KorQuAD 2.0을 이용한 Q/A 시스템 구현

KorQuAD 2.0

- **KorQuAD 2.0**은 KorQuAD 1.0에서 질문답변 20,000+ 쌍을 포함하여 총 **100,000+** 쌍으로 구성된 한국어 Machine Reading Comprehension 데이터셋
- KorQuAD 1.0과는 다르게 1~2 문단이 아닌 **Wikipedia article 전체**에서 답을 찾아야 함
- **매우 긴** 문서들이 있기 때문에 탐색 시간에 대한 고려가 필요
- 표와 리스트도 포함되어 있기 때문에 **HTML tag**를 통한 문서의 구조 이해도 필요

47,957개의 Wikipedia article, 102,960개의 Q/A pair

KorQuAD 2.0 데이터 형태

```
data['data'][0].keys()
```

```
dict_keys(['context', 'qas', 'title', 'url', 'raw_html'])
```

- 하나의 위키피디아 페이지가 ‘context’, ‘qas’, ‘title’, ‘url’, ‘raw_html’의 keys를 가지는 하나의 딕셔너리로 리스트에 정리되어 있음.
- 하나의 json 파일에 1000개의 딕셔너리, 총 39개의 json 파일

KorQuAD 2.0 데이터 형태

‘title’

```
data['data'][0]['title']
```

```
'예고범'
```

‘url’

```
data['data'][0]['url']
```

```
'https://ko.wikipedia.org/wiki/예고범'
```

KorQuAD 2.0 데이터 형태

‘context’

```
data['data'][0]['context']
```


```
'<!DOCTYPE html>₩n<html>₩n<head>₩n<meta>₩n<title>예고범 - 위키백과, 우리 모두의 백과사전</title>₩n₩n₩n<link>₩n₩n<meta>₩n<link>₩n<meta>₩n<meta>₩n<meta>₩n<meta>₩n<link>₩n<link>₩n<link>₩n<link>₩n<link>₩n<link>₩n<link>₩n<link>₩n<link>₩n₩n</head>₩n<body>₩n<div></div>₩n<div></div>₩n<div>₩n<a></a>₩n<div></div>₩n<div>₩n</div>₩n<h1>예고범</h1>₩n<div>₩n<div>위키백과, 우리 모두의 백과사전.</div>₩n<div></div>₩n<div></div>₩n<a>둘러보기로 가기</a>₩n<a>검색하러 가기</a>₩n<div><div><p>《<b>예고범</b>》(<span><a>일본어</a>: </span><span>予告犯</span>)은, 츠츠이 테츠야의 <a>만화</a> 작품이다.₩n</p>₩n<div><input><div><h2>목차</h2><span><label></label></span></div>₩n<ul>₩n<li><a><span>1</span> <span>개요</span></a></li>₩n<li><a><span>2</span> <span>등장인물</span></a>₩n<ul>₩n<li><a><span>2.1</span> <span>예고범 그룹</span></a></li>₩n<li><a><span>2.2</span> <span>경찰 관계자</span></a></li>₩n<li><a><span>2.3</span> <span>타깃이 된 인물 · 기업 · 단체</span></a></li>₩n<li><a><span>2.4</span> <span>그 외</span></a></li>₩n</ul>₩n</li>₩n<li><a><span>3</span> <span>서지 정보</span></a></li>₩n<li><a><span>4</span> <span>스핀오프 작품</span></a>₩n<ul>₩n<li><a><span>4.1</span> <span>예고범-THE COPYCAT-</span></a></li>₩n<li><a><span>4.2</span> <span>예고범-THE CHASER-</span></a></li>₩n</ul>₩n</li>₩n<li><a><span>5</span> <span>영화</span></a>₩n<ul>₩n<li><a><span>5.1</span> <span>캐스트</span></a></li>₩n<li><a><span>5.2</span> <span>스태프</span></a></li>₩n</ul>₩n</li>₩n<li><a><span>6</span> <span>텔레비전 드라마</span></a>₩n<ul>₩n<li><a><span>6.1</span> <span>캐스트</span></a></li>₩n<li><a><span>6.2</span> <span>스태프</span></a></li>₩n<li><a><span>6.3</span> <span>방송 일자</span></a></li>₩n</ul>₩n</li>₩n<li><a><span>7</span> <span>외부 링크</span></a></li>₩n</ul>₩n</div>₩n<h2><span></span><span>개요</span><span><span>[</span><a>편집</a><span>]</span></span></h2>₩n<p>《점프 카이》 (<a>슈에이샤</a>)에서 <a>2011년</a>부터 <a>2013년</a> 9호까지 연재되었다. 단행본은 전 3권.₩n</p>₩n<h2><span></span><span>등장인물</span><span><span>[</span><a>편집</a><span>]</span></span></h2>₩n<h3><span></span><span>예고범 그룹</span><span><span>[</span><a>편집</a><span>]</span></span></h3>₩n<dl><dt>게이츠/오쿠다 히로아키</dt>₩n<dd>예고범 그룹 〈신분시〉의 주범격.</dd>₩n<dt>칸사이/카사이 토모히코</dt>₩n<dd><a>오사카</a> 출신.</dd>₩n<dt>메타보/테라하라 신이치</dt>₩n<dd><a>후쿠오카</a> 출신.</dd>₩n<dt>노비타/키무라 코이치</dt>₩n<dd><a>미야기</a> 출신.</dd></dl>₩n<h3><span></span><span>경찰 관계자</span><span><span>[</span><a>편집</a><span>]</span></span></h3>₩n<dl><dt>요시노 에리카</dt>₩n<dd><a>경시청</a> 사이버 범죄 대책과 반장
```

