

SSAFY AI 코스 최종 프로젝트

SSAFY 학사 규칙 QA 챗봇

4팀

김도형 박종원 엄도윤 정재영

목차

1. 주제 선정
2. 데이터
3. 모델링 및 개요
4. 실제 구현 및 배포
5. 기대효과 및 보완할 점

1. 주제 선정

1. 주제 선정

배경

- SSAFY 학사 규칙 검색의 어려움
- 운영사무국의 업무 증가
- 신속한 정보 획득의 어려움

“예비 13기를 위한 학사 규칙 QA 챗봇”

2. 데이터

2. 데이터 데이터 수집



- OCR을 통해 텍스트
- 생성형 AI를 이용해 json

2. 데이터 규정 json 파일

```
{
  "1_출결": {
    "1-1_출결관리_기준": {
      "공가": {
        "주요_내용": [
          {
            "항목": "국가 관련 소집",
            "세부사항": "예비군/민방위 등",
            "기간": "해당 소요일수"
          },
          {
            "항목": "본인 및 배우자 경조사",
            "세부사항": ["결혼", "사망", "출산 등"],
            "기간": "최대 5일"
          },
          {
            "항목": "취업관련 결석",
            "세부사항": ["면접", "취업자격증 취득", "정진대회 참석 등"],
            "기간": "최대 3일/월 허용"
          },
          {
            "항목": "천재지변 등으로 인해 불가피한 교육 불참",
            "세부사항": ["홍수", "태풍", "지진 등 자연재해"]
          }
        ],
      },
      "비고": [
        "출석 인정",
        "교육지원금 지급"
      ]
    }
  }
}
```

적은 양의 DATA



QA DATA 생성

2. 데이터

QA json 파일

```
{
  "qa_dataset": [
    {
      "category": "출결",
      "qa_pairs": [
        {
          "question": "출석 체크는 몇 시까지 해야 하나요?",
          "answer": "교육 시작 시간인 오전 9시까지 출석 체크를 완료해야 합니다. 18:00까지 시스템 출석체크를 완료해야 하며, 18:30분까지 시스템 출석체크가 완료되면 출석체크가 인정됩니다."
        },
        {
          "question": "배우자가 출산하면 몇 일간 휴가를 받을 수 있나요?",
          "answer": "배우자 출산 시 5일간의 경조사 휴가가 주어집니다. 이는 공가로 처리되어 출석이 인정되며, 교육지원금도 지급됩니다."
        },
        {
          "question": "면접 때문에 결석하게 될 것 같은데 어떻게 해야 하나요?",
          "answer": "취업 관련 결석(면접 등)은 공가로 처리되며, 월 최대 3일까지 허용됩니다. 증빙 서류를 제출해야 하며 출석으로 인정됩니다."
        },
        {
          "question": "지각 3번이 결석 하루로 처리되나요?",
          "answer": "네, 지각/조퇴/외출 3회는 1일 결석으로 처리됩니다. 이 경우 출석이 불인정되며 교육지원금도 일할 차감됩니다."
        },
        {
          "question": "조퇴는 어떻게 처리되나요?",
          "answer": "조퇴는 교육종료 시간(18:00) 이전 퇴실을 의미하며, 지각/조퇴/외출 3회시 1일 결석으로 처리됩니다."
        },
        {
          "question": "9시에 출석했으면 몇 시에 조퇴가 가능한가요?",
          "answer": "9시부터 18시 사이 점심시간 1시간을 제외하고 4시간 이상 자리를 비우면 안 되므로 2시에 조퇴가 가능합니다."
        },
        {
          "question": "9시에 출석했는데 1시에 조퇴하면 결석이 되나요?",
          "answer": "점심 시간을 제외하고 5시간 이상 자리를 비우게 되므로 이는 조퇴가 아닌 결석에 해당됩니다."
        }
      ]
    }
  ]
}
```

