Vấn đề: Nhận dạng các thông tin trong hoá đơn với các thông tin cần thiết: ngày xuất hoá đơn, số hoá đơn, tổng tiền.

* Điều kiện:
  + Các hoá đơn là khác nhau về bố cục, thứ tự các thông tin có trên hoá đơn, và được viết bằng 1 trong 3 dạng: tiếng anh, tiếng việt và cả tiếng anh lẫn tiếng việt.

Mục tiêu: trích xuất được các thông tin cần thiết: ngày xuất hoá đơn, số hoá đơn, tổng tiền.

Phân tích vấn đề, khó khăn và đưa ra đề xuất:

* Vấn đề 1 (trọng tâm):
  + Vấn đề: Bố cục hoá đơn là khác nhau, và được viết khác nhau về dạng ngôn ngữ.
  + Phương án đề ra: vấn đề này liên quan đến dữ liệu, một nguồn dữ liệu cần bao phủ được các bố cục đa dạng của hoá đơn. Cần thu thập thêm dữ liệu ( 1 ảnh là không đủ để mô hình nhận thấy và bao quát) và đảm bảo dữ liệu đủ tốt, không lỗi.
* Vấn đề 2:
  + Vấn đề: Chọn mô hình nhận dạng tốt. Nhưng đặc biệt phải xử lý được ngôn ngữ khác nhau và trong số chúng thì được viết thành bằng tiếng việt, ngôn ngữ này lại không thông dụng, in mờ làm mất đi dấu của từ.
  + Phương án đề ra: sau khi tra google và trang web [Hugging Face](https://huggingface.co/models?pipeline_tag=image-to-text&sort=trending) ( trang này lưu trữ mô hình cho các dạng bài toán khác nhau). Tìm hiểu các đặc điểm từng mô hình thì em tìm được mô hình [naver-clova-ix/donut-base-finetuned-docvqa](https://huggingface.co/naver-clova-ix/donut-base-finetuned-docvqa), mô hình này tập trung vào các vị trí, chứa giá trị không cần tiêu đề (khắc phục về vấn đề ngôn ngữ). Kiến trúc của mô hình:

A diagram of a computer

AI-generated content may be incorrect.

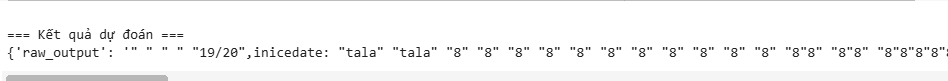
Thực hiện:

* Bước 1: đánh nhãn (bước này em đánh nhãn thủ công)
* Bước 2: fine tuning mô hình được xác định ở vấn đề 2 ở trên, tham khảo cách triển khai [documents](https://huggingface.co/docs/transformers/main/en/model_doc/donut) và tham khảo [trang fine-tuning-donut](https://www.philschmid.de/fine-tuning-donut?utm_source=chatgpt.com).

Một vấn đề gặp phải cực kì lớn đó là dữ liệu không đủ (trong bài này em chỉ có duy nhất 1 tấm ảnh), để train được 1 mô hình AI cần ít nhất khoảng 1 nghìn tấm ảnh khác nhau bao phủ vấn đề.

Do đó, em cho rằng chúng ta cần thu thập thật nhiều dữ liệu để bao quát vấn đề hơn.

Demo kết quả của mô hình, mô hình AI cần có rất nhiều dữ liệu mà em có 1 dữ liệu nên sự dự đoán của nó không tốt lắm, nếu có khoảng 1 nghìn ảnh trở lên thì hoàn toàn mô hình dự đoán chính xác cao:



- Vì mô hình fine tuning này lớn nên em không tải lên nỗi git.

- Để có mô hình cần chạy file: execute.ipynb

- Sau đó dự đoán trên file predict.py