KorQuAD 2.0 데이터 형태

‘raw_html’

```
data['data'][0]['raw_html']
```


```
'<!DOCTYPE html>₩n<html class="client-nojs" dir="ltr" lang="ko">₩n<head>₩n<meta charset="utf-8"/>₩n<title>예고범 - 위키백과, 우리 모
두의 백과사전</title>₩n<script>document.documentElement.className=document.documentElement.className.replace(/(^|₩₩s)client-nojs(₩₩s
|$)/,"$1client-js$2");RLCONF={"wgCanonicalNamespace":"","wgCanonicalSpecialPageName":!1,"wgNamespaceNumber":0,"wgPageName":"예고
범","wgTitle":"예고범","wgCurRevisionId":21882503,"wgRevisionId":21882503,"wgArticleId":1142341,"wgIsArticle":!0,"wgIsRedirect":!
1,"wgAction":"view","wgUserName":null,"wgUserGroups":["*"],"wgCategories":["일본어 표기를 포함한 문서","일본의 만화","추리 만화","범
죄를 소재로 한 작품","테러를 소재로 한 작품","일본의 영화 작품","일본의 범죄 영화","테러리즘을 소재로 한 영화","만화를 바탕으로 한
영화"],"wgBreakFrames":!1,"wgPageContentLanguage":"ko","wgPageContentModel":"wikitext","wgSeparatorTransformTable":["",""],"wgDigitT
ransformTable":["",""],"wgDefaultDateFormat":"ko","wgMonthNames":["","1월","2월","3월","4월","5월","6월","7월","8월","9월","10월","1
1월","12월"],₩n"wgMonthNamesShort":["","1","2","3","4","5","6","7","8","9","10","11","12"],"wgRelevantPageName":"예고범","wgRelevant
ArticleId":1142341,"wgRequestId":"XRp8LgpAICMAABjkYugAAADP","wgCSPNonce":!1,"wgIsProbablyEditable":!0,"wgRelevantPageIsProbablyEdita
ble":!0,"wgRestrictionEdit":[],"wgRestrictionMove":[],"wgMediaViewerOnClick":!0,"wgMediaViewerEnabledByDefault":!0,"wgPopupsReferenc
ePreviews":!1,"wgPopupsConflictsWithNavPopupGadget":!1,"wgVisualEditor":{"pageLanguageCode":"ko","pageLanguageDir":"ltr","pageVarian
tFallbacks":"ko"},"wgMFDisplayWikibaseDescriptions":{"search":!0,"nearby":!0,"watchlist":!0,"tagline":!0},"wgWMESchemaEditAttemptSte
pOversample":!1,"wgPoweredByHHVM":!0,"wgULSCurrentAutonym":"한국어","wgNoticeProject":"wikipedia","wgWikibaseItemId":"Q3407535","wgC
entralAuthMobileDomain":!1,"wgEditSubmitButtonLabelPublish":!0};RLSTATE={"ext.globalCssJs.user.styles":"ready","ext.globalCssJs.sit
e.styles":"ready","site.styles":"ready","noscript":₩n"ready","user.styles":"ready","ext.globalCssJs.user":"ready","ext.globalCssJs.s
ite":"ready","user":"ready","user.options":"loading","user.tokens":"loading","mediawiki.legacy.shared":"ready","mediawiki.legacy.com
monPrint":"ready","mediawiki.toc.styles":"ready","wikibase.client.init":"ready","ext.visualEditor.desktopArticleTarget.noscript":"re
ady","ext.uls.interlanguage":"ready","ext.wikimediaBadges":"ready","ext.3d.styles":"ready","mediawiki.skinning.interface":"ready","s
kins.vector.styles":"ready"};RLPAGEMODULES=["site","mediawiki.page.startup","mediawiki.page.ready","mediawiki.toc","mediawiki.search
```

