as soon as possible.\n\t </p>\n\n\t <p>Once an update is available in glibc, updating that package on your\n\t L inux machine will be enough to alleviate the issue. You do not need to\n\t update PHP, as glibc is a dynamically linked library.</p>\n\n\t <p>\n\t If your Linux distributi on has not published a patched version of glibc,\n\t there is no fix for this issue. H owever, there exists a workaround described in\n\t GLIBC\n\t Vulnerability on Servers Serving PHP whi ch explains a way on how to remove\n\t the problematic character set from glibc. Perform this procedure for every\n\t gconv-modules-extra.conf file that is available on your syst em.</p>\n\n\t <p>PHP users on Windows are not affected.</p>\n\n\t <p>Therefore, a new ver sion of PHP will not be released for this vulnerability.</p>\n\t </div>\n\t \n\t </div>\n\t </article><article class="newsentry">\n\t <header class="title">\n\t <time datetime="2024-04-12T15:42:40+00:00">12 Apr 2024</time>\n\t <h2 class="newstitle">\n\t PHP 8.1.28 Released!\n\t </h2>\n\t </header>\n\t <div class="newscontent">\n\t <div><p>The PHP development te am announces the immediate availability of PHP 8.1.28. This is a security release.</p>\n\t <p>All PHP 8.1 users are encouraged to upgrade to this version.</p>\n\t <p>For source dow nloads of PHP 8.1.28 please visit our downloa ds page,\n\t Windows source and binaries can be found on windows.php.net/download/.\n\t The list of changes is recorded in the <a hr ef="https://www.php.net/ChangeLog-8.php#8.1.28">ChangeLog.\n\t </div>\n\t \n\t </div>\n\t </article><article class="newsentry">\n\t <header class="title">\n\t <time datetime= "2024-04-11T14:34:04+00:00">11 Apr 2024</time>\n\t <h2 class="newstitle">\n\t PHP 8.3.6 Released !\n\t </h2>\n\t </header>\n\t <div class="newscontent">\n\t <div><p>The PHP developme nt team announces the immediate availability of PHP 8.3.6. This is a security release.</p>\n\t <p>All PHP 8.3 users are encouraged to upgrade to this version.</p>\n\t <p>For source downloads of PHP 8.3.6 please visit our downl oads page,\n\t Windows source and binaries can be found on windows.php.net/download/.\n\t The list of changes is recorded in the ChangeLog.\n\t </div>\n\t \n\t </div>\n\t </article><article class="newsentry">\n\t <header class="title">\n\t <time datetime= "2024-04-11T13:37:51+00:00">11 Apr 2024</time>\n\t <h2 class="newstitle">\n\t <a hre f="https://www.php.net/archive/2024.php#2024-04-11-1" id="2024-04-11-1">PHP 8.2.18 Releas ed!\n\t </h2>\n\t </header>\n\t <div class="newscontent">\n\t <div><p>The PHP develop ment team announces the immediate availability of PHP 8.2.18. This is a security release.</p>\n\t <p>All PHP 8.2 users are encouraged to upgrade to this version.</p>\n\t <p>For sou rce downloads of PHP 8.2.18 please visit our downloads page,\n\t Windows source and binaries can be found on windows.php.net/download/.\n\t The list of changes is recorded in th e ChangeLog.\n\t </div>\n\t \n\t </div>\n\t </article><article class="newsentry">\n\t <header class="title">\n\t <time da tetime="2024-03-14T21:02:45+00:00">14 Mar 2024</time>\n\t <h2 class="newstitle">\n\t PHP 8.3.4 R eleased!\n\t </h2>\n\t </header>\n\t <div class="newscontent">\n\t <div><p>The PHP de velopment team announces the immediate availability of PHP 8.3.4. This is a bug fix relea se.</p>\n\t <p>All PHP 8.3 users are encouraged to upgrade to this version.</p>\n\t <p>For source downloads of PHP 8.3.4 please visit our downloads page,\n\t Windows source and binaries can be found on windows.php.net/download/.\n\t The list of changes is recorded in the ChangeLog.\n\t </div>\n\t \n\t </div>\n\t </article><article class="newsentry">\n\t <header class="title">\n\t <time da tetime="2024-03-14T15:21:29+00:00">14 Mar 2024</time>\n\t <h2 class="newstitle">\n\t PHP 8.2.17 Released!\n\t </h2>\n\t </header>\n\t <div class="newscontent">\n\t <div>\n\t <p>T he PHP development team announces the immediate availability of PHP 8.2.17. This is a bug fix release.</p>\n\t <p>All PHP 8.2 users are encouraged to upgrade to this version.</p>\n\t <p>For source downloads of PHP 8.2.17 please visit our downloads page,\n\t Windows source and binaries can be found on windows.php.net/download/.\n\t The list of changes is recorded in the C hangeLog.\n\t </div>\n\t \n\t </div>\n\t </article><article class="newsentry">\n\t <header class="title">\n\t <time datetime="2024-02-15T13:36:20-05:00">15 Feb 2024</ time>\n\t <h2 class="newstitle">\n\t PHP 8.2.16 Released!\n\t </h2>\n\t </header>\n\t <div cl ass="newscontent">\n\t <div><p>The PHP development team announces the immediate availabi

lity of PHP 8.2.16. This is a bug fix release.

All PHP 8.2 users are encouraged to upgrade to this version.

For source downloads of PHP 8.2.16 please visit our [downloads page](https://www.php.net/downloads.php), Windows source and binaries can be found on windows.php.net/download/.

The list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.2.16).

PHP 8.3.3 Released!

The PHP development team announces the immediate availability of PHP 8.3.3. This is a bug fix release.

All PHP 8.3 users are encouraged to upgrade to this version.

For source downloads of PHP 8.3.3 please visit our [downloads page](https://www.php.net/downloads.php), Windows source and binaries can be found on windows.php.net/download/.

The list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.3.3).

PHP 8.2.15 Released!

The PHP development team announces the immediate availability of PHP 8.2.15. This is a bug fix release.

All PHP 8.2 users are encouraged to upgrade to this version.

For source downloads of PHP 8.2.15 please visit our [downloads page](https://www.php.net/downloads.php), Windows source and binaries can be found on windows.php.net/download/.

The list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.2.15).

PHP 8.3.2 Released!

The PHP development team announces the immediate availability of PHP 8.3.2. This is a bug fix release.

All PHP 8.3 users are encouraged to upgrade to this version.

For source downloads of PHP 8.3.2 please visit our [downloads page](https://www.php.net/downloads.php), Windows source and binaries can be found on windows.php.net/download/.

The list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.3.2).

PHP 8.2.14 Released!

The PHP development team announces the immediate availability of PHP 8.2.14. This is a bug fix release.

All PHP 8.2 users are encouraged to upgrade to this version.

For source downloads of PHP 8.2.14 please visit our [downloads page](https://www.php.net/downloads.php), Windows source and binaries can be found on windows.php.net/download/.

The list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.2.14).

PHP 8.1.27 Released!

The PHP development team announces the immediate availability of PHP 8.1.27. This is a bug fix release.

All PHP 8.1 users are encouraged to upgrade to this version.

For source downloads of PHP 8.1.27 please visit our [downloads page](https://www.php.net/downloads.php), Windows source and binaries can be found on windows.php.net/download/.

The list of changes is recorded in the [ChangeLog](https://www.php.net/ChangeLog-8.php#8.1.27).

Older News Entries

The PHP Foundation

The PHP Foundation is a collective of people and organizations, united in the mission to ensure the long-term prosperity of the PHP language.

<https://thephp.foundation/donate/>

Upcoming conferences

- <https://www.php.net/conferences/index.php#2024-08-15-3> PHP Russia 2024
- <https://www.php.net/conferences/index.php#2024-07-19-1> CascadiaPHP 2024
- <https://www.php.net/conferences/index.php#2024-06-11-1> International PHP Conference Munich 2024
- <https://www.php.net/conferences/index.php#2024-06-10-1> Forum PHP 2024 - Paris (France)

User Group Events

[Special](/thanks.php)

```

Thanks</a></p>\n      <div class=\'panel social-media\'>\n          <span class=\'headline\'>So
cial media</span>\n          <div class=\'body\'>\n              <ul>\n                  <li>\n
<a href="https://twitter.com/official_php">\n                      <i class="icon-x-twitter"></i>
\n                      @official_php\n                      </a>\n                  </li>\n                  <li>\n
<a href="https://fosstodon.org/@php">\n                      <i class="icon-mastodon"></i>\n
@php@fosstodon.org\n                      </a>\n                  </li>\n              </ul>\n          </div>\n      </div>\n</div>\n</aside>\n\n</div><!-- layout -->\n\n<footer>\n    <div class="contai
ner footer-content">\n        <div class="row-fluid">\n            <ul class="footmenu">\n
<li><a href="/copyright.php">Copyright &copy; 2001-2024 The PHP Group</a></li>\n            <
li><a href="/my.php">My PHP.net</a></li>\n            <li><a href="/contact.php">Contact</a><
/li>\n            <li><a href="/sites.php">Other PHP.net sites</a></li>\n            <li><a href=
"/privacy.php">Privacy policy</a></li>\n            <li><a href="https://github.com/php/web-p
hp/blob/master/index.php">View Source</a></li>\n            </ul>\n        </div>\n    </div>\n</footer>\n\n    <div class=\'elephpants\'><div class=images></div></div>\n <!-- External
and third party libraries. -->\n<script src="https://code.jquery.com/jquery-3.6.0.min.js
" integrity="sha256-/xUj+3OJU5yExlq6GSYGSk7tPXikynS7ogEvDej/m4=" crossorigin="anonymous"
"></script>\n<script src="/cached.php?t=1707321815&f=/js/ext/hogan-3.0.2.min.js"></scr
ipt>\n<script src="/cached.php?t=1723177202&f=/js/ext/typeahead.jquery.min.js"></scri
pt>\n<script src="/cached.php?t=1723177202&f=/js/ext/FuzzySearch.min.js"></script>\n<script
src="/cached.php?t=1707321815&f=/js/ext/mousetrap.min.js"></script>\n<script s
rc="/cached.php?t=1707321815&f=/js/ext/jquery.scrollTo.min.js"></script>\n<script src
="/cached.php?t=1723177202&f=/js/search.js"></script>\n<script src="/cached.php?t=170
7321815&f=/js/common.js"></script>\n\n<a id="toTop" href="javascript:;"><span id="toT
opHover"></span></a>
\n\n</body>\n</html>\n'

