

○○아카데미(훈련기관명)

## 팀 프로젝트 명 (주제)

TEAM 1조 어벤져스 (팀명)

김○○, 박○○, 이○○, 최○○, 정○○

※ 양식은 예시로 자유롭게 변경 가능하나, 목차 안에 구성된 내용은 포함되도록 작성(해당 템플릿 활용 지양)

## 00 작성요령

---

- ▶ 본 훈련생 포트폴리오 양식은 대표 프로젝트의 팀 별로 각각 작성하여 제출
- ▶ 프로젝트 수행 과정 및 결과에 대해서는 제공된 목차 및 세부 항목별 작성요령을 참조하여 작성하되, 프로젝트 특성에 따라 기본적인 구성을 유지한 상태에서 제공 양식을 보완하거나 추가하여 작성할 수 있음
- ▶ [별첨2] 팀별 프로젝트 수행 결과 작성 양식을 대체할 수 있는 훈련생 포트폴리오, 결과보고서 등 다른 문서가 있는 경우 대체하여 제출 가능
- ▶ 훈련생 포트폴리오에 작성한 내용은 관련 증빙자료를 제출해야 함
- ▶ 작성 예시 및 작성요령 등은 모두 삭제 후 제출

# 목차

01. 프로젝트 개요

02. 프로젝트 팀 구성 및 역할

03. 프로젝트 수행 절차 및 방법

04. 프로젝트 수행 결과

05. 자체 평가 의견

# 01 프로젝트 개요

---

▶ [프로젝트 개요]는 아래와 같은 내용 등으로 구성하여 작성한다.

- 프로젝트 주제 및 선정 배경(기획의도 등)
- 프로젝트 개요(프로젝트 구현 내용, 컨셉, 훈련내용과의 관련성 등)
- 활용 장비 및 재료(개발 환경 등)
- 프로젝트 구조
- 기대 효과

## 02 프로젝트 팀 구성 및 역할

▶ [프로젝트 팀 구성 및 역할]은 프로젝트를 기본 단위로 작성하며 팀원의 수에 따라 칸을 추가/삭제할 수 있다.

- 담당 업무 : 훈련생 별로 해당 프로젝트를 진행하면서 주도적으로 참여한 부분을 중심으로 작성

훈련생	역할	담당 업무
김○○	팀장	▶ 데이터 정제 및 정규화 ▶ 모바일 서비스 테스트
박○○	팀원	▶ 모바일 플랫폼 구현 ▶ 외부 데이터 수집
정○○	팀원	▶ 서비스 시스템 설계 ▶ 텍스트 마이닝

## 03 프로젝트 수행 절차 및 방법

▶ [프로젝트 수행 절차 및 방법]은 프로젝트의 사전 기획과 프로젝트 수행 및 완료 과정으로 나누어서 작성한다.

- 프로젝트 수행 절차를 도식화하여 제시하거나, 더 효과적으로 전달하는 방법 등이 있다면 수정하여 작성 가능
- 기획 단계에서 도출된 주제와 아이디어를 기반으로 실제 프로젝트를 수행한 세부적인 기간과 활동 내용 작성

구분	기간	활동	비고
사전 기획	▶ 0/0(월)~0/0(금)	▶ 프로젝트 기획 및 주제 선정 ▶ 기획안 작성	▶ 아이디어 선정
데이터 수집	▶ 0/0(월)~0/0(금)	▶ 필요 데이터 및 수집 절차 정의 ▶ 외부 데이터 수집	▶ 협약기업 데이터 협조
데이터 전처리	▶ 0/0(월)~0/0(금)	▶ 데이터 정제 및 정규화	
모델링	▶ 0/0(월)~0/0(금)	▶ 모형 구현	▶ 팀별 중간보고 실시
서비스 구축	▶ 0/0(월)~0/0(금)	▶ 모바일 서비스 시스템 설계 ▶ 모바일 플랫폼 구현	▶ 최적화, 오류 수정
총 개발기간	▶ 0/0(월)~0/0(금)(총7주)	-	-