KorQuAD 2.0 데이터 형태

‘qas’

```
data['data'][0]['qas']
```

```
[{'question': '드라마 예고범의 감독은 누구일까?',
  'id': '8089',
  'answer': {'text': '나카무라 요시히로, 히라바야시 카츠토시, 사와다 메구미',
   'html_answer_text': '나카무라 요시히로, 히라바야시 카츠토시, 사와다 메구미',
   'answer_start': 6302,
   'html_answer_start': 21842}}]
```

- 1개 이상의 질문-답변 딕셔너리들이 리스트 형태로 저장
- ‘answer_start’는 ‘context’에서 답변 시작 인덱스
- ‘html_answer_start’는 ‘raw_html’에서의 답변 시작 인덱스

Preprocessing

Html tag

Html outline

```
</head>
<body>
<div></div>
<div></div>
<div>
<a></a>
<div></div>
<div>
</div>
<h1>이윤성 (1944년)</h1>
<div>
<div>위키백과, 우리 모두의 백과사전.</div>
<div></div>
<div></div>
<a>둘러보기로 가기</a>
<a>검색하러 가기</a>
<div><div><table><tbody><tr><th colspan="2"><div><div></div><span>이윤성 <br/> 李允盛</span></div></th></tr><tr><th colspan="2">기본
<td>
<a>대한민국</a></td></tr><tr><th>출생</th>
<td>
1944년 10월 2일<span>(<span>1944-10-02</span>)</span> (74세)<br/><a>일제 강점기</a> <a>함경북도</a> <a>청진부</a></td></tr><tr><th>.
<td>
<a>대한민국</a> <a>인천광역시</a></td></tr><tr><th>학력</th>
<td>
<a>한국외국어대학교</a> 학사</td></tr><tr><th>직업</th>
<td>
<a>정치가</a>, 방송인</td></tr><tr><th>경력</th>
<td>
前 인천학술진흥재단 이사장</td></tr>
<tr><th>군복무</th>
<td>
```

1. 이윤성 (1944년) <body><h1>
    1. 목차 <h2>
    2. 경력[편집] <h2>
    3. 학력[편집] <h2>
        1. 비학위 수료[편집] <h3>
    4. 이력[편집] <h2>
    5. 주요 행적 및 논란[편집] <h2>
        1. 언론인으로서의 행적[편집] <h3>
        2. 천안함 사태가 다행이라는 '망언'에 대한 논란[편집] <h3>
        3. 공무원 연봉 부풀리기 논란[편집] <h3>
            1. 사건의 발단[편집] <h4>
            2. 반론[편집] <h4>
            3. 노조의 반응[편집] <h4>
    6. 역대 선거 결과[편집] <h2>
    7. 각주[편집] <h2>
    8. 외부 링크[편집] <h2>
    9. 둘러보기 메뉴 <h2>
        1. 개인 도구 <h3>
        2. 이름공간 <h3>
        3. 변수 <h3>
        4. 보기 <h3>
        5. 더 보기 <h3>
        6. 검색 <h3>
        7. 둘러보기 <h3>
        8. 도구 <h3>
        9. 인쇄/내보내기 <h3>
        10. 다른 언어 <h3>

Tag embedding

Question & Answer 생성 기준

-텍스트<p>, 표<table>, 리스트<ol, ul, dl>에 존재

-소제목(<h2>, <h3>등) 단위의 문서 기준 생성

문서 수집

- 위키 문서 중에서 **page view 상위 문서 15만** 건 + **임의로 선정된 5만** 건의 페이지 HTML 크롤링
- 수집한 문서 중 질문을 생성할 부분으로 **텍스트(<p>), 표(<table>), 리스트(<ul>, <ol>, <dl>)** 추출