```

As you can see from the printed output, the result contains a lot of HTML tags that need to be cleaned. We can use BeautifulSoup to clean the grabbed text like this:

In [12]:

```

#code block 3
#apply beautifulsoap(used for web scraping) and split the text
soup = BeautifulSoup(html,"html5lib") #"html5lib" parser is a robust parser that can handle
poorly formatted HTML and produces a consistent and accurate representation of the document's
structure.
text = soup.get_text(strip=True)
print(text)

```

PHP: Hypertext PreprocessorDownloadsDocumentationGet InvolvedHelpGetting StartedIntroduct ionA simple tutorialLanguage ReferenceBasic syntaxTypesVariablesConstantsExpressionsOpera torsControl StructuresFunctionsClasses and ObjectsNamespacesEnumerationsErrorsExceptionsF ibersGeneratorsAttributesReferences ExplainedPredefined VariablesPredefined ExceptionsPre defined Interfaces and ClassesPredefined AttributesContext options and parametersSupporte d Protocols and WrappersSecurityIntroductionGeneral considerationsInstalled as CGI binary Installed as an Apache moduleSession SecurityFilesystem SecurityDatabase SecurityError Re portingUser Submitted DataHiding PHPKeeping CurrentFeaturesHTTP authentication with PHPCo okiesSessionsHandling file uploadsUsing remote filesConnection handlingPersistent Databas e ConnectionsCommand line usageGarbage CollectionDTrace Dynamic TracingFunction Reference Affecting PHP's BehaviourAudio Formats ManipulationAuthentication ServicesCommand Line Sp ecific ExtensionsCompression and Archive ExtensionsCryptography ExtensionsDatabase Extens ionsDate and Time Related ExtensionsFile System Related ExtensionsHuman Language and Char acter Encoding SupportImage Processing and GenerationMail Related ExtensionsMathematical ExtensionsNon-Text MIME OutputProcess Control ExtensionsOther Basic ExtensionsOther Servi cesSearch Engine ExtensionsServer Specific ExtensionsSession ExtensionsText ProcessingVar iable and Type Related ExtensionsWeb ServicesWindows Only ExtensionsXML ManipulationGUI E xtensionsKeyboard Shortcuts?This helpjNext menu itemkPrevious menu itemg pPrevious man pa geg nNext man pageGScroll to bottomg gScroll to topg hGoto homepageg sGoto search(current page)/Focus search boxApopular general-purpose scripting languagethat is especially suite d to web development.Fast, flexible and pragmatic, PHP powers everything from your blog t o the most popular websites in the world.What's new in 8.3Download8.3.10·Changelog·Upgrad ing8.2.22·Changelog·Upgrading8.1.29·Changelog·Upgrading15 Aug 2024PHP 8.4.0 Beta 3 now av ailable for testingThe PHP team is pleased to announce the release of PHP 8.4.0, Beta 3.

This is the first beta release, continuing the PHP 8.4 release cycle,

the rough outline of which is specified in thePHP Wiki.For source downloads of P HP 8.4.0, Beta 3 please visit thedownload page.Please carefully test this version and rep ort any issues found in thebug reporting system.Please DO NOT use this version in product ion, it is an early test version.For more information on the new features and other chang es, you can read theNEWSfile

ortheUPGRADINGfile for a complete list of upgrading notes. These files can also be

found in the release archive. The next release will be Beta 4, planned for 29 Aug 2024. The signatures for the release can be found in the manifest on the QA site. Thank you for helping us make PHP better.

01 Aug 2024 PHP 8.4.0 Alpha 4 available for testing. The PHP team is pleased to announce the second testing release of PHP 8.4.0, Alpha 4. This continues the PHP 8.4 release cycle, the rough outline of which is specified in the PHP Wiki. For source downloads of PHP 8.4.0 Alpha 4 please visit the download page. Please carefully test this version and report any issues found in the bug reporting system. Please DO NOT use this version in production, it is an early test version. For more information on the new features and other changes, you can read the NEWS file, or the UPGRADING file for a complete list of upgrading notes. These files can also be found in the release archive. The next release will be Beta 1, planned for 15 Aug 2024. The signatures for the release can be found in the manifest on the QA site. Thank you for helping us make PHP better.

01 Aug 2024 PHP 8.2.22 Released! The PHP development team announces the immediate availability of PHP 8.2.22. This is a bug fix release. All PHP 8.2 users are encouraged to upgrade to this version. For source downloads of PHP 8.2.22 please visit our downloads page, Windows source and binaries can be found on windows.php.net/download/. The list of changes is recorded in the ChangeLog.

01 Aug 2024 PHP 8.3.10 Released! The PHP development team announces the immediate availability of PHP 8.3.10. This is a bug fix release. All PHP 8.3 users are encouraged to upgrade to this version. For source downloads of PHP 8.3.10 please visit our downloads page, Windows source and binaries can be found on windows.php.net/download/. The list of changes is recorded in the ChangeLog.

18 Jul 2024 PHP 8.4.0 Alpha 2 available for testing. The PHP team is pleased to announce the second testing release of PHP 8.4.0, Alpha 2. This continues the PHP 8.4 release cycle, the rough outline of which is specified in the PHP Wiki. For source downloads of PHP 8.4.0 Alpha 2 please visit the download page. Please carefully test this version and report any issues found in the bug reporting system. Please DO NOT use this version in production, it is an early test version. For more information on the new features and other changes, you can read the NEWS file, or the UPGRADING file for a complete list of upgrading notes. These files can also be found in the release archive. The next release will be Alpha 3, planned for 1 Aug 2024. The signatures for the release can be found in the manifest on the QA site. Thank you for helping us make PHP better.

05 Jul 2024 PHP 8.4.0 Alpha 1 available for testing. The PHP team is pleased to announce the first testing release of PHP 8.4.0, Alpha 1. This starts the PHP 8.4 release cycle, the rough outline of which is specified in the PHP Wiki. For source downloads of PHP 8.4.0 Alpha 1 please visit the download page. Please carefully test this version and report any issues found using the bug tracking system. Please DO NOT use this version in production, it is an early test version. For more information on the new features and other changes, you can read the NEWS file, or the UPGRADING file for a complete list of upgrading notes. These files can also be found in the release archive. The next release will be Alpha 2, planned for 18 Jul 2024. The signatures for the release can be found in the manifest on the QA site. Thank you for helping us make PHP better.

04 Jul 2024 PHP 8.3.9 Released! The PHP development team announces the immediate availability of PHP 8.3.9. This is a bug fix release. All PHP 8.3 users are encouraged to upgrade to this version. For source downloads of PHP 8.3.9 please visit our downloads page, Windows source and binaries can be found on windows.php.net/download/. The list of changes is recorded in the ChangeLog.

04 Jul 2024 PHP 8.2.21 Released! The PHP development team announces the immediate availability of PHP 8.2.21. This is a bug fix release. All PHP 8.2 users are encouraged to upgrade to this version. For source downloads of PHP 8.2.21 please visit our downloads page, Windows source and binaries can be found on windows.php.net/download/. The list of changes is recorded in the ChangeLog.

06 Jun 2024 PHP 8.2.20 Released! The PHP development team announces the immediate availability of PHP 8.2.20. This is a security release. All PHP 8.2 users are encouraged to upgrade to this version. For source downloads of PHP 8.2.20 please visit our downloads page, Windows source and binaries can be found on windows.php.net/download/. The list of changes is recorded in the ChangeLog.

06 Jun 2024 PHP 8.3.8 Released! The PHP development team announces the immediate availability of PHP 8.3.8. This is a security release. All PHP 8.3 users are encouraged to upgrade to this version. For source downloads of PHP 8.3.8 please visit our downloads page, Windows source and binaries can be found on windows.php.net/download/. The list of changes is recorded in the ChangeLog.

06 Jun 2024 PHP 8.1.29 Released! The PHP development team announces the immediate availability of PHP 8.1.29. This is a security release. All PHP 8.1 users are encouraged to upgrade to this version. For source downloads of PHP 8.1.29 please visit our downloads page, Windows source and binaries can be found on windows.php.net/download/. The list of changes is recorded in the ChangeLog.

09 May 2024 PHP 8.3.7 Released! The PHP development team announces the immediate availability of PHP 8.3.7. This is a bug fix release. All PHP 8.3 users are encouraged to upgrade to this version. For source downloads of PHP 8.3.7 please visit our downloads page, Windows source and binaries can be found on windows.php.net/download/. The list of changes is recorded in the ChangeLog.

09 May 2024 PHP 8.2.19 Released! The PHP development team announces the immediate availability of PHP 8.2.19. This is a bug fix release.

ase.All PHP 8.2 users are encouraged to upgrade to this version.For source downloads of P
HP 8.2.19 please visit ourdownloads page,

Windows source and binaries can be found onwindows.php.net/download/.

The list of changes is recorded in the[ChangeLog](#).24 Apr 2024Statement on glibc/iconv
VulnerabilityEDIT 2024-04-25: Clarified when a PHP application is vulnerable to this bug.
Recently, a bug in glibc version 2.39 and older (CVE-2024-2961) was uncovered

where a buffer overflow in character set conversion to the ISO-2022-CN-EXT character set
can result in remote code execution.This specific buffer overflow in glibc is exploitable
through PHP,

which exposes the iconv functionality of glibc to do character set
conversions via the iconv extension.

Although the bug is exploitable in the context of the PHP

Engine, the bug is not in PHP. It is also not directly exploitable

remotely.The bug is exploitable, if and only if,

the PHP application calls iconv functions or filters with user-supplied character sets.

Applications are not vulnerable if: Glibc security updates from the distribution have been i
nstalled.Or the iconv extension is not loaded.Or the vulnerable character set has been re
moved from gconv-modules-extra.conf.Or the application passes only specifically allowed c
haracter sets to iconv.Moreover, when using a user-supplied character set,

it is good practice for applications to accept only

specific charsets that have been explicitly allowed by the application.

One example of how this can be done is by using an allow-list and the `array_search()`
function

to check the encoding before passing it to iconv.

For example: `array_search($charset, $allowed_list, true)`There are numerous reports o
nline with titles like "Mitigating the

iconv Vulnerability for PHP (CVE-2024-2961)" or "PHP Under Attack". These

titles are misleading as this is not a bug in PHP itself.If your PHP application is vulne
rable, we first recommend to check if your Linux distribution

has already published patched variants of glibc.Debian,

CentOS, and others, have already done so, and please upgrade as soon as possible.On
ce an update is available in glibc, updating that package on your

Linux machine will be enough to alleviate the issue. You do not need to

update PHP, as glibc is a dynamically linked library.If your Linux distribution has
not published a patched version of glibc,

there is no fix for this issue. However, there exists a workaround described in GLIB

C

Vulnerability on Servers Serving PHP which explains a way on how to remove

the problematic character set from glibc. Perform this procedure for every

`gconv-modules-extra.conf` file that is available on your system.PHP users on Windows are
not affected.Therefore, a new version of PHP will not be released for this vulnerability.

