

Bài tập Xử lý văn bản

- Bài 1: Tìm hiểu thư viện nltk: <https://www.nltk.org/>, thực hiện các công việc tiền xử lý một văn bản với thư viện nltk.
 - Tách câu, Tách từ,
 - Tính tần suất xuất hiện các từ
 - Xóa từ dừng
 - Tìm gốc của một từ
 - v.v.

Tham khảo

- <https://www.youtube.com/watch?v=FLZvOKSCkxY&list=PLQVvvaa0QuDf2JswnfGkliBlnZnIC4HL>
- <https://www.youtube.com/watch?v=lqPWZL5f-7g&list=PLcTXcpndN-Sl9eYrKM6jtcOTgC52EJnqH>

Bài tập Xử lý văn bản

- Bài 2: Viết hàm/chương trình chuẩn hóa 1 đoạn văn tiếng Việt (tự lựa chọn ngôn ngữ lập trình)

Ví dụ đầu vào, đầu ra của hàm/chương trình như sau:

Đoạn văn Đầu vào: “Tỷ giá trung tâm giữa đồng Việt Nam (VND) và đô la Mỹ (USD) được NHNN công bố hôm 22/9 ở mức 24.060 đồng mỗi USD. Giá USD tại các ngân hàng thương mại hôm 23/9 được giao dịch quanh mức 24.160-24.530 đồng/USD (mua vào - bán ra).”



Đoạn văn Kết quả: “*tỷ giá trung tâm giữa đồng **việt nam về nợ đê** và đô la **mỹ u ét đê** được **ngân hàng nhà nước** công bố hôm **hai mươi hai tháng chín** ở mức **hai mươi tư nghìn không trăm sáu mươi đồng mỗi u ét đê**. **giá u ét đê** tại các ngân hàng thương mại hôm **hai mươi ba tháng chín** được giao dịch quanh mức **hai mươi tư nghìn một trăm sáu mươi tới hai mươi tư nghìn năm trăm ba mươi đồng trên u ét đê** mua vào và bán ra”*