3. 모델링 및 개요

3. 모델링 및 개요

청크화

```
{
  "1_출결": {
    "1-1_출결관리_기준": {
      "공가": {
        "주요_내용": [
          {
            "항목": "국가 관련 소집",
            "세부사항": "예비군/민방위 등",
            "기간": "해당 소요일수"
          },
          {
            "항목": "본인 및 배우자 경조사",
            "세부사항": ["결혼", "사망", "출산 등"],
            "기간": "최대 5일"
          },
          {
            "항목": "취업관련 결석",
            "세부사항": ["면접", "취업자격증 취득", "정진대회 참석 등"],
            "기간": "최대 3일/월 허용"
          },
          {
            "항목": "천재지변 등으로 인해 불가피한 교육 불참",
            "세부사항": ["홍수", "태풍", "지진 등 자연재해"]
          }
        ],
        "비고": [
          "출석 인정",
          "교육지원금 지급"
        ]
      }
    }
  }
}
```

→
“재귀적 청크화”

Id: “1_출결”
Content: {“1_출결”: {...}}

Id: “1_출결/1-1_출결관리_기준”
Content: {“1-1_출결관리_기준”: {...}}

Id: “1_출결/1-1_출결관리_기준/공가”
Content: {“공가”: {“주요_내용”: [...]}}

3. 모델링 및 개요

청크화

```
{
  "qa_dataset": [
    {
      "category": "출결",
      "qa_pairs": [
        {
          "question": "출석 체크는 몇 시까지 해야 하나요?",
          "answer": "교육 시작 시간인 오전 9시까지 출석 체크를 완료해야 합니다. 18:00까지 시스템 출석체크를 완료해야 하며, 18:30분까지 시스템 출석체크가 완료도",
        },
        {
          "question": "배우자가 출산하면 몇 일간 휴가를 받을 수 있나요?",
          "answer": "배우자 출산 시 5일간의 경조사 휴가가 주어집니다. 이는 공가로 처리되어 출석이 인정되며, 교육지원금도 지급됩니다."
        },
        {
          "question": "면접 때문에 결석하게 될 것 같은데 어떻게 해야 하나요?",
          "answer": "취업 관련 결석(면접 등)은 공가로 처리되며, 월 최대 3일까지 허용됩니다. 중빙 서류를 제출해야 하며 출석으로 인정됩니다."
        },
        {
          "question": "지각 3번이 결석 하루로 처리되나요?",
          "answer": "네, 지각/조퇴/외출 3회는 1일 결석으로 처리됩니다. 이 경우 출석이 불인정되며 교육지원금도 일할 차감됩니다."
        },
        {
          "question": "조퇴는 어떻게 처리되나요?",
          "answer": "조퇴는 교육종료 시간(18:00) 이전 퇴실을 의미하며, 지각/조퇴/외출 3회시 1일 결석으로 처리됩니다."
        },
        {
          "question": "9시에 출석했으면 몇 시에 조퇴가 가능한가요?",
          "answer": "9시부터 18시 사이 점심시간 1시간을 제외하고 4시간 이상 자리를 비우면 안 되므로 2시에 조퇴가 가능합니다."
        },
        {
          "question": "9시에 출석했는데 1시에 조퇴하면 결석이 되나요?",
          "answer": "점심 시간을 제외하고 5시간 이상 자리를 비우게 되므로 이는 조퇴가 아닌 결석에 해당됩니다."
        }
      ]
    }
  ],
}
```



Id: “qa_출결_0”

Content: {

“question”: “출석체크는…”,

“answer”: “8시 59분까지…”,

“category”: “출결”

}

3. 모델링 및 개요

임베딩과 캐싱

- text-embedding-ada-002을 이용한 임베딩
- API 호출은 시간과 비용이 높음
- 임베딩 캐싱 사용

3. 모델링 및 개요

검색 프로세스: 임베딩 기반 초기 검색

Question: “출석 체크는 몇 시까지이고, 지각하면 어떻게 되나요?”



