**CHƯƠNG 3: CÀI ĐẶT GIẢI PHÁP**

1. **Cài đặt các thư viện**

Giống như các chương trình máy học khác, Data-to-Text Generation with Content Selection and Planning cũng cần có môi trường hoạt động bằng cách cài đặt các thư viện:

* Numpy
* Nltk
* Tensoflow

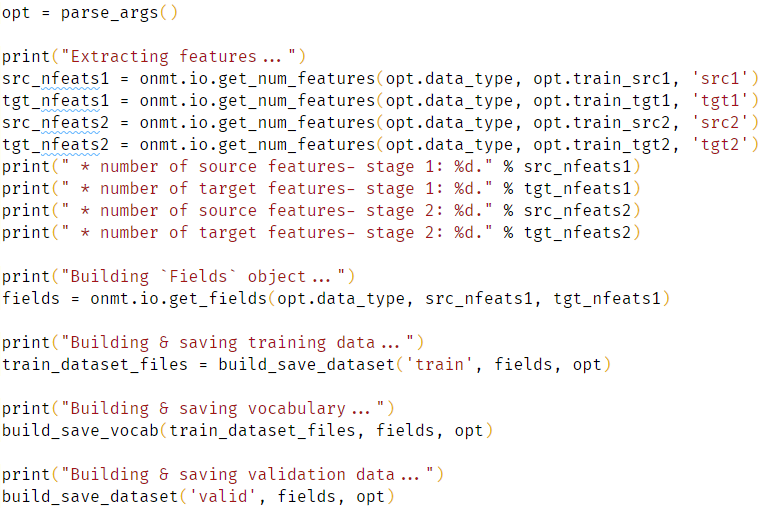
Ngoài ra, mô hình còn sử dụng python phiên bản 2.7 và Pytorch phiên bản 0.3.1, CUDA 8.0. Bên cạnh đó còn có các model được thiết lập sẵn như: AudioEncoder, Conv2Conv, ConvMultiStepAttention, Gate, CopyGenerator, GlobalAttention, ImageEncoder, PointerAttention, SRU, StackedRNN, StructuredAttention, Tranformer, UtilClass, WeightNorm.

1. **Tạo dữ liệu đầu vào**

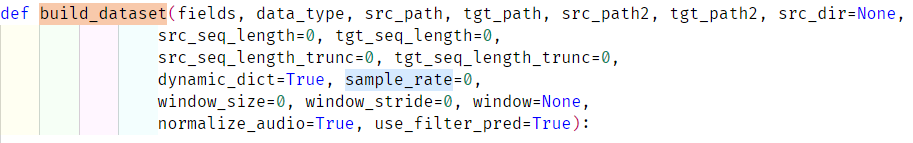
Đầu tiên phải khai báo biến $BASE để làm biến môi trường trong suốt quá trình cài đặt mô hình. Đồng thời tạo thư mục preprocess để lưu trữ dữ liệu đã chuyển đổi thành các véc-tơ.

Hình. Quá trình chuyển đổi dữ liệu ban đầu thành dữ liệu hệ thống

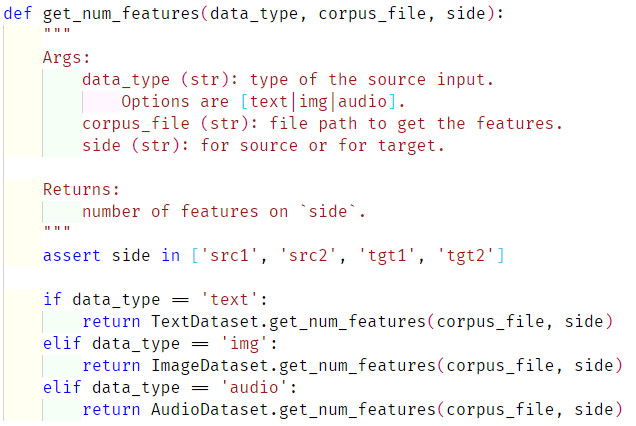
Dựa vào những thuật toán đã được trình bày trong thư mục onmt.io của dự án, mô hình sẽ áp dụng những thuật toán này vào nhiệm vụ mã hóa dữ liệu.



