**pos tagging in nlp**

* Mục tiêu chính của gắn thẻ Part-of-Speech (POS) là xác định nhóm ngữ pháp của một từ đã cho. Cho dù đó là một danh từ, đại từ, tính từ, động từ, trạng từ, giới từ, vv dựa trên bối cảnh. Gắn thẻ POS tìm kiếm các mối quan hệ trong câu và gán một thẻ tương ứng cho từ đó.



* Gắn thẻ POS là một giải pháp học tập có giám sát sử dụng các tính năng như từ trước đó, từ tiếp theo, chứ cái viết hoa đầu tiên… NLTK có chức năng nhận thẻ POS và nó hoạt động sau quá trình mã hóa.

**import** nltk  
sent = **"Albert Einstein was born in Ulm, Germany in 1879."**tokens=nltk.word\_tokenize(sent)  
print(tokens)

['Albert', 'Einstein', 'was', 'born', 'in', 'Ulm', ',', 'Germany', 'in', '1879', '.']

print(nltk.pos\_tag(tokens))

[('Albert', 'NNP'), ('Einstein', 'NNP'), ('was', 'VBD'), ('born', 'VBN'), ('in', 'IN'), ('Ulm', 'NNP'), (',', ','), ('Germany', 'NNP'), ('in', 'IN'), ('1879', 'CD'), ('.', '.')]

*# Chúng ta có thể biết thêm chi tiết về bất kỳ thẻ POS nào bằng cách sử dụng help funciton của NLTK như sau.*nltk.help.upenn\_tagset(**"NN"**)

NN: noun, common, singular or mass

common-carrier cabbage knuckle-duster Casino afghan shed thermostat

investment slide humour falloff slick wind hyena override subhumanity

machinist ...