질문 임베딩



“코사인 유사도”

청크 임베딩

상위 10개의 청크 선택

3. 모델링 및 개요

검색 프로세스: Cross-encoder 재순위화

[질문, 청크] 쌍 생성(배치 크기 32)



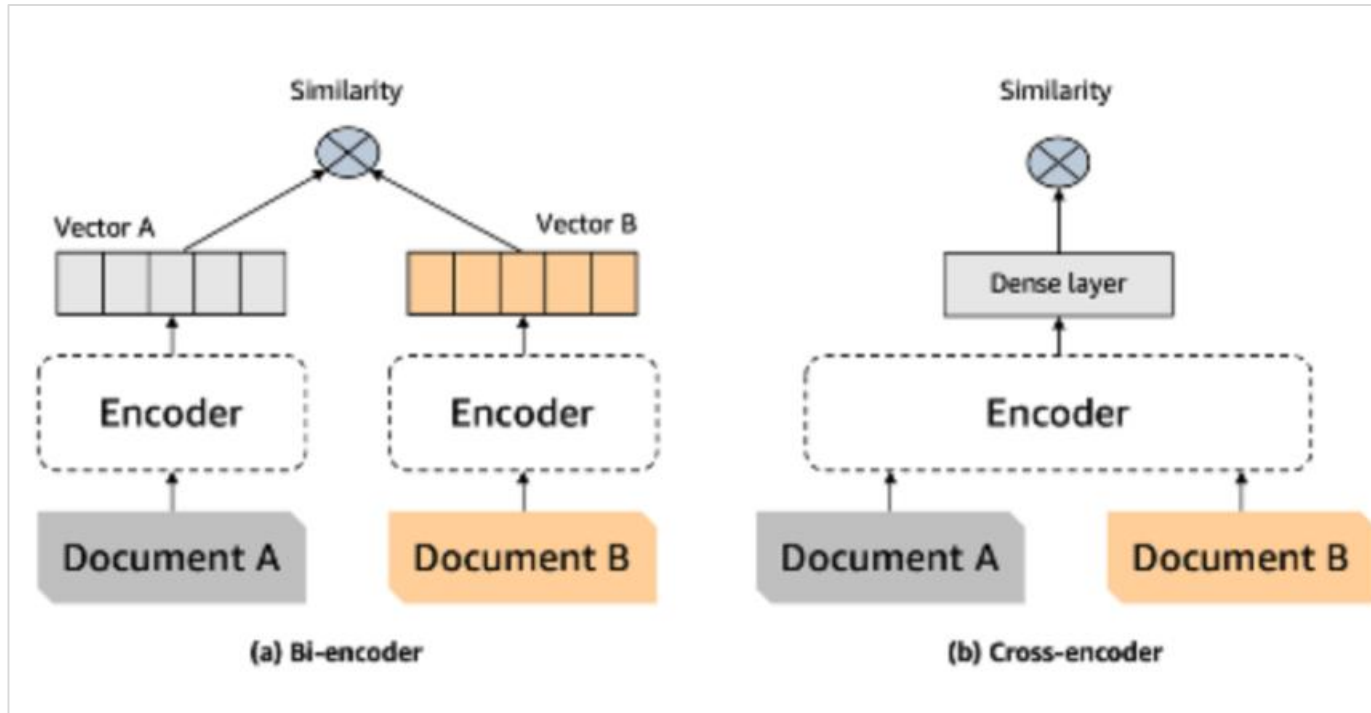
Cross-encoder 평가
문맥적 관련성 계산



최종 상위 3개 선택
Gpt-40-mini 컨텐스로 구성

3. 모델링 및 개요

재순위화 장점



Stage1 (Bi-encoder)

- 각각 인코딩 후 유사도 판단(빠르지만 단순) -> 초기 10개 선택

Stage2 (Cross-encoder)

- 전체 문맥 고려한 정확한 관련성 평가(정확하지만 느림) -> 최종 3개 선택

4. 실제 구현 및 배포

4. 실제 구현 및 배포

<https://ssafychat.kro.kr/>

5. 기대효과 및 보완할 점

5. 기대효과 및 보완할 점

기대효과

- 신속하고 정확한 규정 정보 획득
- 반복적 문의 응대 업무 감소
- 데이터 기반 규정 개선
- 규정 해석의 일관성 확보

5. 기대효과 및 보완할 점

보완할 점

- 데이터 확장 필요성
- 성능 모니터링
- 캐싱 전략 개선



감사합니다.