자연어처리 기말시험

- 1. RNN과 LSTM에 대해 다음 물음에 답하라.
 - a. Vanilla RNN과 비교했을 때 LSTM이 가지는 장-단점을 설명하라.
 - b. Encoder-decoder 방식의 기계번역 시스템에 적용했을 때 LSTM 시스템이 보이는 한계를 설명하라.
- 2. 다음은 Keras에서 text processing에 사용되는 함수들이다. 이 함수들의 기능과 각 함수 의 입-출력 데이터가 어떻게 표현되는지 설명하라.
 - fit on texts()
 - texts_to_sequences()
 - pad_sequences()
- 3. Skip-gram 훈련 방식에 대해 다음 물음에 답하라.
 - a. Skip-gram에 의해 얻고자 하는 출력 데이터 내용을 설명하라. 이 결과의 장점은 무 엇인가?
 - b. 다음 문장이 훈련 과정에서 어떤 절차로 입력되는지 설명하라.
 - "One hundred balls are tossed into fifty boxes"
- 4. 영한번역시스템에서 Sentence Piece 방식을 사용하려면 어떤 절차를 거쳐야 하는지 설명하라. 이 방식을 사용했을 때와 사용하지 않았을 때의 차이점을 설명하라.
- 5. a. 500개의 셀로 구성되는 vanilla RNN 레이어를 만든다고 한다. 이 레이어의 파라미터 의 개수가 어떻게 되는지 설명하라.
 - b. 500개의 셀로 구성되는 LSTM 레이어를 만든다고 한다. 이 레이어의 파라미터의 개수가 어떻게 되는지 설명하라.
- 6. seq2seq 모델에서의 attention 기능에 대해 다음 물음에 답하라.
 - a. Attention이 계산되는 수식과 계산되는 빈도수를 설명하라.
 - b. Attention 기능을 도입했을 때 기계번역 성능이 향상되는 이유를 설명하라.
 - c. Attention 기능이 훈련 데이터 양과 훈련시간에 미치는 영향을 설명하라.

- 7. Transformer에 대해 다음 물음에 답하라.
 - a. Transformer에서의 self-attention과 초기 Bahdanau attention의 차이점을 설명하라. 두 방식의 장-단점을 설명하라.
 - b. Transformer에서 사용하는 residual network의 의미를 설명하라. 이 구조를 사용하는 이유를 설명하라.
- 8. BERT에 대해 다음 물음에 답하라.
 - a. 한국어 처리를 위해 BERT를 학습시키는 과정을 설명하라.
 - b. 질의 응답을 위해 한국어 BERT 시스템을 활용한다고 했을 때 시스템을 훈련시키는 과정을 설명하라.