**Bài 10: Natural Language Processing - Lập trình AI bằng Python**

Natural Language Processing (NLP) đề cập đến phương pháp giao tiếp AI với các hệ thống thông minh bằng ngôn ngữ tự nhiên như tiếng Anh.

Xử lý ngôn ngữ tự nhiên là bắt buộc khi bạn muốn một hệ thống thông minh như robot thực hiện theo hướng dẫn của bạn, khi bạn muốn nghe quyết định từ hệ thống chuyên gia lâm sàng dựa trên đối thoại, v.v.

Lĩnh vực NLP liên quan đến việc làm cho máy tính thực hiện các nhiệm vụ hữu ích với ngôn ngữ tự nhiên mà con người sử dụng. Đầu vào và đầu ra của một hệ thống NLP có thể là speech và writen Text

**1. Thành phần NLP**

Ta sẽ tìm hiểu về các thành phần khác nhau của NLP. Có hai thành phần của NLP bao gồm :

**a. Natural Language Understanding (NLU) :**

Nó bao gồm các nhiệm vụ sau:

* Ánh xạ đầu vào đã cho bằng ngôn ngữ tự nhiên thành các biểu diễn hữu ích.
* Phân tích các khía cạnh khác nhau của ngôn ngữ.

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

Đó là quá trình tạo ra các cụm từ và câu có nghĩa dưới dạng ngôn ngữ tự nhiên từ một số biểu diễn bên trong. Nó liên quan đến các thành phần sau :

* **Text Planning** : Điều này bao gồm truy xuất nội dung có liên quan từ cơ sở kiến thức.
* **Sentence planning :**Điều này bao gồm việc lựa chọn các từ cần thiết, tạo thành các cụm từ có nghĩa, thiết lập giọng điệu của câu.
* **Text Realization :**Đây là ánh xạ kế hoạch câu thành cấu trúc câu.

**2. Những khó khăn trong NLU**

NLU rất phong phú về hình thức và cấu trúc, tuy nhiên, nó là mơ hồ. Có thể có nhiều mức độ mơ hồ khác nhau :

* **Sự mơ hồ về ngữ nghĩa** : Nó ở cấp độ rất sơ khai chẳng hạn như cấp độ từ. Ví dụ, coi từ “board” là danh từ hay động từ?
* **Mức độ không rõ ràng về cú pháp**: Một câu có thể được phân tích cú pháp theo nhiều cách khác nhau. Ví dụ: "Anh ấy nâng con bọ có mũ màu đỏ." - Anh ta dùng nắp để nhấc con bọ lên hay anh ta nhấc một con bọ cánh cứng có nắp màu đỏ?
* **Tham chiếu không rõ ràng :**Đề cập đến điều gì đó bằng cách sử dụng đại từ. Ví dụ, Rima đã đến Gauri. Cô ấy nói, "Tôi mệt mỏi." - Chính xác là ai đang mệt?

**3. Thuật ngữ NLP :**

Bây giờ chúng ta hãy xem một vài thuật ngữ quan trọng trong thuật ngữ NLP :

* **Phonology** : Nghiên cứu về tổ chức âm thanh một cách hệ thống.
* **Morphology** : Nghiên cứu về cấu tạo của từ từ các đơn vị có nghĩa ban đầu.
* **Morpheme :**Đơn vị nghĩa nguyên thủy trong một ngôn ngữ.
* **Syntax :**Đề cập đến việc sắp xếp các từ để tạo thành một câu. Nó cũng liên quan đến việc xác định vai trò cấu trúc của các từ trong câu và trong các cụm từ.
* **Semantics :** Liên quan đến nghĩa của từ và cách kết hợp các từ thành các cụm từ và câu có nghĩa.
* **Pragmatics :**đề cập đến việc sử dụng và hiểu các câu trong các tình huống khác nhau và cách thức giải thích câu bị ảnh hưởng.
* **Discourse :**đề cập đến việc câu ngay trước đó có thể ảnh hưởng như thế nào đến việc giải thích câu tiếp theo.
* **World Knowledge :**Bao gồm những kiến thức chung về thế giới.

**4. Các bước trong NLP :**

**a. Phân tích từ vựng :**

Nó liên quan đến việc xác định và phân tích cấu trúc của từ. Từ vựng của một ngôn ngữ có nghĩa là tập hợp các từ và cụm từ trong một ngôn ngữ. Phân tích từ vựng là phân chia toàn bộ đoạn văn bản thành các đoạn văn, câu và từ.

**b. Phân tích cú pháp :**

Nó bao gồm việc phân tích các từ trong câu để tìm ngữ pháp và sắp xếp các từ theo cách thể hiện mối quan hệ giữa các từ. Câu như "The school go to boy" bị từ chối bởi bộ phân tích cú pháp tiếng Anh.

**c. Phân tích ngữ nghĩa :**

Nó rút ra ý nghĩa chính xác hoặc nghĩa từ điển từ văn bản. Văn bản được kiểm tra xem có ý nghĩa hay không. Nó được thực hiện bằng cách ánh xạ các cấu trúc cú pháp và các đối tượng trong miền tác vụ. Bộ phân tích ngữ nghĩa bỏ qua câu chẳng hạn như "hot ice-cream".

**d. Tích hợp bài giảng :**

Ý nghĩa của bất kỳ câu nào phụ thuộc vào nghĩa của câu ngay trước nó. Ngoài ra nó còn mang ý nghĩa câu thành công ngay lập tức.

**e. Phân tích thực dụng :**

Trong lúc này, những gì đã nói sẽ được diễn giải lại dựa trên ý nghĩa thực sự của nó. Nó liên quan đến việc tạo ra những khía cạnh của ngôn ngữ đòi hỏi kiến thức thế giới thực.