질문 - 답변 생성

- 크라우드 소싱을 통해 질문-답변 **80,000+** 쌍 제작
- 작업자는 위키 전체 문서가 아니라 **소제목 단위의 문서**를 보고 질문-답변 생성

Tag embedding

텍스트, 표, 리스트 tags 추출시 이슈(예: 이윤성(1944년))

question

5. 주요 행적 및 논란[편집] <h2>
   1. 언론인으로서의 행적[편집] <h3>
   2. 천안함 사태가 다행이라는 '망언'에 대한 논란[편집] <h3>

answer

천안함 사태가 다행이라는 '망언'에 대한 논란 [편집]

전국 동시 지방 선거를 앞두고 있던 2010년 5월 31일, 이윤성 의원은 여의도 당사에서 열린 수도권 선대위원장 기자간담회에서 인천시장 선거의 판세를 분석했다. 그런데 여기서 그는 "또 하나 반가운 것은 10명의 기초단체장 가운데 옹진군은 무투표로 당선됐고 나머지 9군데도 좀 어렵다고 생각했는데 **다행히 천안함 사태가 바로 인천 앞바다(에서 일어났)다**"며 "그렇기 때문에 다른 계층보다 느끼는 바가 달라 기초단체장도 1~2곳의 경합 지역을 빼놓고는 다 우세 지역으로 궤도에 진입했다"고 발언해 구설수에 올랐다.[6]

천안함 사태가 다행이라는 그의 발언에 대해 이후 각 당으로부터 비난이 쏟아졌다. 민주당 김유정 선대위 대변인은 "자신의 지역구인 인천 앞바다에서 천안함 사태가 발생한 것이 다행이고 행운이었다는 말을 서슴없이 쏟아내는 이 의원은 인천시민을 대표할 자격이 없는 사람"이라며 "망언"을 철회하고 희생자와 유가족, 그리고 국민에 사죄할 것을 촉구했다.[6] 미래연합의 오형석 대변인도 논평에서 "이 의원의 망언을 규탄한다"며 "이번 발언은 국가적 불행인 천안함 사건이 한나라당에는 곧 행복이었음을 고백한 것이나 다름없다. 이번 망언에 대한 책임으로 의원직을 즉각 사퇴하고 국민들에게 사죄하라"고 촉구했다.[7] 자유선진당의 박선영 대변인도"망발도 유분수"라면서 6월 1일 논평을 통해 "천안함 사태가 한나라당 선거 판세에 도움이 됐다는 이유로 '반가웠고', 또 '다행히'라니. 구천을 떠도는 영령들과 국민 모두가 가슴을 치며 통곡할 일"이라고 비판했다.[8][9]

```
<h3><span></span><span>천안함 사태가 다행이라는 '망언'에 대한 논란</span><span>
<span>[</span><a>편집</a><span>]</span></span></h3>
<p>전국 동시 지방 선거를 앞두고 있던 2010년 5월 31일, 이윤성 의원은 여의도 당사에
서 열린 수도권 선대위원장 기자간담회에서 인천시장 선거의 판세를 분석했다. 그런데
여기서 그는 "또 하나 반가운 것은 10명의 기초단체장 가운데 옹진군은 무투표로 당선
됐고 나머지 9군데도 좀 어렵다고 생각했는데<b> 다행히 천안함 사태가 바로 인천 앞바
다(에서 일어났)다</b>" 며 "그렇기 때문에 다른 계층보다 느끼는 바가 달라 기초단체
장도 1~2곳의 경합 지역을 빼놓고는 다 우세 지역으로 궤도에 진입했다" 고 발언해 구
설수에 올랐다.<sup><a>[6]</a></sup>
</p><p>천안함 사태가 다행이라는 그의 발언에 대해 이후 각 당으로부터 비난이 쏟아졌
다. 민주당 김유정 선대위 대변인은 "자신의 지역구인 인천 앞바다에서 천안함 사태가
발생한 것이 다행이고 행운이었다는 말을 서슴없이 쏟아내는 이 의원은 인천시민을 대
표할 자격이 없는 사람" 이라며 "망언"을 철회하고 희생자와 유가족, 그리고 국민에 사
죄할 것을 촉구했다.<sup><a>[6]</a></sup> 미래연합의 오형석 대변인도 논평에서 "이
의원의 망언을 규탄한다" 며 "이번 발언은 국가적 불행인 천안함 사건이 한나라당에는
곧 행복이었음을 고백한 것이나 다름없다. 이번 망언에 대한 책임으로 의원직을 즉각
사퇴하고 국민들에게 사죄하라" 고 촉구했다.<sup><a>[7]</a></sup> 자유선진당의 박선
영 대변인도 "망발도 유분수" 라면서 6월 1일 논평을 통해 "천안함 사태가 한나라당
선거 판세에 도움이 됐다는 이유로 '반가웠고', 또 '다행히' 라니. 구천을 떠도는
영령들과 국민 모두가 가슴을 치며 통곡할 일" 이라고 비판했다.<sup><a>[8]</a></sup>
<sup><a>[9]</a></sup>
</p>
```

