BERT:

B: Bidirectional

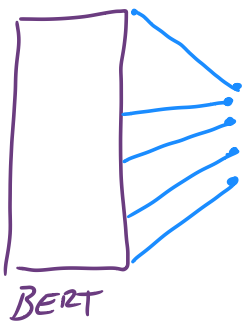
E: Encoder

R: Representations from

T: Transformers

→ Pretrain for unlabeled text (Unsupervised)

→ Joining left and right contexts



We only need one more layer to use this network for different tasks

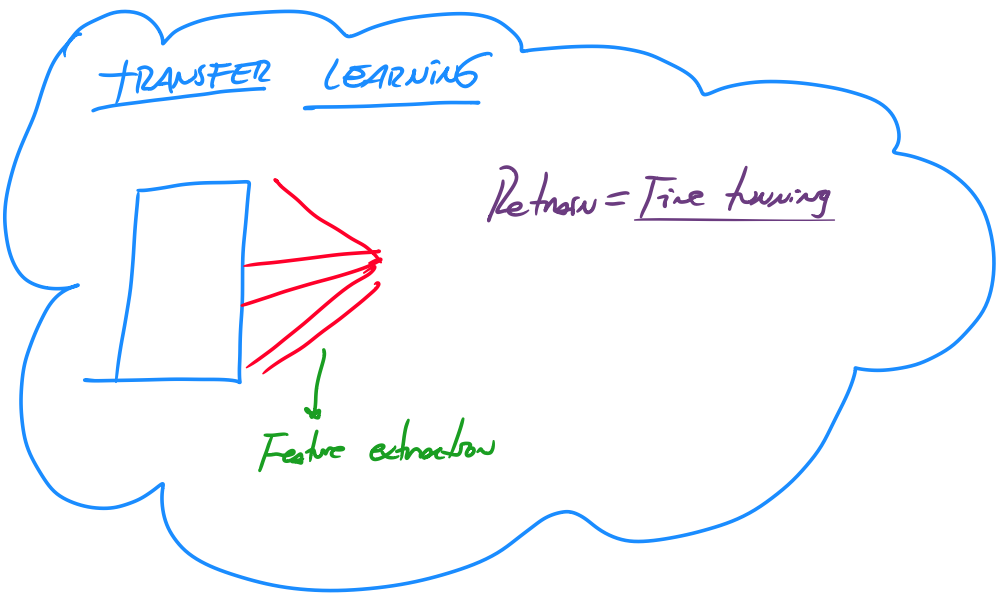
Well to: → Named entity recognition
→ Question answering

there are two existing strategies for applying pre-trained language representations.

1) Feature based: Using the pre-trained representations as additional features.

2) Fine-tuning: Introduces minimal task-specific parameters and is trained on the downstream tasks by simple fine tuning all pretrained parameters.

Its very important consider context from both sides.



BERT: two steps:

1) Pre-training

2) Fine-tuning