Hình. Sử dụng các thuật toán để chuyển đổi dữ liệu

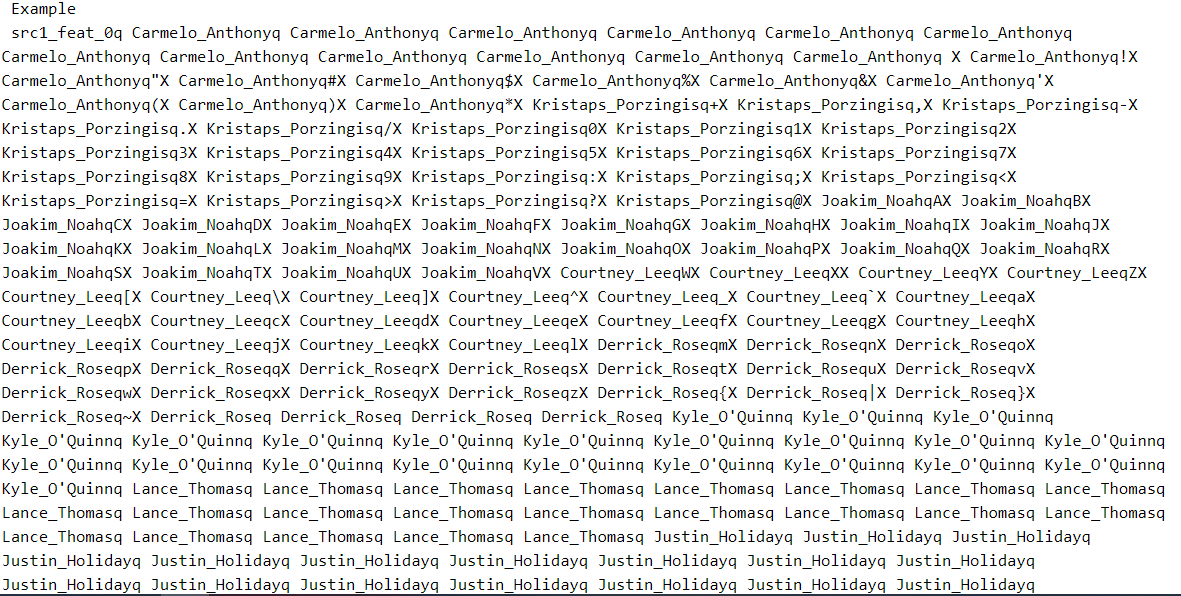


Hình. Hàm xây dụng dữ liệu



Hình. Hàm lấy dữ liệu đầu vào

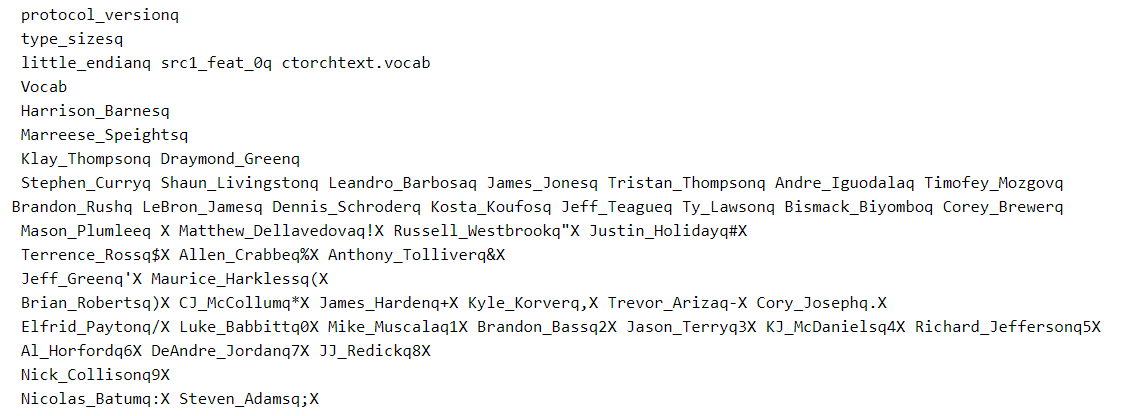
Sau khi thực hiện kết thúc quá trình mã hóa dữ liệu, tập dữ liệu sẽ được chia thành ba tập tin nằm trong thư mục preprocess: roto.train.1, roto.valid.1, roto.vocab.



