

ADVANCE PYTHON – LAB ASSIGNMENT

15-07-25

STRING OPERATIONS IN PYTHON

22MID0127

GOKULA J

The screenshot displays a Linux desktop environment. The top window is a web browser showing the Gutenberg.org page for the eBook 'The glad lady' by Amy E. Blanchard. The page content includes the title, author, release date (July 13, 2025), and a detailed copyright notice. The bottom window is a JupyterLab notebook titled 'Untitled6'. The notebook contains Python code for fetching and processing the eBook text. The code includes imports for 'urllib.request' and 'nltk.tokenize', followed by a request to the Gutenberg.org URL. The output shows the raw text of the eBook, including the title and author information. A syntax error is visible in the final cell of the notebook, indicating an issue with the 'word_tokenize' function call.

```
[8]: from urllib import request
[9]: url="https://www.gutenberg.org/cache/epub/76495/pg76495.txt"
[10]: response=request.urlopen(url)
[11]: raw=response.read().decode("utf8")
[12]: print(len(raw))
[13]: 381845
[14]: !pip install nltk
Requirement already satisfied: nltk in /home/software/software/lib/python3.12/site-packages (3.8.1)
Requirement already satisfied: click in /home/software/software/lib/python3.12/site-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in /home/software/software/lib/python3.12/site-packages (from nltk) (1.4.2)
Requirement already satisfied: regex==2021.8.3 in /home/software/software/lib/python3.12/site-packages (from nltk) (2023.10.3)
Requirement already satisfied: tqdm in /home/software/software/lib/python3.12/site-packages (from nltk) (4.66.4)
[15]: import nltk
[16]: from nltk.tokenize import word_tokenize url="https://www.gutenberg.org/cache/epub/76495/pg76495.txt"
response=request.urlopen(url)
response=request.urlopen(url)
tokens=word_tokenize(raw)
cell In[18]: line 1
from nltk.tokenize import word_tokenize url="https://www.gutenberg.org/cache/epub/76495/pg76495.txt"
SyntaxError: invalid syntax
```

```
Activities Jul 15 9:05 AM
Home x Untitled6 x what are taken in a par x gutenberg.org/cache/e x WhatsApp x +
localhost:8888/notebooks/Untitled6.ipynb

jupyter Untitled6 Last Checkpoint: 56 minutes ago
File Edit View Run Kernel Settings Help Trusted
+ - X Copy Paste Undo Redo Code

JupyterLab Python 3 (ipykernel)

[2]: n1="welcome"
    n2="VIT"
    print(n1+"to"+n2)

    welcome to VIT

[3]: print(n1+ " to "+n2)

    welcome to VIT

[4]: print(n1.title())

    Welcome

[5]: print(n1.upper())

    WELCOME

[7]: !pip install jieba

Collecting jieba
  Downloading jieba-0.42.1.tar.gz (19.2 MB)
    Preparing metadata (setup.py) ... done
Building wheels for collected packages: jieba
  Building wheel for jieba (setup.py) ... done
  Created wheel for jieba: filename=jieba-0.42.1-py3-none-any.whl size=19314459 sha256=b226afc5895d65771e2a224888e97465d581bcfa172e8d28387
  Stored in directory: /home/matlab/.cache/pip/wheels/08/a1/a3/5c8ac52c2f5782ffffc34c95c57c8e5ec3063dc69541e7c
Successfully built jieba
Installing collected packages: jieba
Successfully installed jieba-0.42.1

[8]: from urllib import request

[9]: url="https://www.gutenberg.org/cache/epub/76495/pg76495.txt"

[10]: response=request.urlopen(url)

[11]: raw=response.read().decode("utf8")

[12]: print(len(raw))

    381845

[14]: !pip install nltk

Requirement already satisfied: nltk in /home/software/software/lib/python3.12/site-packages (3.8.1)
Requirement already satisfied: click in /home/software/software/lib/python3.12/site-packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in /home/software/software/lib/python3.12/site-packages (from nltk) (1.4.2)
Requirement already satisfied: regex<2021.8.3 in /home/software/software/lib/python3.12/site-packages (from nltk) (2023.10.3)
Requirement already satisfied: tqdm in /home/software/software/lib/python3.12/site-packages (from nltk) (4.66.4)

[23]: import nltk

[25]: nltk.download('punkt')

[nltk_data] Downloading package punkt to /home/matlab/nltk data...
[nltk_data] Unzipping tokenizers/punkt.zip.

[25]: True

[26]: from nltk.tokenize import word_tokenize

    url="https://www.gutenberg.org/cache/epub/76495/pg76495.txt"
    response=request.urlopen(url)
    tokens=word_tokenize(raw)
    print(tokens[:200])

    ['\u00ffThe', 'Project', 'Gutenberg', 'eBook', 'of', 'The', 'glad', 'lady', 'This', 'eBook', 'is', 'for', 'the', 'use', 'of', 'anyone', 'an', 'ywhere', 'in', 'the', 'United', 'States', 'and', 'most', 'other', 'parts', 'of', 'the', 'world', 'at', 'no', 'cost', 'and', 'with', 'almost', 'no', 'restrictions', ' WHATSOEVER', 'You', 'may', 'copy', 'it', 'give', 'it', 'away', 'or', 're-use', 'it', 'under', 'the', 'terms', 'of', 'the', 'Project', 'Gutenberg', 'License', 'included', 'with', 'this', 'eBook', 'or', 'online', 'at', 'www.gutenberg.org', 'If', 'you', 'are', 'not', 'located', 'in', 'the', 'United', 'States', 'you', 'will', 'have', 'to', 'check', 'the', 'laws', 'of', 'the', 'country', 'where', 'you', 'are', 'located', 'before', 'using', 'this', 'eBook', 'The', 'glad', 'lady', 'Author', 'Amy', 'E', 'Blanchard', 'Release', 'date', 'July', '13', '2025', 'eBook', '76495', 'Language', 'English', 'Original', 'publication', 'Boston', 'Dana', 'Estes', 'Company', '1910', 'Credits', 'Carol', 'Brown', 'Adriana', 'and', 'the', 'Online', 'Distributed', 'Proofreading', 'Team', 'at', 'https://www.pgdp.net', 'This', 'book', 'was', 'produced', 'from', 'images', 'made', 'available', 'by', 'the', 'HathiTrust', 'Digital', 'Library', 'ART', 'OF', 'THE', 'PROJECT', 'GUTENBERG', 'EBOOK', 'THE', 'GLAD', 'LADY', 'BY', 'AMY', 'E.', 'BLANCHARD', 'Author', 'BEHIND', 'THE', 'HEDGE', 'AND', 'WAITED', 'I', 'THE', 'GLAD', 'LADY', 'BY', 'AMY', 'E.', 'BLANCHARD', 'Author']

[27]: print(len(tokens))

    83080

[28]: print(type(raw))

    <class 'str'>

[ ]:
```