## 04 프로젝트 수행 결과

---

- ▶ [프로젝트 수행 결과]는 프로젝트 결과물이 도출된 과정을 세부적으로 기록
- 예시는 하나의 사례로 간단하게 제시한 것이므로 프로젝트의 성격에 따라 보다 자세하게 기록하며, 결과를 서술하는 과정에서는 활용된 기술(구현 방법), 핵심기능, 검증 결과\* 등을 상세히 기재한다.  
\* 예. 빅데이터 직종의 경우 정확도 등
  - 프로젝트의 결과는 그 과정이 잘 드러날 수 있도록 가공 과정부터 활용까지 전체적인 프로세스를 확인할 수 있도록 단계별로 작성  
\* 첨부 자료 예시: 결과물 사진, 시연 동영상 등 프로젝트의 우수성이 드러날 수 있는 자료

## 프로젝트 수행 결과

## 결과 제시 ① 탐색적 분석 및 전처리

## ▶ 학습 데이터 소개 (Train/dev set)

- LG CNS KORQUAD 질의응답 형식
  - Context : 10,645개
  - QA 쌍 : 66,181개

- Tokenizing : Okt
- Regular Expression : 불용어가 많아 필수 한글, 영어, 숫자만 추출
- Embedding : 단어 임베딩(Glove) – 단어 사이 문맥상 유사성 이해
- Vocabulary

```
insert paragraph idx (1~28): 1
insert qa idx (1~2): 1
```

Title: 활동저위

paragraph(1) context:

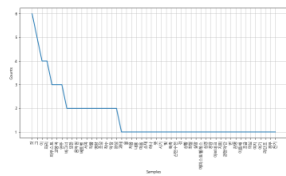
활동 전위 (action potential, 活動電位)는 근육·신경 등 흥분성 세포의 흥분에 의한 세포막의 일시적인 전위변화를 가리키며 동적전위이다. 세포막에 존재하는 나트륨 칼륨 등의 여러 이온 펌프의 활동에 의해 세포 안팎의 이온 조성은 차이가 있는데, 이러한 이온 조성차 90 mV의 음전위(정지전위)를 나타낸다. 신경·근육 등의 흥분성 세포가 흥분하면 세포막 안팎의 극성(極性)이 바뀌어 세포내가 30~40 mV 된다. 이 전위변화는 몇 밀리초(ms) 정도의 빠른 시간 안에 회복되므로 스파이크 전위(spike potential)라고도 하며, 회복기에 보이는 ter potential)과 구별된다. 이 전위변화에 따라 국소전류가 발생하며 1~100m/s의 속도로 흥분이 전달된다. 이 전위변화를 올리우스(na<sup>+</sup>)를 탈분극현상으로 설명한다.

1st qa:

question: 근육, 신경 등 흥분성 세포의 흥분에 의한 세포막의 일시적인 전위변화를 가리키며 동작전위라고도 하는 것은 무엇인가?

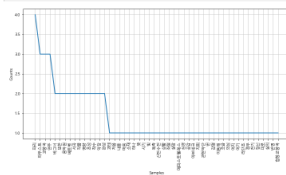
answers: 활동 전위

answers start 0 , end 5



```
In [37]: stop_words = ['것', '그', '이', '을', '것', '']
ko = [each_word for each_word in ko if each_word not in stop_words]
ko
```

```
In [38]: figo = nltk.Text(token_co, name='Trsain_context_4')
ko = nltk.Text(ko, name='Tcontext4')
ko.vocab()
plt.figure(figsize=(12,6))
ko.plot(50)
nltk.show()
```



```
okt = Okt()
words = okt.morphs(all_words, stem=True)
print(words)
```

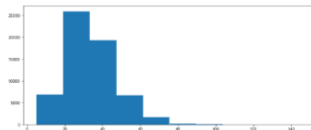
```
# 불용어 제거
stop_words = set(['은', '는', '이', '가', '하', '마', '것', '들', '있', '있', '되', '수', '보', '주', '등',
clean_words = [token for token in words if not token in stop_words]
```

['81',  
',',  
'4',  
',',  
'2010년',  
'9월',  
'1일',  
'2',  
NF']

```
In [45]: plt.figure(figsize=(12,5))
```