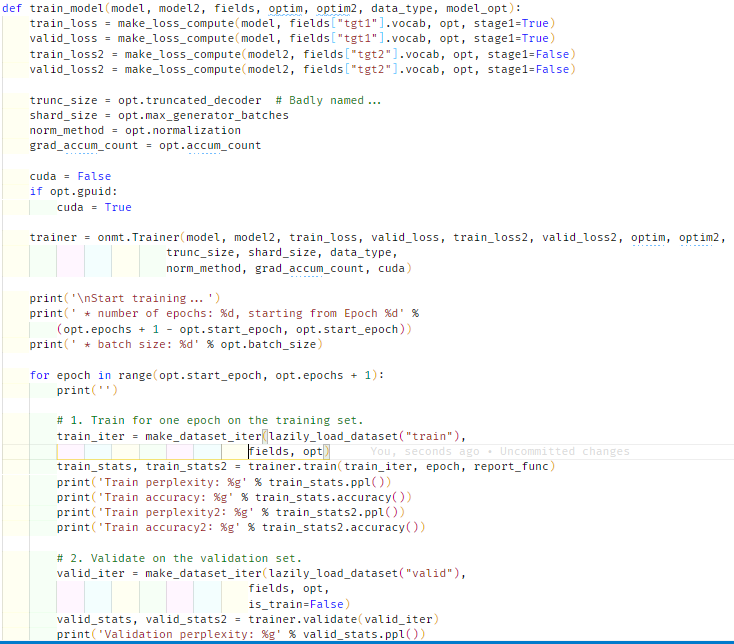
Hình. Dữ liệu đã được chuyển đổi

1. **Xây dựng mô hình**

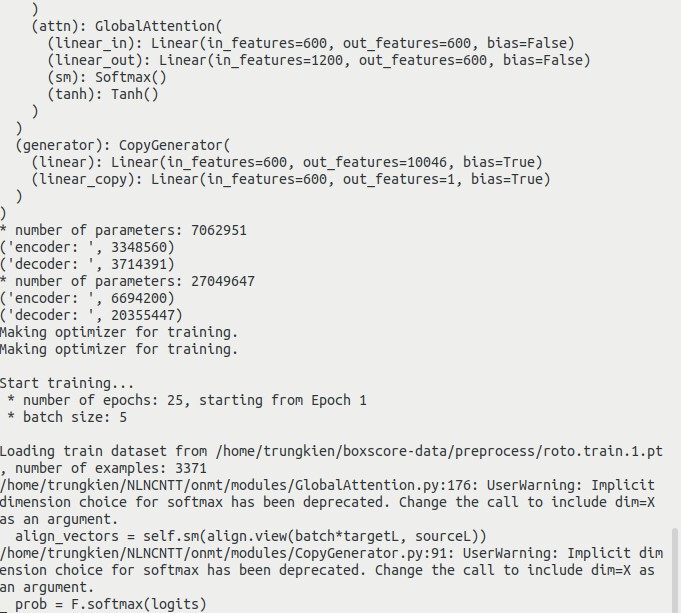
Sau khi đã có dữ liệu đã được mã hóa, mô hình sẽ được xây dụng dựa trên model kết hợp giữa Neural Content Planning (NPC) + Conditional Copy (CC).



Hình. Model NPC+CC



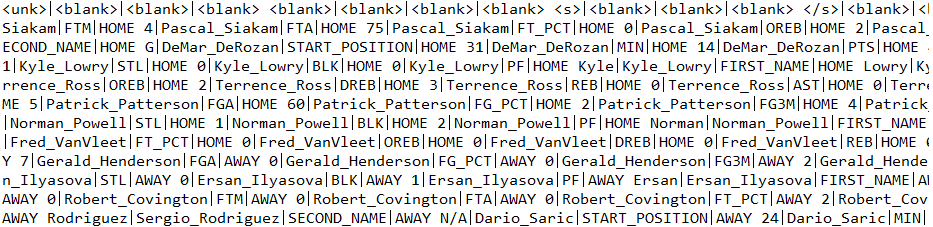
Hinh. Hàm xử lý xây dựng dữ liệu



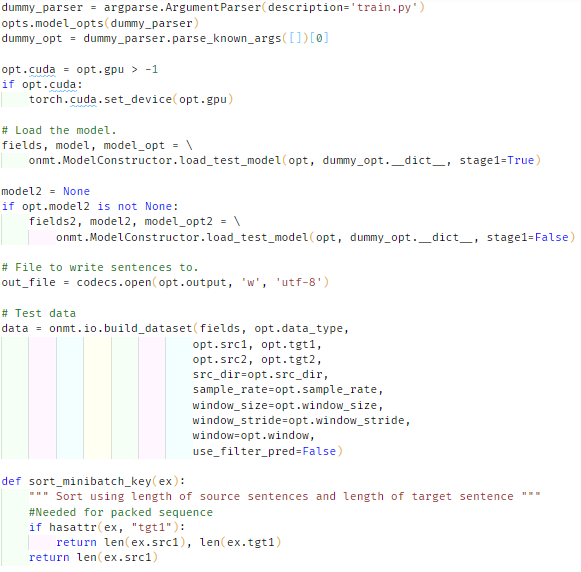
Hình. Quá trình xây dựng mô hình

1. **Tạo văn bản**

Để tạo được văn bản, chương trình cần tạo ra kế hoạch nội dung dựa vào mô hình đã được xây dựng bằng model NPC+CC.