Tag embedding

텍스트, 표, 리스트 tags 추출시 이슈

- 소제목(<h2>, <h3>등) 단위의 문서 기준 생성 → tag parsing 후 소제목과 연결의 어려움

예) question 생성된 곳 <h2> answer 생성된 곳 <p>

→ tag별 추출시 <h2>등 과의 연결 불가

→ <h2>,<h3>와 <p>,<table>,<list>를 sequential하게 연결 필요

Approach

Approach

1. HTML 문서로 되어 있는 context 데이터를 공백 없이 붙어 있는 태그를 분리하고, 그 이후 문장의 각 단어와 태그를 공백을 기준으로 분리하여 어절 단위로 tokenize

1. Bi-LSTM 모델을 사용하여 predict 단계시 기존 정답의 start, end index를 구해 MSE loss를 통해 regression으로 학습

Html 파싱 & 토큰 추출

 시장 재임 기간 시에서 추진하는 사업이 개인의 업적인 것처럼 하여 홍보 동영상을 수차례 개인 SNS에 게재한 혐의(공직선거법 위반)로 기소되었다. 2019년 2월 14일 춘천지법 강릉지원에서 열린 1심에서 공직선거법 위반 혐의에 대해 벌금 70만 원을 선고받았다. 2019년 5월 29일 서울고법 춘천재판부에서 열린 2심에서 공직선거법 위반 혐의에 대해 무죄를 선고받았다.

- BeautifulSoup html.parser 사용하여 html tag와 text 분리
- 어절 단위를 token으로 처리, split() 하여 사용
- 각 tag(p, table, list)마다 포함 text 시작과 끝 지점 index화

- 예) 심규언

  tag와 text를 space로 구분

  tag 뒤 text 시작 부분과 끝 부분 index화

```
# <p> tag
p_tagging = [len(context)]
pstart = [i+1 for i, j in enumerate(context) if j == '<p>']
pend = [i-1 for i, j in enumerate(context) if j == '</p>']

for s, e in zip(pstart, pend):
  p_tagging.append((s,e))

p_tagging
```

```
[1665, (269, 291), (647, 693), (985, 984), (1026, 1025)]
```

```
context[647]
```

```
'시장'
```

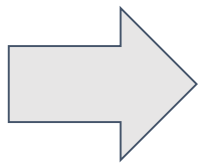
```
context[693]
```

```
'선고받았다.'
```

Answer end index 음절 단위 추출

```
data[1]['qas'][0]
```

```
{'answer': {'answer_end': 3881,
  'answer_start': 3873,
  'html_answer_start': 16093,
  'html_answer_text': '20,890 표',
  'text': '20,890 표'},
 'id': '36615',
 'question': '심규언은 17대 지방 선거에서 몇 표를 득표하였는가?'}
```



```
data[1]['context'][3873:3881]
```

```
'20,890 표'
```

```
class BaseClassifier(nn.Module):

    def __init__(self, num_context, num_query, embed_dim, hid_dim, fc_hid_dim = 300, context_maxlen = 10000,
                 dropout=0.1, num_layers=1, batch_first=True, bidirectional=True):

        super(BaseClassifier, self).__init__()

        self.context_len = context_maxlen
        self.concat_dim = hid_dim * 4 if bidirectional else hid_dim * 2
        self.context_embed = nn.Embedding(num_context, embed_dim)
        self.query_embed = nn.Embedding(num_query, embed_dim)

        self.p_emb = nn.Embedding(2, embed_dim)
        self.table_emb = nn.Embedding(2, embed_dim)
        self.list_emb = nn.Embedding(2, embed_dim)

        self.lstm_c = nn.LSTM(embed_dim, hid_dim, dropout=dropout, num_layers=num_layers, batch_first=batch_first, bidirectional=bidirectional)
        self.lstm_q = nn.LSTM(embed_dim, hid_dim, dropout=dropout, num_layers=num_layers, batch_first=batch_first, bidirectional=bidirectional)
        self.fc_start = nn.Sequential(nn.Linear(self.concat_dim, fc_hid_dim), nn.Linear(fc_hid_dim, 1))
        self.fc_end = nn.Sequential(nn.Linear(self.concat_dim, fc_hid_dim), nn.Linear(fc_hid_dim, 1))
```