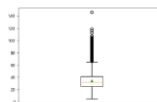
```
plt.hist(quest ion_length)
```

Out[45]: (array([6.6340e+03, 2.5895e+04, 1.9221e+04, 6.6360e+03, 1.6340e+03,  
1.8600e+02, 3.7000e+01, 9.0000e+00, 1.0000e+00, 2.0000e+00]),  
array([ 5., 19.1, 39.2, 47.3, 61.4, 75.5, 89.6, 103.7, 117.8,  
131.9, 146. ]),  
on a list of 10 Patch objects)



```
in [46]: plt.boxplot(question_length)
```

```
Out [46]: {'shiskers': [matplotlib.lines.Line2D at 0x15e793d78f0,
matplotlib.lines.Line2D at 0x15e793dc300],
'caps': [matplotlib.lines.Line2D at 0x15e793dc660,
matplotlib.lines.Line2D at 0x15e793dc960],
'boxes': [matplotlib.lines.Line2D at 0x15e793d400,
matplotlib.lines.Line2D at 0x15e793dc0f0],
'handles': [matplotlib.lines.Line2D at 0x15e793d3900],
'wheels': [matplotlib.lines.Line2D at 0x15e793d0000]}
```



```
In [39]: data = ko.vocab(), most_common(100)
```

[illegible]

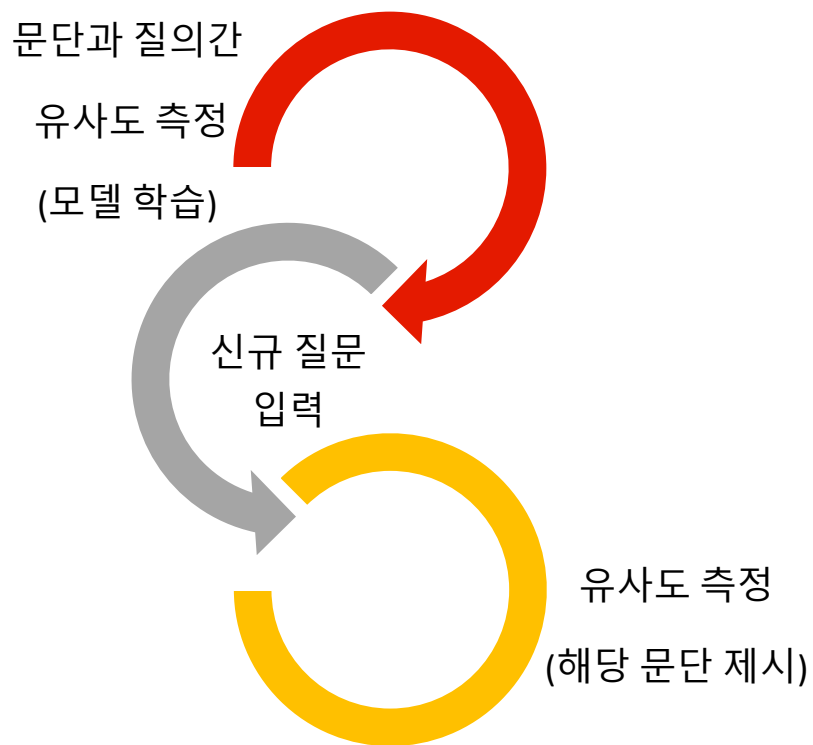


## 04 프로젝트 수행 결과

결과 제시 ② 모델 개요

### ▶ LSTM(Long short-term memory)

- 피드백 루프를 순환하면서 주어진 입력에 관한 신경망 출력을 방지하기 위해 고안된 순환 신경망  
(RNN: Recurrent Neural Network)

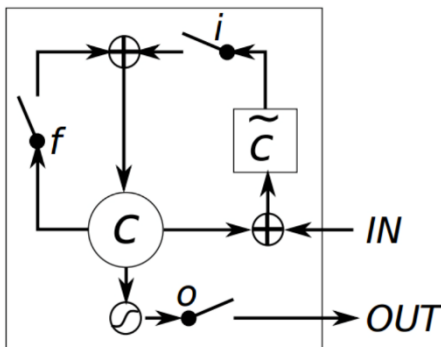


# 04 프로젝트 수행 결과

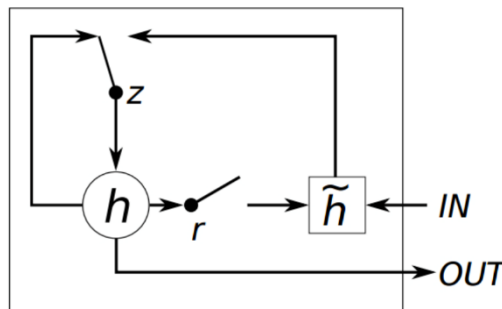
## 결과 제시 ③ 모델 선정 및 분석

### ▶ LSTM(Long short-term memory)

- 2Layer LSTM : 문단, 질문
- 코사인 유사도(문서)