12 Apr 2024PHP 8.1.28 Released!The PHP development team announces the immediate availabil
ity of PHP 8.1.28. This is a security release.All PHP 8.1 users are encouraged to upgrade
to this version.For source downloads of PHP 8.1.28 please visit ourdownloads page,

Windows source and binaries can be found onwindows.php.net/download/.

The list of changes is recorded in the[ChangeLog](#).11 Apr 2024PHP 8.3.6 Released!The PHP dev
elopment team announces the immediate availability of PHP 8.3.6. This is a security relea
se.All PHP 8.3 users are encouraged to upgrade to this version.For source downloads of PH
P 8.3.6 please visit ourdownloads page,

Windows source and binaries can be found onwindows.php.net/download/.

The list of changes is recorded in the[ChangeLog](#).11 Apr 2024PHP 8.2.18 Released!The PHP de
velopment team announces the immediate availability of PHP 8.2.18. This is a security rel
ease.All PHP 8.2 users are encouraged to upgrade to this version.For source downloads of
PHP 8.2.18 please visit ourdownloads page,

Windows source and binaries can be found onwindows.php.net/download/.

The list of changes is recorded in the[ChangeLog](#).14 Mar 2024PHP 8.3.4 Released!The PHP dev
elopment team announces the immediate availability of PHP 8.3.4. This is a bug fix releas
e.All PHP 8.3 users are encouraged to upgrade to this version.For source downloads of PHP
8.3.4 please visit ourdownloads page,

Windows source and binaries can be found onwindows.php.net/download/.

The list of changes is recorded in the[ChangeLog](#).14 Mar 2024PHP 8.2.17 Released!The PHP de
velopment team announces the immediate availability of PHP 8.2.17. This is a bug fix rele
ase.All PHP 8.2 users are encouraged to upgrade to this version.For source downloads of P
HP 8.2.17 please visit ourdownloads page,

Windows source and binaries can be found onwindows.php.net/download/.

The list of changes is recorded in the[ChangeLog](#).15 Feb 2024PHP 8.2.16 Released!The
PHP development team announces the immediate availability of PHP 8.2.16. This is a bug fi
x release.All PHP 8.2 users are encouraged to upgrade to this version.For source download
s of PHP 8.2.16 please visit ourdownloads page,

Windows source and binaries can be found onwindows.php.net/download/.

The list of changes is recorded in the[ChangeLog](#).15 Feb 2024PHP 8.3.3 Released!The PHP dev
elopment team announces the immediate availability of PHP 8.3.3. This is a bug fix releas

e.All PHP 8.3 users are encouraged to upgrade to this version.For source downloads of PHP 8.3.3 please visit ourdownloads page,
Windows source and binaries can be found onwindows.php.net/download/.
The list of changes is recorded in theChangeLog.18 Jan 2024PHP 8.2.15 Released!The PHP development team announces the immediate availability of PHP 8.2.15. This is a bug fix release.All PHP 8.2 users are encouraged to upgrade to this version.For source downloads of PHP 8.2.15 please visit ourdownloads page,
Windows source and binaries can be found onwindows.php.net/download/.
The list of changes is recorded in theChangeLog.18 Jan 2024PHP 8.3.2 Released!The PHP development team announces the immediate availability of PHP 8.3.2. This is a bug fix release.All PHP 8.3 users are encouraged to upgrade to this version.For source downloads of PHP 8.3.2 please visit ourdownloads page,
Windows source and binaries can be found onwindows.php.net/download/.
The list of changes is recorded in theChangeLog.21 Dec 2023PHP 8.2.14 Released!The PHP development team announces the immediate availability of PHP 8.2.14. This is a bug fix release.All PHP 8.2 users are encouraged to upgrade to this version.For source downloads of PHP 8.2.14 please visit ourdownloads page,
Windows source and binaries can be found onwindows.php.net/download/.
The list of changes is recorded in theChangeLog.21 Dec 2023PHP 8.1.27 Released!The PHP development team announces the immediate availability of PHP 8.1.27. This is a bug fix release.All PHP 8.1 users are encouraged to upgrade to this version.For source downloads of PHP 8.1.27 please visit ourdownloads page,
Windows source and binaries can be found onwindows.php.net/download/.
The list of changes is recorded in theChangeLog.Older News EntriesThe PHP FoundationThe PHP Foundation is a collective of people and organizations, united in the mission to ensure the long-term prosperity of the PHP language.DonateUpcoming conferencesPHP Russia 2024CascadiaPHP 2024International PHP Conference Munich 2024Forum PHP 2024 - Paris (France)User Group EventsSpecial ThanksSocial media@official_php@php@fosstodon.orgCopyright © 2001-2024 The PHP GroupMy PHP.netContactOther PHP.net sitesPrivacy policyView Source

Now we have a clean text from the crawled web page. Awesome, right? Finally, let's convert that text into tokens by splitting the text like this:

In [13]:

```
#code block 4
#split the text into tokens
tokens = [t for t in text.split()]
print(tokens)
clean_tokens = tokens[:]
print(clean_tokens)
```

```
['PHP:', 'Hypertext', 'PreprocessorDownloadsDocumentationGet', 'InvolvedHelpGetting', 'StartedIntroductionA', 'simple', 'tutorialLanguage', 'ReferenceBasic', 'syntaxTypesVariablesConstantsExpressionsOperatorsControl', 'StructuresFunctionsClasses', 'and', 'ObjectsNamespacesEnumerationsErrorsExceptionsFibersGeneratorsAttributesReferences', 'ExplainedPredefined', 'VariablesPredefined', 'ExceptionsPredefined', 'Interfaces', 'and', 'ClassesPredefined', 'AttributesContext', 'options', 'and', 'parametersSupported', 'Protocols', 'and', 'WrappersSecurityIntroductionGeneral', 'considerationsInstalled', 'as', 'CGI', 'binaryInstalled', 'as', 'an', 'Apache', 'moduleSession', 'SecurityFilesystem', 'SecurityDatabase', 'SecurityError', 'ReportingUser', 'Submitted', 'DataHiding', 'PHPKeeping', 'CurrentFeaturesHTTP', 'authentication', 'with', 'PHPCookiesSessionsHandling', 'file', 'uploadsUsing', 'remote', 'filesConnection', 'handlingPersistent', 'Database', 'ConnectionsCommand', 'line', 'usageGarbage', 'CollectionDTrace', 'Dynamic', 'TracingFunction', 'ReferenceAffecting', 'PHP's', 'BehaviourAudio', 'Formats', 'ManipulationAuthentication', 'ServicesCommand', 'Line', 'Specific', 'ExtensionsCompression', 'and', 'Archive', 'ExtensionsCryptography', 'ExtensionsDatabase', 'ExtensionsDate', 'and', 'Time', 'Related', 'ExtensionsFile', 'System', 'Related', 'ExtensionsHuman', 'Language', 'and', 'Character', 'Encoding', 'SupportImage', 'Processing', 'and', 'GenerationMail', 'Related', 'ExtensionsMathematical', 'ExtensionsNon-Text', 'MIME', 'OutputProcess', 'Control', 'ExtensionsOther', 'Basic', 'Extensionsother', 'ServicesSearch', 'Engine', 'ExtensionsServer', 'Specific', 'ExtensionsSession', 'ExtensionsText', 'ProcessingVariable', 'and', 'Type', 'Related', 'ExtensionsWeb', 'ServicesWindows', 'Only', 'ExtensionsXML', 'ManipulationGUI', 'ExtensionsKeyboard', 'Shortcuts?This', 'helpjNext', 'menu', 'itemkPrevious', 'menu', 'itemg', 'pPrevious', 'man', 'pageg', 'nNext', 'man', 'pageGScroll', 'to', 'bottomg', 'gScroll', 'to', 'topg', 'hGoto', 'homepageg', 'sGoto', 'search(current', 'page)/Focus', 'search', 'boxApopular', 'general-purpose', 'scripting', 'languagethat', 'is', 'especially', 'suited', 'to', 'web', 'development.Fast', 'flexible', 'and', 'pragmatic', 'PHP', 'powers', 'everything', 'from', 'your', 'blog', 'to', 'the', 'most', 'popular', 'websites', 'in', 'the', 'world.What's', 'new', 'in', '8.3Download8.3.10·Changelog·Upgrading8.2.22·Changelog·Upgrading8.1.29·Changelog·Upgrading15', 'Aug', '2024PHP', '8.4.0', 'Beta', '3', 'now', 'available', 'for', 'testingThe
```