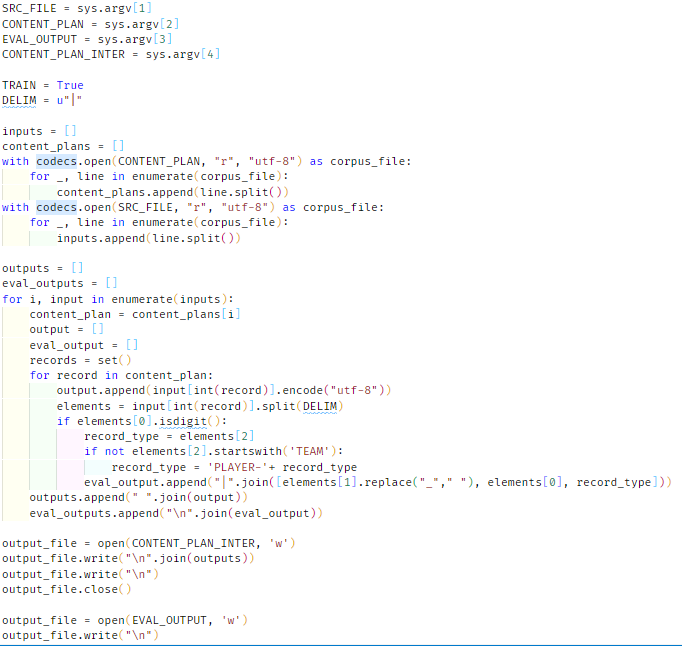


Hình. Nội dung kế hoạch



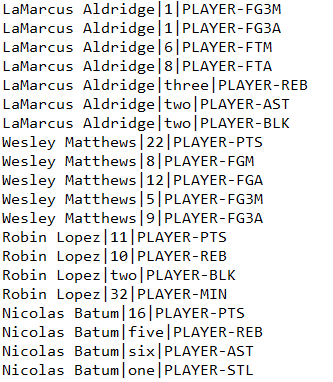
Hình. Hàm lập kế hoạch

Tiếp theo sẽ lập kế hoạch nội dung với các bản ghi từ đầu vào với các chỉ số.

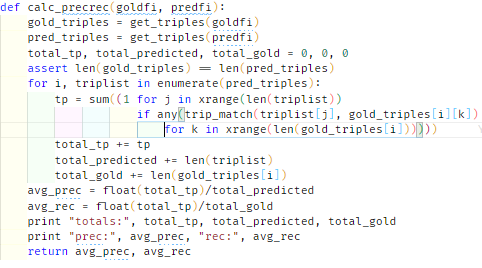


Hình.Hàm lập kế hoạch nội dung với các bản ghi từ đầu vào với các chỉ số

Sau đó sẽ đánh giá độ chính xác của kế hoạch nội dung trong giai đoạn đầu tiên.

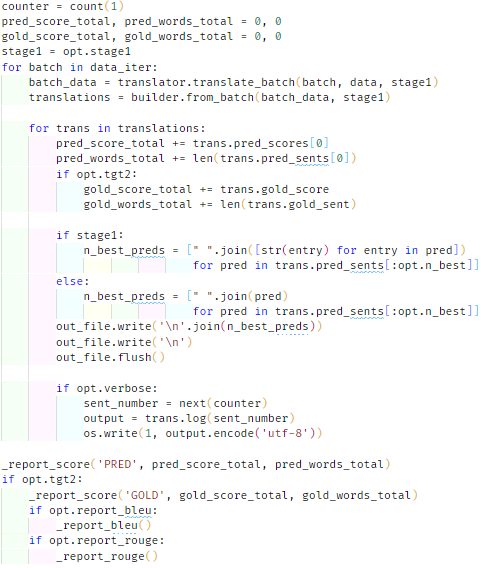


Hình. Nội dung đánh giá



Hình. Hàm đáh giá

Bước cuối cùng trong quá trình là tóm tắt dữ liệu đầu ra thành văn bản.



Hình. Hàm xử lý tóm tắt