html tag marking 정보를
embedding으로 사용

bi-LSTM

Tag Embeddings

Results

Train 결과

Q: 상속 결격자의 요건은 어떻게 되는가?

정답: 고의로 직계존속, 피상속인, 그의 배우자 또는 상속의 선순위자나 동상속인^{<a>[6]}을 살해하거나 살해하려 한 자
고의로 직계존속, 피상속인과 그 배우자에게 상해를 가하여 사망에 이르게 한 자
사기 또는 강박으로 피상속인의 상속에 관한 유언 또는 유언의 철회를 방해한 자
사기 또는 강박으로 피상속인의 상속에 관한 유언을 하게 한 자
피상속인의 상속에 관한 유언서를 위변조, 파기 또는 은닉한자

A: 고의로 직계존속, 피상속인, 그의 배우자 또는 상속의 선순위자나 동상속인^{<a>[6]}을 살해하거나 살해하려 한 자
고의로 직계존속, 피상속인과 그 배우자에게 상해를 가하여 사망에 이르게 한 자
사기 또는 강박으로 피상속인의 상속에 관한 유언 또는 유언의 철회를 방해한 자
사기 또는 강박으로 피상

Q: 강만수는 어느 대학교를 나왔을까?

정답: 한양대학교
A:

<h2>생애
[<a>편집
]</h2>
<p>성지공고, <a>한양대학교를 나와 국가대표 레프트 공격수로 뛰며 <a>대한민국을 대표하는 강 스파이커로 명성을 떨쳤다.

p와 list tag 안에서 생성된 답은 답 주변으로 비교적 잘 예측하는 것을 보임.

Train 결과

Q: 2009년 와글와글 꼬꼬맘은 어떤 화들을 방송했을까?

정답: <table>
<tbody><tr>
<th>회차</th>
<th>방송일</th>
<th>부제
</th></tr>
<tr>
<th rowspan="2">1화
</th>
<td rowspan="2"><a>11월 18일</td>
<td>꼬르륵 배고픈 병아리
</td></tr>
<tr>
<td>소중한 트로피
...중략...
<th rowspan="2">7화
</th>
<td rowspan="2"><a>12월 30일</td>
<td>음악 가족 돈돈 패밀리
</td></tr>
<tr>
<td>엄마가 되고 싶어
</td></tr>
</tbody></table>

A: 병아리들의 비밀 작전
</td></tr>
<tr>
<td>마당에서의 하루
</td></tr>
<tr>
<th rowspan="2">5화
</th>
<td rowspan="2"><a>12월 16일</td>
<td>사랑해요 꼬옥
</td></tr>
<tr>
<td>비 님 어서 오세요
</td></tr>
<tr>
<th rowspan="2">6화
</th>
<td rowspan="2"><a>12월 23일</td>
<td>방가방가 어린이집
</td></tr>
<tr>
<td>두근두근 보물택배

table tag의 경우 답의 처음 index를 잘 잡아내지 못함.

Difficulties

- 원래는 입력된 context의 음절 단위 길이 만큼 LSTM에서 logit을 구하여 softmax를 적용해 분류 문제로 학습하는 것이 일반적이나, 여기서는 html이 포함된 context의 길이가 너무 길어 regression으로 학습하는 것으로 변경함.

- html 각 head(h2, h3)당 해당되는 tag(p, list, table)의 관계를 표시하는 것의 한계가 있었음.

- Train 시 table tag의 answer가 다른 tag보다 정확하지 않았음. → table 정보에 대한 전처리 필요

Thank you