','PHP','team','is','pleased','to','announce','the','release','of','PHP','8.4.0','Beta','3.','This','is','the','first','beta','release','continuing','the','PHP','8.4','release','cycle','the','rough','outline','of','which','is','specified','in','thePHP','Wiki.For','source','downloads','of','PHP','8.4.0','Beta','3','please','visit','thedownload','page.Please','carefully','test','this','version','and','report','any','issues','found','in','thebug','reporting','system.Please','DO','NOT','use','this','version','in','production','it','is','an','early','test','version.For','more','information','on','the','new','features','and','other','changes','you','can','read','theNEWSfile','or','theUPGRADINGfile','for','a','complete','list','of','upgrading','notes.','These','files','can','also','be','found','in','the','release','archive.The','next','release','will','be','Beta','4','planned','for','29','August','2024.The','signatures','for','the','release','can','be','found','inthe','manifestor','onthe','QA','site.Thank','you','for','helping','us','make','PHP','better.01','Aug','2024PHP','8.4.0','Alpha','4','available','for','testingThe','PHP','team','is','pleased','to','announce','the','second','testing','release','of','PHP','8.4.0','Alpha','4.','This','continues','the','PHP','8.4','release','cycle','the','rough','outline','of','which','is','specified','in','thePHP','Wiki.For','source','downloads','of','PHP','8.4.0','Alpha','4','please','visit','thedownload','page.Please','carefully','test','this','version','and','report','any','issues','found','in','thebug','reporting','system.Please','DO','NOT','use','this','version','in','production','it','is','an','early','test','version.For','more','information','on','the','new','features','and','other','changes','you','can','read','theNEWSfile','or','theUPGRADINGfile','for','a','complete','list','of','upgrading','notes.','These','files','can','also','be','found','in','the','release','archive.The','next','release','will','be','Beta','1','planned','for','15','Aug','2024.The','signatures','for','the','release','can','be','found','inthe','manifestor','onthe','QA','site.Thank','you','for','helping','us','make','PHP','better.01','Aug','2024PHP','8.2.22','Released!The','PHP','development','team','announces','the','immediate','availability','of','PHP','8.2.22.','This','is','a','bug','fix','release.All','PHP','8.2','users','are','encouraged','to','upgrade','to','this','version.For','source','downloads','of','PHP','8.2.22','please','visit','ourdownloads','page','Windows','source','and','binaries','can','be','found','onwindows.php.net/download/.','The','list','of','changes','is','recorded','in','theChangeLog.01','Aug','2024PHP','8.3.10','Released!The','PHP','development','team','announces','the','immediate','availability','of','PHP','8.3.10.','This','is','a','bug','fix','release.All','PHP','8.3','users','are','encouraged','to','upgrade','to','this','version.For','source','downloads','of','PHP','8.3.10','please','visit','ourdownloads','page','Windows','source','and','binaries','can','be','found','onwindows.php.net/download/.','The','list','of','changes','is','recorded','in','theChangeLog.18','Jul','2024PHP','8.4.0','Alpha','2','available','for','testingThe','PHP','team','is','pleased','to','announce','the','second','testing','release','of','PHP','8.4.0','Alpha','2.','This','continues','the','PHP','8.4','release','cycle','the','rough','outline','of','which','is','specified','in','thePHP','Wiki.For','source','downloads','of','PHP','8.4.0','Alpha','2','please','visit','thedownload','page.Please','carefully','test','this','version','and','report','any','issues','found','in','thebug','reporting','system.Please','DO','NOT','use','this','version','in','production','it','is','an','early','test','version.For','more','information','on','the','new','features','and','other','changes','you','can','read','theNEWSfile','or','theUPGRADINGfile','for','a','complete','list','of','upgrading','notes.','These','files','can','also','be','found','in','the','release','archive.The','next','release','will','be','Alpha','3','planned','for','1','Aug','2024.The','signatures','for','the','release','can','be','found','inthe','manifestor','onthe','QA','site.Thank','you','for','helping','us','make','PHP','better.05','Jul','2024PHP','8.4.0','Alpha','1','available','for','testingThe','PHP','team','is','pleased','to','announce','the','first','testing','release','of','PHP','8.4.0','Alpha','1.','This','starts','the','PHP','8.4','release','cycle','the','rough','outline','of','which','is','specified','in','thePHP','Wiki.For','source','downloads','of','PHP','8.4.0','Alpha','1','please','visit','thedownload','page.Please','carefully','test','this','version','and','report','any','issues','found','usingthe','bug','tracking','system.Please','DO','NOT','use','this','version','in','production','it','is','an','early','test','version.For','more','information','on','the','new','features','and','other','changes','you','can','read','theNEWSfile','or','theUPGRADINGfile','for','a','complete','list','of','upgrading','notes.','These','files','can','also','be','found','in','the','release','archive.The','next','release','will','be','Alpha','2','planned','for','18','Jul','2024.The','signatures','for','the','release','can','be','found','in','themanifestor','on','theQA','site.Thank','you','for','helping','us','make','PHP','better.04','Jul','2024PHP','8.3.9','Released!The','PHP','development','team','announces','the','immediate','availability','of','PHP','8.3.9.','This','is','a','bug','fix','release.All','PHP','8.3','users','are','encouraged','to','upgrade','to','this','version.For','source','downloads','of','PHP','8.3.9','please','visi

t', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.04', 'Jul', '2024PHP', '8.2.21', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.21.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.21', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.06', 'Jun', '2024PHP', '8.2.20', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.20.', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.20', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.06', 'Jun', '2024PHP', '8.3.8', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.8.', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.8', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.06', 'Jun', '2024PHP', '8.1.29', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.1.29.', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.1', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.1.29', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.09', 'May', '2024PHP', '8.3.7', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.7.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.7', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.09', 'May', '2024PHP', '8.2.19', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.19.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.19', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.24', 'Apr', '2024Statement', 'on', 'glibc/iconv', 'VulnerabilityEDIT', '2024-04-25:', 'Clarified', 'when', 'a', 'PHP', 'application', 'is', 'vulnerable', 'to', 'this', 'bug.Recently', 'a', 'bug', 'inlibc version', '2.39', 'and', 'older', '(CVE-2024-2961)', 'was', 'uncovered', 'where', 'a', 'buffer', 'overflow', 'in', 'character', 'set', 'conversionstothe', 'ISO-2022-CN-EXT', 'character', 'set', 'can', 'result', 'in', 'remote', 'code', 'execution.This', 'specific', 'buffer', 'overflow', 'in', 'glibc', 'is', 'exploitable', 'through', 'PHP', 'which', 'exposes', 'the', 'iconv', 'functionality', 'of', 'glibc', 'to', 'do', 'character', 'set', 'conversions', 'via', 'theiconv', 'extension.', 'Although', 'the', 'bug', 'is', 'exploitable', 'in', 'the', 'context', 'of', 'the', 'PHP', 'Engine', 'the', 'bug', 'is', 'not', 'in', 'PHP.', 'It', 'is', 'also', 'not', 'directly', 'exploitable', 'remotely.The', 'bug', 'is', 'exploitable,if', 'and', 'only', 'if', 'the', 'PHP', 'application', 'calls', 'iconvfunctionsorfilterswith', 'user-supplied', 'character', 'sets.Applications', 'arenotvulnerable', 'if:Glibc', 'security', 'updates', 'from', 'the', 'distribution', 'have', 'been', 'installed.Or', 'the', 'iconv', 'extension', 'is', 'not', 'loaded.Or', 'the', 'vulnerable', 'character', 'set', 'has', 'been', 'removed', 'from', 'gconv-modules-extra.conf.Or', 'the', 'application', 'passes', 'only', 'specifically', 'allowed', 'character', 'sets', 'to', 'iconv.Moreover', 'when', 'using', 'a', 'user-supplied', 'character', 'set', 'it', 'is', 'good', 'practice', 'for', 'applications', 'to', 'accept', 'only', 'specific', 'character sets', 'that', 'have', 'been', 'explicitly', 'allowed', 'by', 'the', 'application.', 'One', 'example', 'of', 'how', 'this', 'can', 'be', 'done', 'is', 'by', 'using', 'an', 'allow-list', 'and', 'thearray_search()function', 'to', 'check', 'the', 'encoding', 'before', 'passing', 'it', 'to', 'iconv.', 'For', 'example:array_search(\$charset, \$allowed_list, true)There', 'are', 'numerous', 'reports', 'online', 'with', 'titles', 'like', '"Mitigating', 'the', 'iconv', 'Vulnerability', 'for', 'PHP', '(CVE-2024-2961)"', 'or', '"PHP', 'Under', 'Attack".', 'These', 'titles', 'are', 'misleading', 'as', 'this', 'isnota', 'bug', 'in', 'PHP', 'itself.If', 'your', 'PHP', 'application', 'is', 'vulnerable', 'we', 'first', 'recommend', 'to', 'check', 'if', 'your', 'Linux', 'distribution', 'has', 'already', 'published', 'patched', 'variants', 'of', 'glibc.Debian', 'CentOS', 'and', 'others', 'have', 'already', 'done', 'so', 'and', 'please', 'upgrade', 'as', 'soon', 'as', 'possible.Once', 'an', 'update', 'is', 'available', 'in', 'glibc', 'updating', 'that', 'package', 'on', 'your', 'Linux', 'machine', 'will', 'be', 'enough', 'to', 'alleviate', 'the', 'issue.', 'You', 'do', 'not', 'need', 'to', 'update', 'PHP', 'as', 'glibc', 'is', 'a', 'dynamically', 'linked', 'library.If', 'your', 'Linux', 'distribution', 'has', 'not', 'publi

shed', 'a', 'patched', 'version', 'of', 'glibc', 'there', 'is', 'no', 'fix', 'for', 'this', 'issue.', 'However', 'there', 'exists', 'a', 'workaround', 'described', 'inGLIBC', 'Vulnerability', 'on', 'Servers', 'Serving', 'PHPwhich', 'explains', 'a', 'way', 'on', 'how', 'to', 'remove', 'the', 'problematic', 'character', 'set', 'from', 'glibc.', 'Perform', 'this', 'procedure', 'for', 'every', 'gconv-modules-extra.conf', 'file', 'that', 'is', 'available', 'on', 'your', 'system.PHP', 'users', 'on', 'Windows', 'are', 'not', 'affected. Therefore', 'a', 'new', 'version', 'of', 'PHP', 'will', 'not', 'be', 'released', 'for', 'this', 'vulnerability.12', 'Apr', '2024PHP', '8.1.28', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.1.28.', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.1', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.1.28', 'please', 'visit', 'ourdownloads', 'page,', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.11', 'Apr', '2024PHP', '8.3.6', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.6.', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.6', 'please', 'visit', 'ourdownloads', 'page,', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.11', 'Apr', '2024PHP', '8.2.18', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.18.', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.18', 'please', 'visit', 'ourdownloads', 'page,', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.14', 'Mar', '2024PHP', '8.3.4', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.4.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.4', 'please', 'visit', 'ourdownloads', 'page,', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.14', 'Mar', '2024PHP', '8.2.17', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.17.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.17', 'please', 'visit', 'ourdownloads', 'page,', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.15', 'Feb', '2024PHP', '8.2.16', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.16.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.16', 'please', 'visit', 'ourdownloads', 'page,', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.15', 'Feb', '2024PHP', '8.3.3', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.3.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.3', 'please', 'visit', 'ourdownloads', 'page,', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.18', 'Jan', '2024PHP', '8.2.15', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.15.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.15', 'please', 'visit', 'ourdownloads', 'page,', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.18', 'Jan', '2024PHP', '8.3.2', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.2.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.2', 'please', 'visit', 'ourdownloads', 'page,', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.21', 'Dec', '2023PHP', '8.2.14', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.14.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.14', 'please', 'visit', 'ourdownloads', 'page,', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.21', 'Dec', '2023PHP', '8.1.27', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.1.27.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.1', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.1.27', 'please

ase', 'visit', 'ourdownloads', 'page,', 'Windows', 'source', 'and', 'binaries', 'can', 'b
e', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recor
ded', 'in', 'theChangeLog.Older', 'News', 'EntriesThe', 'PHP', 'FoundationThe', 'PHP', 'F
oundation', 'is', 'a', 'collective', 'of', 'people', 'and', 'organizations,', 'united', 'i
n', 'the', 'mission', 'to', 'ensure', 'the', 'long-term', 'prosperity', 'of', 'the', 'PH
P', 'language.DonateUpcoming', 'conferencesPHP', 'Russia', '2024CascadiaPHP', '2024Intern
ational', 'PHP', 'Conference', 'Munich', '2024Forum', 'PHP', '2024', '-', 'Paris', '(Fran
ce)User', 'Group', 'EventsSpecial', 'ThanksSocial', 'media@official_php@php@fosstodon.org
Copyright', '©', '2001-2024', 'The', 'PHP', 'GroupMy', 'PHP.netContactOther', 'PHP.net',
'sitesPrivacy', 'policyView', 'Source']

