ADVANCE PYTHON – LAB ASSIGNMENT   15-07-25

STRING OPERATIONS IN PYTHON

22MID0127

GOKULA J

Top browser window — Jupyter Notebook (Untitled6):

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

     welcometoVIT

[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)
                                          19.2/19.2 MB 5.5 MB/s eta 0:00:00m eta 0:00:01[36m0:00:01
       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=b226afc5805dd5771e2a2224088e97465d581bcfa172edd28307
     982925ebef3c
       Stored in directory: /home/matlab/.cache/pip/wheels/08/a1/a3/5c8ac52cc2f5782ffffc34c95c57c8e5ecb3063dc69541ee7c
     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")

[13]: print(len(raw))

      381845
```

Bottom screen — continuation of the Jupyter Notebook:

```
      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])

      ['\ufeffThe', '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', 'almos
      t', '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', '.', 'Title', ':', 'The', 'glad', 'lady', 'Author',
      ':', 'Amy', 'E', 'Blanchard', 'Release', 'date', ':', 'July', '13', ',', '2025', '[', 'eBook', '#', '76495', ']', 'Language', ':', 'Englis
      h', 'Original', 'publication', ':', 'Boston', ':', 'Dana', 'Estes', '&', 'Company', ':', '1910', 'Credits', ':', 'Carol', 'Brown', ',', 'Aa
      ron', 'Adrignola', 'and', 'the', 'Online', 'Distributed', 'Proofreading', 'Team', 'at', 'https', ':', '//www.pgdp.net', '(', 'This', 'boo
      k', 'was', 'produced', 'from', 'images', 'made', 'available', 'by', 'the', 'HathiTrust', 'Digital', 'Library', '.', ')', '*', '*', '*', 'ST
      ART', 'OF', 'THE', 'PROJECT', 'GUTENBERG', 'EBOOK', 'THE', 'GLAD', 'LADY', '*', '*', '*', '[', 'Illustration', ':', '"', 'PATTY', 'CREPT',
      'BEHIND', 'THE', 'HEDGE', 'AND', 'WAITED', '"', ']', 'THE', 'GLAD', 'LADY', 'BY', 'AMY', 'E.', 'BLANCHARD', 'Author']

[27]: print(len(tokens))

      83080

[28]: print(type(raw))

      <class 'str'>
```