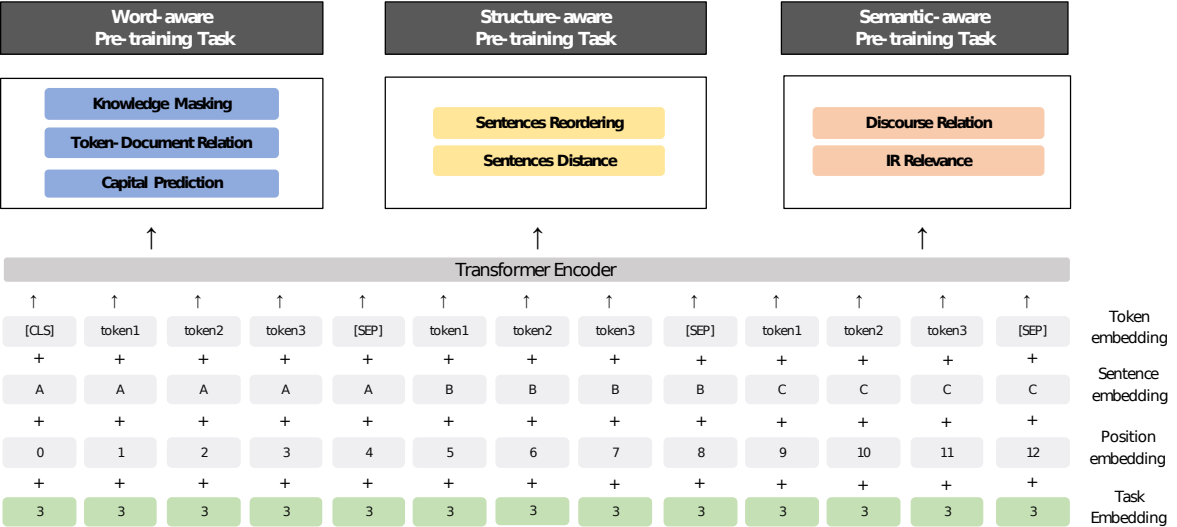
Word-aware Pre-training Task
Structure-aware Pre-training Task
Semantic-aware Pre-training Task
Knowledge Masking
Token-Document Relation
Capital Prediction
Sentences Reordering
Sentences Distance
Discourse Relation
IR Relevance
Transformer Encoder
[CLS] token1 token2 token3 [SEP] token1 token2 token3 [SEP] token1 token2 token3 [SEP]
A A A A A B B B B C C C C
0 1 2 3 4 5 6 7 8 9 10 11 12
3 3 3 3 3 3 3 3 3 3 3 3 3
Token embedding
Sentence embedding
Position embedding
Task Embedding