['PHP:', 'Hypertext', 'PreprocessorDownloadsDocumentationGet', 'InvolvedHelpGetting', 'St
artedIntroductionA', 'simple', 'tutorialLanguage', 'ReferenceBasic', 'syntaxTypesVariable
sConstantsExpressionsOperatorsControl', 'StructuresFunctionsClasses', 'and', 'ObjectsName
spacesEnumerationsErrorsExceptionsFibersGeneratorsAttributesReferences', 'ExplainedPredef
ined', 'VariablesPredefined', 'ExceptionsPredefined', 'Interfaces', 'and', 'ClassesPredef
ined', 'AttributesContext', 'options', 'and', 'parametersSupported', 'Protocols', 'and',
'WrappersSecurityIntroductionGeneral', 'considerationsInstalled', 'as', 'CGI', 'binaryIns
talled', 'as', 'an', 'Apache', 'moduleSession', 'SecurityFilesystem', 'SecurityDatabase',
'SecurityError', 'ReportingUser', 'Submitted', 'DataHiding', 'PHPKeeping', 'CurrentFeatur
esHTTP', 'authentication', 'with', 'PHPCookiesSessionsHandling', 'file', 'uploadsUsing',
'remote', 'filesConnection', 'handlingPersistent', 'Database', 'ConnectionsCommand', 'lin
e', 'usageGarbage', 'CollectionDTrace', 'Dynamic', 'TracingFunction', 'ReferenceAffecting
, 'PHP's', 'BehaviourAudio', 'Formats', 'ManipulationAuthentication', 'ServicesCommand',
'Line', 'Specific', 'ExtensionsCompression', 'and', 'Archive', 'ExtensionsCryptography',
'ExtensionsDatabase', 'ExtensionsDate', 'and', 'Time', 'Related', 'ExtensionsFile', 'Syst
em', 'Related', 'ExtensionsHuman', 'Language', 'and', 'Character', 'Encoding', 'SupportIm
age', 'Processing', 'and', 'GenerationMail', 'Related', 'ExtensionsMathematical', 'Extens
ionsNon-Text', 'MIME', 'OutputProcess', 'Control', 'ExtensionsOther', 'Basic', 'Extension
sOther', 'ServicesSearch', 'Engine', 'ExtensionsServer', 'Specific', 'ExtensionsSession',
'ExtensionsText', 'ProcessingVariable', 'and', 'Type', 'Related', 'ExtensionsWeb', 'Servi
cesWindows', 'Only', 'ExtensionsXML', 'ManipulationGUI', 'ExtensionsKeyboard', 'Shortcuts
?This', 'helpjNext', 'menu', 'itemkPrevious', 'menu', 'itemg', 'pPrevious', 'man', 'pageg
, 'nNext', 'man', 'pageGScroll', 'to', 'bottomg', 'gScroll', 'to', 'topg', 'hGoto', 'hom
epageg', 'sGoto', 'search(current', 'page)/Focus', 'search', 'boxApopular', 'general-purp
ose', 'scripting', 'languagethat', 'is', 'especially', 'suited', 'to', 'web', 'developmen
t.Fast,', 'flexible', 'and', 'pragmatic,', 'PHP', 'powers', 'everything', 'from', 'your',
'blog', 'to', 'the', 'most', 'popular', 'websites', 'in', 'the', 'world.What's', 'new', 'i
n', '8.3Download8.3.10·Changelog·Upgrading8.2.22·Changelog·Upgrading8.1.29·Changelog·Upg
rading15', 'Aug', '2024PHP', '8.4.0', 'Beta', '3', 'now', 'available', 'for', 'testingThe
, 'PHP', 'team', 'is', 'pleased', 'to', 'announce', 'the', 'release', 'of', 'PHP', '8.4.
0', 'Beta', '3.', 'This', 'is', 'the', 'first', 'beta', 'release,', 'continuing', 'the',
'PHP', '8.4', 'release', 'cycle,', 'the', 'rough', 'outline', 'of', 'which', 'is', 'speci
fied', 'in', 'thePHP', 'Wiki.For', 'source', 'downloads', 'of', 'PHP', '8.4.0', 'Beta',
'3', 'please', 'visit', 'thedownload', 'page.Please', 'carefully', 'test', 'this', 'versi
on', 'and', 'report', 'any', 'issues', 'found', 'in', 'thebug', 'reporting', 'system.Plea
se', 'DO', 'NOT', 'use', 'this', 'version', 'in', 'production,', 'it', 'is', 'an', 'early
, 'test', 'version.For', 'more', 'information', 'on', 'the', 'new', 'features', 'and', 'o
ther', 'changes,', 'you', 'can', 'read', 'theNEWSfile', 'or', 'theUPGRADINGfile', 'for',
'a', 'complete', 'list', 'of', 'upgrading', 'notes.', 'These', 'files', 'can', 'also', 'b
e', 'found', 'in', 'the', 'release', 'archive.The', 'next', 'release', 'will', 'be', 'Bet
a', '4,', 'planned', 'for', '29', 'August', '2024.The', 'signatures', 'for', 'the', 'rele
ase', 'can', 'be', 'found', 'inthe', 'manifestor', 'onthe', 'QA', 'site.Thank', 'you', 'f
or', 'helping', 'us', 'make', 'PHP', 'better.01', 'Aug', '2024PHP', '8.4.0', 'Alpha', '4',
, 'available', 'for', 'testingThe', 'PHP', 'team', 'is', 'pleased', 'to', 'announce', 'th
e', 'second', 'testing', 'release', 'of', 'PHP', '8.4.0', 'Alpha', '4.', 'This', 'contin
ues', 'the', 'PHP', '8.4', 'release', 'cycle,', 'the', 'rough', 'outline', 'of', 'which',
'is', 'specified', 'in', 'thePHP', 'Wiki.For', 'source', 'downloads', 'of', 'PHP', '8.4.0',
'Alpha', '4', 'please', 'visit', 'thedownload', 'page.Please', 'carefully', 'test', 't
his', 'version', 'and', 'report', 'any', 'issues', 'found', 'in', 'thebug', 'reporting',
'system.Please', 'DO', 'NOT', 'use', 'this', 'version', 'in', 'production,', 'it', 'is',
'an', 'early', 'test', 'version.For', 'more', 'information', 'on', 'the', 'new', 'feature
s', 'and', 'other', 'changes,', 'you', 'can', 'read', 'theNEWSfile,', 'or', 'theUPGRADING
file', 'for', 'a', 'complete', 'list', 'of', 'upgrading', 'notes.', 'These', 'files', 'ca
n', 'also', 'be', 'found', 'in', 'the', 'release', 'archive.The', 'next', 'release', 'wil
l', 'be', 'Beta', '1,', 'planned', 'for', '15', 'Aug', '2024.The', 'signatures', 'for', 't
he', 'release', 'can', 'be', 'found', 'inthe', 'manifestor', 'onthe', 'QA', 'site.Thank',
, 'you', 'for', 'helping', 'us', 'make', 'PHP', 'better.01', 'Aug', '2024PHP', '8.2.22',
'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availabil
ity', 'of', 'PHP', '8.2.22.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2
, 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source',
'downloads', 'of', 'PHP', '8.2.22', 'please', 'visit', 'ourdownloads', 'page,', 'Windows',
, 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The

, 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.01', 'Aug', '2024PHP', '8.3.10', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.10.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.10', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/./', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.18', 'Jul', '2024PHP', '8.4.0', 'Alpha', '2', 'available', 'for', 'testingThe', 'PHP', 'team', 'is', 'pleased', 'to', 'announce', 'the', 'second', 'testing', 'release', 'of', 'PHP', '8.4.0', 'Alpha', '2.', 'This', 'continues', 'the', 'PHP', '8.4', 'release', 'cycle', 'the', 'rough', 'outline', 'of', 'which', 'is', 'specified', 'in', 'thePHP', 'Wiki.For', 'source', 'downloads', 'of', 'PHP', '8.4.0', 'Alpha', '2', 'please', 'visit', 'thedownload', 'page.P', 'lease', 'carefully', 'test', 'this', 'version', 'and', 'report', 'any', 'issues', 'found', 'in', 'thebug', 'reporting', 'system.Please', 'DO', 'NOT', 'use', 'this', 'version', 'in', 'production', 'it', 'is', 'an', 'early', 'test', 'version.For', 'more', 'information', 'on', 'the', 'new', 'features', 'and', 'other', 'changes', 'you', 'can', 'read', 'the', 'NEWSfile', 'or', 'theUPGRADINGfile', 'for', 'a', 'complete', 'list', 'of', 'upgrading', 'notes.', 'These', 'files', 'can', 'also', 'be', 'found', 'in', 'the', 'release', 'archive.The', 'next', 'release', 'will', 'be', 'Alpha', '3', 'planned', 'for', '1', 'Aug', '2024.The', 'signatures', 'for', 'the', 'release', 'can', 'be', 'found', 'inthe', 'manifestor', 'onthe', 'QA', 'site.Thank', 'you', 'for', 'helping', 'us', 'make', 'PHP', 'better.05', 'Jul', '2024PHP', '8.4.0', 'Alpha', '1', 'available', 'for', 'testingThe', 'PHP', 'team', 'is', 'pleased', 'to', 'announce', 'the', 'first', 'testing', 'release', 'of', 'PHP', '8.4.0', 'Alpha', '1.', 'This', 'starts', 'the', 'PHP', '8.4', 'release', 'cycle', 'the', 'rough', 'outline', 'of', 'which', 'is', 'specified', 'in', 'thePHP', 'Wiki.For', 'source', 'downloads', 'of', 'PHP', '8.4.0', 'Alpha', '1', 'please', 'visit', 'thedownload', 'page.Please', 'carefully', 'test', 'this', 'version', 'and', 'report', 'any', 'issues', 'found', 'usingthe', 'bug', 'tracking', 'system.Please', 'DO', 'NOT', 'use', 'this', 'version', 'in', 'production', 'it', 'is', 'an', 'early', 'test', 'version.For', 'more', 'information', 'on', 'the', 'new', 'features', 'and', 'other', 'changes', 'you', 'can', 'read', 'theNEWSfile', 'or', 'theUPGRADINGfile', 'for', 'a', 'complete', 'list', 'of', 'upgrading', 'notes.', 'These', 'files', 'can', 'also', 'be', 'found', 'in', 'the', 'release', 'archive.The', 'next', 'release', 'will', 'be', 'Alpha', '2', 'planned', 'for', '18', 'Jul', '2024.The', 'signatures', 'for', 'the', 'release', 'can', 'be', 'found', 'in', 'themanifestor', 'on', 'theQA', 'site.Thank', 'you', 'for', 'helping', 'us', 'make', 'PHP', 'better.04', 'Jul', '2024PHP', '8.3.9', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.9.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.9', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/./', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.04', 'Jul', '2024PHP', '8.2.21', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.21.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.21', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/./', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.06', 'Jun', '2024PHP', '8.2.20', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.20.', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.20', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/./', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.06', 'Jun', '2024PHP', '8.3.8', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.8.', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.8', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/./', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.06', 'Jun', '2024PHP', '8.1.29', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.1.29.', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.1', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.1.29', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/./', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.09', 'May', '2024PHP', '8.3.7', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.7.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.7', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/./', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.09', 'May', '2024PHP', '8.2.19', 'Rel

Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.19.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.19', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.24', 'Apr', '2024Statement', 'on', 'glibc/iconv', 'VulnerabilityEDIT', '2024-04-25:', 'Clarified', 'when', 'a', 'PHP', 'application', 'is', 'vulnerable', 'to', 'this', 'bug.Recently', 'a', 'bug', 'inlibc version', '2.39', 'and', 'older', '(CVE-2024-2961)', 'was', 'uncovered', 'where', 'a', 'buffer', 'overflow', 'in', 'character', 'set', 'conversionstothe', 'ISO-2022-CN-EXT', 'character', 'set', 'can', 'result', 'in', 'remote', 'code', 'execution.This', 'specific', 'buffer', 'overflow', 'in', 'glibc', 'is', 'exploitable', 'through', 'PHP', 'which', 'exposes', 'the', 'iconv', 'functionality', 'of', 'glibc', 'to', 'do', 'character', 'set', 'conversions', 'via', 'theiconv', 'extension.', 'Although', 'the', 'bug', 'is', 'exploitable', 'in', 'the', 'context', 'of', 'the', 'PHP', 'Engine', 'the', 'bug', 'is', 'not', 'in', 'PHP.', 'It', 'is', 'also', 'not', 'directly', 'exploitable', 'remotely.The', 'bug', 'is', 'exploitable,if', 'and', 'only', 'if', 'the', 'PHP', 'application', 'calls', 'iconv functionsorfilterswith', 'user-supplied', 'character', 'sets.Applications', 'arenotvulnerable', 'if:Glibc', 'security', 'updates', 'from', 'the', 'distribution', 'have', 'been', 'installed.Or', 'the', 'iconv', 'extension', 'is', 'not', 'loaded.Or', 'the', 'vulnerable', 'character', 'set', 'has', 'been', 'removed', 'from', 'gconv-modules-extra.conf.Or', 'the', 'application', 'passes', 'only', 'specifically', 'allowed', 'character', 'sets', 'to', 'iconv.Moreover', 'when', 'using', 'a', 'user-supplied', 'character', 'set', 'it', 'is', 'good', 'practice', 'for', 'applications', 'to', 'accept', 'only', 'specific', 'character sets', 'that', 'have', 'been', 'explicitly', 'allowed', 'by', 'the', 'application.', 'One', 'example', 'of', 'how', 'this', 'can', 'be', 'done', 'is', 'by', 'using', 'an', 'allow-list', 'and', 'thearray_search()function', 'to', 'check', 'the', 'encoding', 'before', 'passing', 'it', 'to', 'iconv.', 'For', 'example:array_search(\$charset', '\$allowed_list', 'true)There', 'are', 'numerous', 'reports', 'online', 'with', 'titles', 'like', '"Mitigating', 'the', 'iconv', 'Vulnerability', 'for', 'PHP', '(CVE-2024-2961)"', 'or', '"PHP', 'Under', 'Attack".', 'These', 'titles', 'are', 'misleading', 'as', 'this', 'isnota', 'bug', 'in', 'PHP', 'itself.If', 'your', 'PHP', 'application', 'is', 'vulnerable', 'we', 'first', 'recommend', 'to', 'check', 'if', 'your', 'Linux', 'distribution', 'has', 'already', 'published', 'patched', 'variants', 'of', 'glibc.Debian', 'CentOS', 'and', 'others', 'have', 'already', 'done', 'so', 'and', 'please', 'upgrade', 'as', 'soon', 'as', 'possible.Once', 'an', 'update', 'is', 'available', 'in', 'glibc', 'updating', 'that', 'package', 'on', 'your', 'Linux', 'machine', 'will', 'be', 'enough', 'to', 'alleviate', 'the', 'issue.', 'You', 'do', 'not', 'need', 'to', 'update', 'PHP', 'as', 'glibc', 'is', 'a', 'dynamically', 'linked', 'library.If', 'your', 'Linux', 'distribution', 'has', 'not', 'published', 'a', 'patched', 'version', 'of', 'glibc', 'there', 'is', 'no', 'fix', 'for', 'this', 'issue.', 'However', 'there', 'exists', 'a', 'workaround', 'described', 'inGLIBC', 'Vulnerability', 'on', 'Servers', 'Serving', 'PHPwhich', 'explains', 'a', 'way', 'on', 'how', 'to', 'remove', 'the', 'problematic', 'character', 'set', 'from', 'glibc.', 'Perform', 'this', 'procedure', 'for', 'every', 'gconv-modules-extra.conf', 'file', 'that', 'is', 'available', 'on', 'your', 'system.PHP', 'users', 'on', 'Windows', 'are', 'not', 'affected.Therefore', 'a', 'new', 'version', 'of', 'PHP', 'will', 'not', 'be', 'released', 'for', 'this', 'vulnerability.12', 'Apr', '2024PHP', '8.1.28', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.1.28.', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.1', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.1.28', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.11', 'Apr', '2024PHP', '8.3.6', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.6', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.6', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.11', 'Apr', '2024PHP', '8.2.18', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.18.', 'This', 'is', 'a', 'security', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.18', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.14', 'Mar', '2024PHP', '8.3.4', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.4.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.4', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.14', 'Mar', '2024PHP', '8.2.17', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability',

, 'of', 'PHP', '8.2.17.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.17', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.15', 'Feb', '2024PHP', '8.2.16', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.16.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.16', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.15', 'Feb', '2024PHP', '8.3.3', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.3.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.3', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.18', 'Jan', '2024PHP', '8.2.15', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.15.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.15', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.18', 'Jan', '2024PHP', '8.3.2', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.3.2.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.3', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.3.2', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.21', 'Dec', '2023PHP', '8.2.14', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.2.14.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.2', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.2.14', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.21', 'Dec', '2023PHP', '8.1.27', 'Released!The', 'PHP', 'development', 'team', 'announces', 'the', 'immediate', 'availability', 'of', 'PHP', '8.1.27.', 'This', 'is', 'a', 'bug', 'fix', 'release.All', 'PHP', '8.1', 'users', 'are', 'encouraged', 'to', 'upgrade', 'to', 'this', 'version.For', 'source', 'downloads', 'of', 'PHP', '8.1.27', 'please', 'visit', 'ourdownloads', 'page', 'Windows', 'source', 'and', 'binaries', 'can', 'be', 'found', 'onwindows.php.net/download/.', 'The', 'list', 'of', 'changes', 'is', 'recorded', 'in', 'theChangeLog.Older', 'News', 'EntriesThe', 'PHP', 'FoundationThe', 'PHP', 'Foundation', 'is', 'a', 'collective', 'of', 'people', 'and', 'organizations', 'united', 'in', 'the', 'mission', 'to', 'ensure', 'the', 'long-term', 'prosperity', 'of', 'the', 'PHP', 'language.DonateUpcoming', 'conferencesPHP', 'Russia', '2024CascadiaPHP', '2024International', 'PHP', 'Conference', 'Munich', '2024Forum', 'PHP', '2024', '-', 'Paris', '(France)User', 'Group', 'EventsSpecial', 'ThanksSocial', 'media@official_php@php@fosstodon.org', 'Copyright', '©', '2001-2024', 'The', 'PHP', 'GroupMy', 'PHP.netContactOther', 'PHP.net', 'sitesPrivacy', 'policyView', 'Source']

Stopwords are commonly used words that are filtered out from text during text preprocessing in natural language processing (NLP) tasks. These words are considered to be of little value in many NLP analyses because they are very frequent and don't carry significant meaning by themselves. Filtering out stopwords can help improve the efficiency of text analysis and reduce noise in the data.

Examples of stopwords in English include words like "the," "and," "is," "of," "in," "to," "for," "that," and so on.

