**Natural Language Processing.**

1. **Giới thiệu.**

Natural Language Processing (NLP) là một bộ phận thuộc Computer Science, nghiên cứu các phương pháp về việc giao tiếp giữa người và một hệ thống thông minh sử dụng ngôn ngữ tự nhiên, chẳng hạn tiếng Anh.

Việc xử lý ngôn ngữ tự nhiên là cần thiết khi chúng ta muốn một hệ thống thông minh, chẳng hạn robot, thực hiện một số việc mà chúng ta yêu cầu thông qua giọng nói hoặc khi ta muốn dịch một đoạn văn bản từ ngôn ngữ này sang ngôn ngữ khác sao cho tự nhiên.

NLP quan tâm chủ yếu đến việc khiến máy tính thực hiện một số tác vụ hữu ích trên ngôn ngữ tự nhiên của con người. Dữ liệu của một hệ thống NLP thường là một đoạn thoại hoặc văn bản.

1. **Các thành phần của NLP.**

NLP có 2 thành phần:

1. **Natural Language Understading (NLU):**

Việc hiểu một ngôn ngữ tự nhiên yêu cầu các công việc sau:

* Ánh xạ dữ liệu đầu vào là ngôn ngữ tự nhiên thành các hình thức hữu ích khác.
* Phân tích các khía cạnh khác nhau của ngôn ngữ.

1. **Natural Language Generation (NLG):**

NLG là quá trình tạo ra các đoạn văn bản hoặc các câu văn có nghĩa trong ngôn ngữ tự nhiên từ các dữ liệu sẵn có:

* Text planning – truy xuất nội dung có liên quan từ cơ sở dữ liệu.
* Sentence planning – lựa chọn các từ ngữ cần thiết, tạo thành các câu văn có nghĩa, thiết lập âm điệu cho câu văn.
* Text realization – chưa dịch được.

Thành phần NLU được xem như khó hơn thành phần NLG.

1. **Các khó khăn đối với NLU.**

Cấu trúc và ngữ nghĩa của ngôn ngữ tự nhiên vô cùng đa dạng:

* Sự mơ hồ trong từ vựng: một từ có thể mang nhiều nghĩa hoặc nhiều loại từ, chẳng hạn từ “process” vừa có thể là danh từ, vừa có thể là động từ.
* Sự mơ hồ trong cú pháp: một câu văn có thể hiểu theo nhiều cách, ví dụ: câu “Children make delicious snacks” vừa có thể hiểu là “Trẻ em làm những món ăn vặt ngon” và cũng có thể hiểu là “Trẻ em là những món ăn vặt ngon”.
* Sự mơ hồ về chủ thể: ví dụ trong câu “Bob is doing homework with Bill. Suddenly, he shouts “Eureka” and runs to the bathroom.” thì từ “he” đanh nói đến ai?
* Một từ hoặc câu văn có thể mang nhiều nghĩa.  
  Ví du: từ “orange” vừa mang nghĩa là “trái cam”, vừa mang nghĩa là “màu cam”.
* Nhiều từ hoặc câu văn khác nhau có thể mang cùng một nghĩa.  
  Ví dụ: cả hai câu “No one in the class has passed the exam” và “Everyone in the class has failed the exam” đều mang nghĩa là “mọi người trong lớp đều không qua được kì kiểm tra”.