(a) Long Short-Term Memory



(b) Gated Recurrent Unit

```
def lstm_model(embedding_size, vocab_size):
    context = layers.Input(shape=(None,), dtype='int32', name='context')
    question = layers.Input(shape=(None,), dtype='int32', name='question')

    embedding = layers.Embedding(
        mask_zero=True,
        input_dim=vocab_size,
        output_dim=embedding_size,
        # weights=[w2v_weights],
        # trainable=False
    )

    # lstm_1 = layers.LSTM(units=512, return_sequences=True)
    # lstm_2 = layers.LSTM(units=512, return_sequences=True)
    # lstm_3 = layers.LSTM(units=512, return_sequences=False)

    # sum_a = lstm_3(lstm_2(lstm_1(embedding(context))))
    # sum_b = lstm_3(lstm_2(lstm_1(embedding(question))))

    lstm_1 = layers.LSTM(units=512, return_sequences=True)
    lstm_2 = layers.LSTM(units=512, return_sequences=False)

    sum_a = lstm_2(lstm_1(embedding(context)))
    sum_b = lstm_2(lstm_1(embedding(question)))

    sim = layers.dot([sum_a, sum_b], axes=1, normalize=True)
    sim = layers.Activation(activation='sigmoid')(sim)
    sim_model = models.Model(
        inputs=[context, question],
        outputs=[sim],
    )
    # sim_model.compile(loss='binary_crossentropy', optimizer='Adam')
    sim_model.compile(loss='binary_crossentropy', optimizer='rmsprop')

    embedding_model = models.Model(
        inputs=[context],
        outputs=[sum_a]
    )
    return sim_model, embedding_model
```

```
lstm, lstm_embedding = lstm_model(embedding_size=EMBEDDING_SIZE, vocab_size=VOCAB_SIZE)
lstm.summary()
lstm.fit_generator(
    data_generator(batch_size=128),
```

# 04 프로젝트 수행 결과

## 결과 제시 ④ 모델 평가 및 개선

### ▶ LSTM(Long short-term memory)

- 3Layer LSTM 으로 변경

- 옵티마이저 조정

: Adam -> 'rmsprop'로 변경

>> 학습속도 및 유사도 개선

	dist	question	result
0	0.873321	바그너는 파우스트를 읽고 감동받아서 무엇을 했나요	1839년 바그너는 괴테의 파우스트를 처음 읽고 그 내용에 마음이 끌려 이를 소재로
1	0.889332	바그너는 교향곡 작곡을 언제 그만뒀는가	1839년 바그너는 괴테의 파우스트를 처음 읽고 그 내용에 마음이 끌려 이를 소재로
2	0.897146	바그너가 파우스트 서곡을 쓸 때 가장 영향을 받았던 곡은	1839년 바그너는 괴테의 파우스트를 처음 읽고 그 내용에 마음이 끌려 이를 소재로
3	0.928316	파우스트 서곡 1악장을 부활시킨 것은 누구인가	한편 1840년부터 바그너와 알고 지내던 리스트가 잊혀져 있던 1악장을 부활시켜 1852년에 바이마르에서 연주했다. 이것을 계기로 바그너도 이 작품에 다시 관심을 갖게 되었
4	0.934321	리스트가 바그너를 처음 만났던 연도는	한편 1840년부터 바그너와 알고 지내던 리스트가 잊혀져 있던 1악장을 부활시켜 1852년에 바이마르에서 연주했다. 이것을 계기로 바그너도 이 작품에 다시 관심을 갖게 되었

```
lstm, lstm_embedding = lstm_model(embedding_size=EMBEDDING_SIZE, vocab_size=VOCAB_SIZE)
lstm.summary()
lstm.fit_generator(
    data_generator(batch_size=128),
    epochs=10,
    steps_per_epoch=100,
)
```