Count word frequency The text is much better now. Let's calculate the frequency distribution of those tokens using Python NLTK. There is a function in NLTK called `FreqDist()` does the job:

In [14]:

```
#code block 5

#remove the simpe repetitive english words using stopwords
sr = stopwords.words('english')
for token in tokens:
    if token in stopwords.words('english'):
        clean_tokens.remove(token)
freq = nltk.FreqDist(clean_tokens)
```

```
#del_tokens = freq.copy()
```

If you search the output, you'll find that the most frequent token is PHP. You can plot a graph for those tokens using plot function like this:

In [15]:

```
#code block 6
#print and plot the 10 tokens with high frequency
for key, val in freq.items():
    print (str(key) + ':' + str(val))
freq.plot(10, cumulative=False)
```

```
PHP::1
Hypertext:1
PreprocessorDownloadsDocumentationGet:1
InvolvedHelpGetting:1
StartedIntroductionA:1
simple:1
tutorialLanguage:1
ReferenceBasic:1
syntaxTypesVariablesConstantsExpressionsOperatorsControl:1
StructuresFunctionsClasses:1
ObjectsNamespacesEnumerationsErrorsExceptionsFibersGeneratorsAttributesReferences:1
ExplainedPredefined:1
VariablesPredefined:1
ExceptionsPredefined:1
Interfaces:1
ClassesPredefined:1
AttributesContext:1
options:1
parametersSupported:1
Protocols:1
WrappersSecurityIntroductionGeneral:1
considerationsInstalled:1
CGI:1
binaryInstalled:1
Apache:1
moduleSession:1
SecurityFilesystem:1
SecurityDatabase:1
SecurityError:1
ReportingUser:1
Submitted:1
DataHiding:1
PHPKeeping:1
CurrentFeaturesHTTP:1
authentication:1
PHPCookiesSessionsHandling:1
file:2
uploadsUsing:1
remote:2
filesConnection:1
handlingPersistent:1
Database:1
ConnectionsCommand: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
```


ExtensionsCryptography: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
bottomg:1
gScroll:1
topg:1
hGoto:1
homepageg:1
sGoto:1
search(current:1
page)/Focus:1
search:1
boxApopular:1
general-purpose:1
scripting:1
languagethat:1
especially:1
suited:1
web:1
development.Fast,:1
flexible:1
pragmatic,:1
PHP:114
powers:1
everything:1
blog:1
popular:1
websites:1
world.What's:1
new:6
8.3Download8.3.10·Changelog·Upgrading8.2.22·Changelog·Upgrading8.1.29·Changelog·Upgrading

15:1
Aug:6
2024PHP:22
8.4.0:7
Beta:5
3:2
available:6
testingThe:4
team:24
pleased:4
announce:4
release:20
8.4.0,:5
3.:1
This:24
first:3
beta:1
release,:1
continuing:1
8.4:4
cycle,:4
rough:4
outline:4
specified:4
thePHP:4
Wiki.For:4
source:44
downloads:24
please:25
visit:24
thedownload:4
page.Please:4
carefully:4
test:8
version:10
report:4
issues:4
found:32
thebug:3
reporting:3
system.Please:4
DO:4
NOT:4
use:4
production,:4
early:4
version.For:24
information:4
features:4
changes,:4
read:4
theNEWSfile:1
theUPGRADINGfile:4
complete:4
list:24
upgrading:4
notes.:4
These:5
files:4
also:5
archive.The:4
next:4
4,:1
planned:4
29:1
August:1
2024.The:4
signatures:4
inthe:3
manifestor:3
onthe:3
QA:3

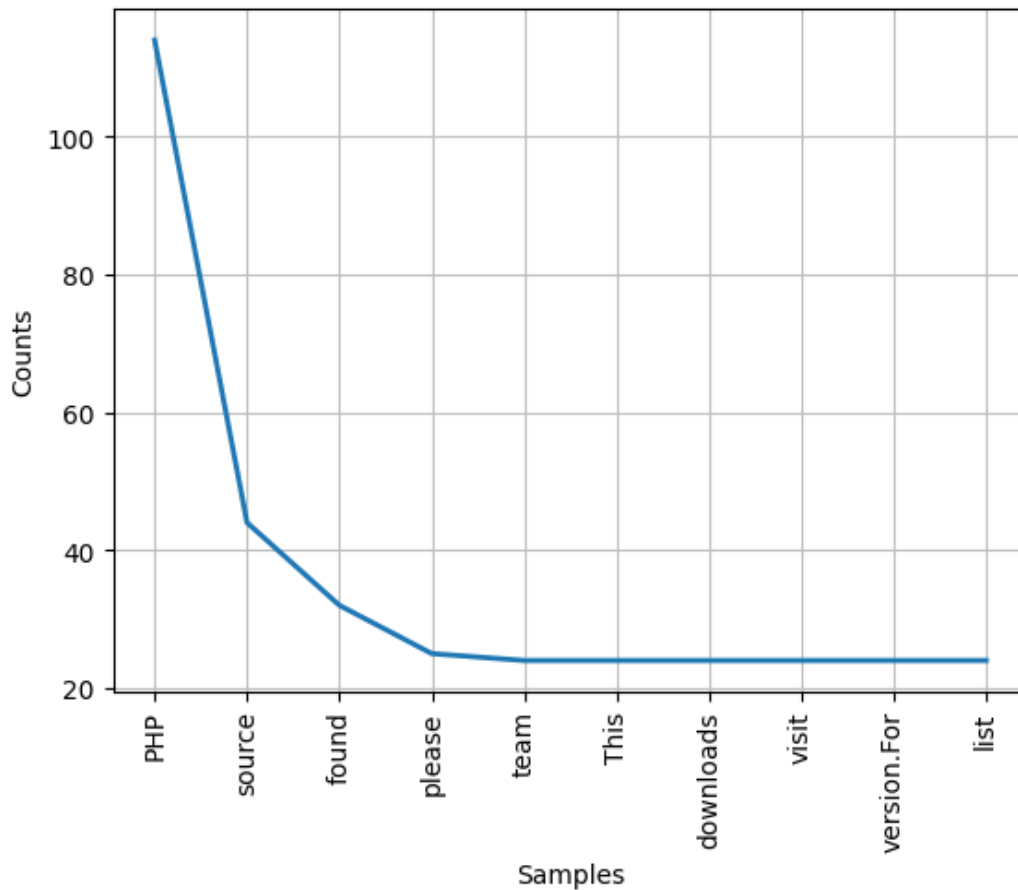
site.Thank:4
helping:4
us:4
make:4
better.01:2
Alpha:11
4:2
second:2
testing:3
4.:1
continues:2
theNEWSfile,:3
1,:1
15:1
8.2.22:2
Released!The:20
development:20
announces:20
immediate:20
availability:20
8.2.22.:1
bug:20
fix:15
release.All:20
8.2:9
users:21
encouraged:20
upgrade:21
ourdownloads:20
page,:20
Windows:21
binaries:20
onwindows.php.net/download/.:20
The:21
changes:20
recorded:20
theChangeLog.01:1
8.3.10:2
8.3.10.:1
8.3:8
theChangeLog.18:3
Jul:5
2:2
2.:1
3,:1
1:3
better.05:1
1.:1
starts:1
usingthe:1
tracking:1
2,:1
18:1
themanifestor:1
theQA:1
better.04:1
8.3.9:2
8.3.9.:1
theChangeLog.04:1
8.2.21:2
8.2.21.:1
theChangeLog.06:3
Jun:3
8.2.20:2
8.2.20.:1
security:7
8.3.8:2
8.3.8.:1
8.1.29:2
8.1.29.:1
8.1:3
theChangeLog.09:2

May:2
8.3.7:2
8.3.7.:1
8.2.19:2
8.2.19.:1
theChangeLog.24:1
Apr:4
2024Statement:1
glibc/iconv:1
VulnerabilityEDIT:1
2024-04-25.:1
Clarified:1
application:4
vulnerable:2
bug.Recently,:1
inglibcversion:1
2.39:1
older:1
(CVE-2024-2961):1
uncovered:1
buffer:2
overflow:2
character:8
set:5
conversionstothe:1
ISO-2022-CN-EXT:1
result:1
code:1
execution.This:1
specific:2
glibc:3
exploitable:3
PHP,:2
exposes:1
iconv:3
functionality:1
conversions:1
via:1
theiconv:1
extension.:1
Although:1
context:1
Engine,:1
PHP.:1
It:1
directly:1
remotely.The:1
exploitable,if:1
if,:1
calls:1
iconvfunctionsorfilterswith:1
user-supplied:2
sets.Applications:1
arenotvulnerable:1
if:Glibc:1
updates:1
distribution:3
installed.Or:1
extension:1
loaded.Or:1
removed:1
gconv-modules-extra.conf.Or:1
passes:1
specifically:1
allowed:2
sets:1
iconv.Moreover,:1
using:2
set,:1
good:1
practice:1
applications:1

accept:1
charsets:1
explicitly:1
application.:1
One:1
example:1
done:2
allow-list:1
thearray_search() function:1
check:2
encoding:1
passing:1
iconv.:1
For:1
example:array_search(\$charset,:1
\$allowed_list,:1
true)There:1
numerous:1
reports:1
online:1
titles:2
like:1
"Mitigating:1
Vulnerability:2
(CVE-2024-2961)":1
"PHP:1
Under:1
Attack".:1
misleading:1
isnota:1
itself.If:1
vulnerable,:1
recommend:1
Linux:3
already:2
published:2
patched:2
variants:1
glibc.Debian,:1
CentOS,:1
others,:1
so,:1
soon:1
possible.Once:1
update:2
glibc,:2
updating:1
package:1
machine:1
enough:1
alleviate:1
issue.:2
You:1
need:1
dynamically:1
linked:1
library.If:1
However,:1
exists:1
workaround:1
described:1
inGLIBC:1
Servers:1
Serving:1
PHPwhich:1
explains:1
way:1
remove:1
problematic:1
glibc.:1
Perform:1
procedure:1

every:1
gconv-modules-extra.conf:1
system.PHP:1
affected.Therefore,:1
released:1
vulnerability.12:1
8.1.28:2
8.1.28.:1
theChangeLog.11:2
8.3.6:2
8.3.6.:1
8.2.18:2
8.2.18.:1
theChangeLog.14:2
Mar:2
8.3.4:2
8.3.4.:1
8.2.17:2
8.2.17.:1
theChangeLog.15:2
Feb:2
8.2.16:2
8.2.16.:1
8.3.3:2
8.3.3.:1
Jan:2
8.2.15:2
8.2.15.:1
8.3.2:2
8.3.2.:1
theChangeLog.21:2
Dec:2
2023PHP:2
8.2.14:2
8.2.14.:1
8.1.27:2
8.1.27.:1
theChangeLog.Older:1
News:1
EntriesThe:1
FoundationThe:1
Foundation:1
collective:1
people:1
organizations,:1
united:1
mission:1
ensure:1
long-term:1
prosperity:1
language.DonateUpcoming:1
conferencesPHP:1
Russia:1
2024CascadiaPHP:1
2024International:1
Conference:1
Munich:1
2024Forum:1
2024:1
-:1
Paris:1
(France)User:1
Group:1
EventsSpecial:1
ThanksSocial:1
media@official_php@php@fosstodon.orgCopyright:1
©:1
2001-2024:1
GroupMy:1
PHP.netContactOther:1
PHP.net:1
sitesPrivacy:1

policyView:1
Source:1



Out[15]:

<Axes: xlabel='Samples', ylabel='Counts'>

PIE CHART REPRESENTATION

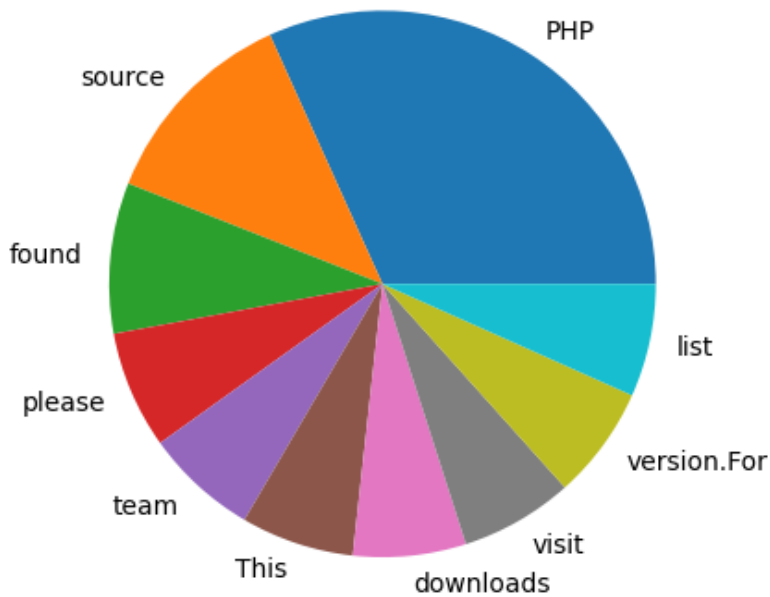
In [16]:

```
#code block 7
#pie chart representation for the 10 high frequency tokens from spaces web apge
import matplotlib.pyplot as plt
a = dict(freq)
sort = sorted(a, key=a.get,reverse=True)
count =0
b = {}
for i in sort:
    b[i] = a[i]
    count+= 1
    if count == 10:
        break
x = b.keys()
v = b.values()
plt.pie(v,labels=x)
```

Out[16]:

```
([<matplotlib.patches.Wedge at 0x7f544d577af0>,
 <matplotlib.patches.Wedge at 0x7f544d577a30>,
 <matplotlib.patches.Wedge at 0x7f544d4e02e0>,
 <matplotlib.patches.Wedge at 0x7f544d4e0760>,
 <matplotlib.patches.Wedge at 0x7f544d4e0be0>,
 <matplotlib.patches.Wedge at 0x7f544d4e1060>,
 <matplotlib.patches.Wedge at 0x7f544d4e14e0>,
 <matplotlib.patches.Wedge at 0x7f544d4e1960>,
 <matplotlib.patches.Wedge at 0x7f544d4e1de0>,
 <matplotlib.patches.Wedge at 0x7f544d4e2260>],
 [Text(0.5965441492138629, 0.9241942858721367, 'PHP'),
 Text(-0.7963087403903862, 0.7588757408020609, 'source'),
 Text(0.1334337500750075, 0.1357000000140000, 'found'),
 Text(0.1334337500750075, 0.1357000000140000, 'please'),
 Text(0.1334337500750075, 0.1357000000140000, 'team'),
 Text(0.1334337500750075, 0.1357000000140000, 'This'),
 Text(0.1334337500750075, 0.1357000000140000, 'downloads'),
 Text(0.1334337500750075, 0.1357000000140000, 'visit'),
 Text(0.1334337500750075, 0.1357000000140000, 'version.For'),
 Text(0.1334337500750075, 0.1357000000140000, 'list')])
```

```
Text(-1.0949075927528056, 0.1057230501463933, 'found'),
Text(-1.0120740956375274, -0.43093621910844476, 'please'),
Text(-0.7412743933683253, -0.8127190619989305, 'team'),
Text(-0.3454069061146697, -1.0443629968589907, 'This'),
Text(0.11051284885699807, -1.094434516194327, 'downloads'),
Text(0.5472188965508005, -0.9542282113088695, 'visit'),
Text(0.8887857472734954, -0.6481202785313037, 'version.For'),
Text(1.0758287633699581, -0.22933048621120317, 'list'))]
```



BAR GRAPH REPRESENTATION

In [17]:

#EXAMPLE

```
import nltk
nltk.download('punkt') # Download necessary data
from nltk.tokenize import word_tokenize
import matplotlib.pyplot as plt
from collections import Counter

# Sample text
text = "NLTK is a powerful library for natural language processing. NLTK provides tools f
or text analysis."

# Tokenize the text
tokens = word_tokenize(text)

# Count word frequencies
word_counts = Counter(tokens)
print(word_counts)
# Separate words and their frequencies
words = list(word_counts.keys())
frequencies = list(word_counts.values())

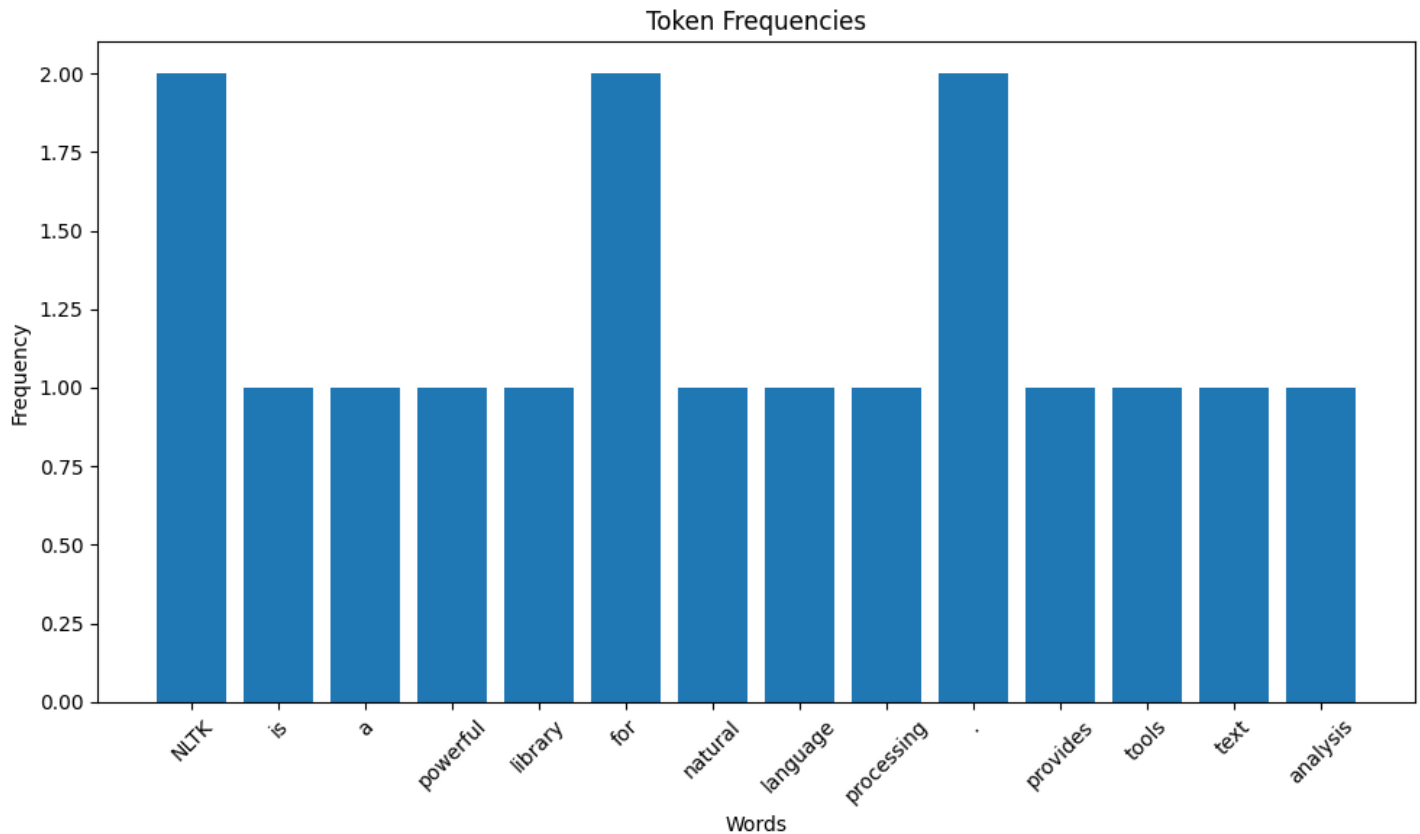
# Create a bar graph
plt.figure(figsize=(10,6))
plt.bar(words, frequencies)
plt.xlabel('Words')
plt.ylabel('Frequency')
plt.title('Token Frequencies')
plt.xticks(rotation=45)

# Display the graph
plt.tight_layout()
plt.show()
```



```
[nltk_data] Downloading package punkt to /root/nltk_data...  
[nltk_data] Package punkt is already up-to-date!
```

```
Counter({'NLTK': 2, 'for': 2, '.': 2, 'is': 1, 'a': 1, 'powerful': 1, 'library': 1, 'natu  
ral': 1, 'language': 1, 'processing': 1, 'provides': 1, 'tools': 1, 'text': 1, 'analysis'  
: 1})
```



Question 4

Work on code block 2 and use the same code concept below to parse data from any live website that talk about SpaceX Technologies and plot any graph for those tokens of your own choice which is not given above.

Note: Please keep in mind that do not put all code in one block split in a fashionable way.

In [20]:

```
import requests  
from bs4 import BeautifulSoup  
import nltk  
from nltk.tokenize import word_tokenize  
import matplotlib.pyplot as plt  
from collections import Counter  
  
# Download necessary NLTK data  
nltk.download('punkt')  
  
# Fetch the data from the website  
url = 'https://www.spacex.com/' # Example SpaceX website  
response = requests.get(url)  
html = response.text  
  
# Parse the HTML content  
soup = BeautifulSoup(html, 'html.parser')  
  
# Extract text from the parsed HTML  
text = soup.get_text()
```

```

# Tokenize the text into words
tokens = word_tokenize(text)

# Filter out non-alphabetic tokens and convert to lowercase
words = [word.lower() for word in tokens if word.isalpha()]

# Compute word frequencies
word_freq = Counter(words)

# Extract the most common words and their frequencies
most_common_words = word_freq.most_common(10)

# Split the words and counts for plotting
words, counts = zip(*most_common_words)

# Generate numeric positions for each word (for the x-axis)
x_positions = range(len(words))

# Plot the data
plt.figure(figsize=(10, 6))
plt.scatter(x_positions, counts, color='blue', s=100) # s is the size of the points
plt.xticks(x_positions, words, rotation=45)
plt.xlabel('Words')
plt.ylabel('Frequency')
plt.title('Scatter Plot of Top 10 Most Common Words on SpaceX Website')
plt.show()

```

```

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!

```