Layer (type)	Output Shape	Param #	Connected to
context (InputLayer)	(None, None)	0	
question (InputLayer)	(None, None)	0	
embedding_1 (Embedding)	(None, None, 100)	25000000	context [0] [0] question [0] [0]
lstm_1 (LSTM)	(None, None, 512)	1255424	embedding_1 [0] [0] embedding_1 [1] [0]
lstm_2 (LSTM)	(None, None, 512)	2099200	lstm_1 [0] [0] lstm_1 [1] [0]
lstm_3 (LSTM)	(None, 512)	2099200	lstm_2 [0] [0] lstm_2 [1] [0]
dot_2 (Dot)	(None, 1)	0	lstm_3 [0] [0] lstm_3 [1] [0]

Total params: 30,453,824  
Trainable params: 30,453,824  
Non-trainable params: 0

Epoch 1/10  
95/100 [=====>...] - ETA: 31:48 - loss: 7.82 - ETA: 27:36 - loss: 4.57 - ETA: 26:14 - loss: 3.31 - ETA: 25:37 - loss: 2.76 - ETA: 25:08 - loss: 2.40 - ETA: 24:53 - loss: 2.16 - ETA: 25:33 - loss: 1.98 - ETA: 26:40 - loss: 1.83 - ETA: 27:17 - loss: 1.70 - ET  
A: 27:02 - loss: 1.65 - ETA: 26:30 - loss: 1.56 - ETA: 25:49 - loss: 1.49 - ETA: 25:17 - loss: 1.43 - ETA: 25:14 - loss: 1.38 - ETA: 25:26  
- loss: 1.34 - ETA: 26:08 - loss: 1.30 - ETA: 26:08 - loss: 1.26 - ETA: 25:47 - loss: 1.23 - ETA: 25:22 - loss: 1.20 - ETA: 24:51 - loss:  
1.18 - ETA: 24:31 - loss: 1.15 - ETA: 24:06 - loss: 1.13 - ETA: 23:52 - loss: 1.11 - ETA: 23:57 - loss: 1.10 - ETA: 24:00 - loss: 1.08 - ET  
A: 23:42 - loss: 1.07 - ETA: 23:18 - loss: 1.05 - ETA: 22:52 - loss: 1.04 - ETA: 22:29 - loss: 1.03 - ETA: 22:07 - loss: 1.02 - ETA: 21:50  
- loss: 1.01 - ETA: 21:42 - loss: 1.00 - ETA: 21:35 - loss: 0.99 - ETA: 21:17 - loss: 0.98 - ETA: 20:58 - loss: 0.97 - ETA: 20:36 - loss:  
0.97 - ETA: 20:16 - loss: 0.96 - ETA: 19:58 - loss: 0.95 - ETA: 19:45 - loss: 0.95 - ETA: 19:38 - loss: 0.94 - ETA: 19:24 - loss: 0.94 - ET  
A: 19:05 - loss: 0.93 - ETA: 18:46 - loss: 0.93 - ETA: 18:25 - loss: 0.92 - ETA: 18:04 - loss: 0.92 - ETA: 17:46 - loss: 0.91 - ETA: 17:33  
- loss: 0.91 - ETA: 17:21 - loss: 0.90 - ETA: 17:02 - loss: 0.90 - ETA: 16:40 - loss: 0.89 - ETA: 16:17 - loss: 0.89 - ETA: 15:55 - loss:  
0.89 - ETA: 15:34 - loss: 0.88 - ETA: 15:16 - loss: 0.88 - ETA: 15:03 - loss: 0.88 - ETA: 14:46 - loss: 0.87 - ETA: 14:26 - loss: 0.87 - ET

## 04 프로젝트 수행 결과

결과 제시 ⑤ 시연 동영상

---

※ 별도 첨부 가능



## 05 자체 평가 의견

---

▶ [자체 평가 의견]은 프로젝트 결과물에 대한 프로젝트 기획 의도와의 부합 정도 및 실무 활용 가능 정도, 달성도, 완성도 등 훈련기관 또는 훈련생의 자체적인 평가 의견과 느낀 점을 작성한다.

- 개인 또는 우리 팀이 잘한 부분과 아쉬운 점을 작성한다.
  - \* 예. 모델 평가 결과, 정확도가 00.00%로 정확도 향상을 위해 모델 추후 개선 필요
- 프로젝트를 수행하면서 느낀 점이나 경험한 성과에 대하여 기재할 수 있으며, 경력 계획 등과 연관시켜 팀별 공통 의견 또는 개인 의견을 자유롭게 작성한다.