

## 과제 기술서: 일상 대화 요약(외부 데이터 및 데이터 증강 사용 불가)

### 1. 과제 개요

‘일상 대화 요약’ 과제는 다자간 대화에서 발생하는 여러 주제를 식별하고 각 주제에 대한 요약문을 생성하는 것을 목표로 한다. 일상 대화 요약 과제의 데이터는 2명 이상의 화자와 여러 주제가 등장하는 일상 대화를 다루고 있으며, 모델은 대화 내 등장한 주제의 핵심 내용을 효율적으로 추출하여 요약문을 생성해야 한다. 해당 과제를 통해 모델의 일상 대화 이해 능력과 화자 식별 능력, 더불어 대화 요약 능력을 측정할 수 있으며 주제별 요약 기술은 정보의 분류와 검색, 지식 관리, 의사결정 지원 등에 유용하게 활용된다는 점에서 의의가 큰 과제이다.

항목	내용
대화	<p>화자1: 우리 그때 처음에 나 데리고 가 줬던 초밥집 기억나?</p> <p>화자2: Company-name 상암동에 있는 데?</p> <p>화자1: 어 나 진짜 깜짝 놀랐잖아. 나는</p> <p>화자2: 그때까지 초밥은 그냥 회전 초밥집? 너가 데리고 가 줘서 오마카세 처음 알았어.</p> <p>화자1: 그때 나는 애는 초밥이 뭔지 모르나 보더라는 생각이 들어서 데리고 갔지</p> <p>화자2: 무시하는 거야?</p> <p>...</p>
주제	초밥집
요약문	이 대화에서 화자 1과 2는 초밥집에 대해 이야기했습니다. 1은 2에게 자신을 데리고 가 줬던 초밥집에서 오마카세를 처음 알게 되었고 신세계였다고 이야기했습니다. 그리고 음식은 배 채우려고 먹는 게 아니라 맛있는 음식을 먹어야 한다고 생각했다고 이야기했습니다(...)

표 1 일상 대화 요약 과제의 예시

### 2. 과제 정의

이 과제의 참여자는 여러 화자가 포함된 대화에서 주제 전환을 정확하게 파악하고, 각 주제에 대한 주요 논의 사항을 요약해야 한다. 요약문 생성 시 주제별 결론, 중요한 의견, 주요 사건 등을 포함한다. 주어진 대화는 여러 주제를 포함하는 일상 대화로 이루어져 있다. 요약은 전체 대화에 대해 하는 것이 아니라 주어진 주제에 대해 주제별 요약문을 작성한다. 주제는 키워드(keyword) 형태로 주어지며, 때로는 두 개 이상의 키워드가 동시에 주어질 수 있다.

정량 평가 지표는 루지(ROUGE), 버트스코어(BERTScore), 블루트(BLEURT) 세 지표의 평균이다. 루지-1(ROUGE-1) 점수는 생성된 요약문과 사람이 작성한 참조 요약문(정답 요약문) 사이에서 일치하는 단어의 수를 바탕으로 정밀도(Precision), 재현율(Recall), 조화 평균 점수(F1 score) 세 가지 주요 측정치로 계산된다. 버트스코어(BERTScore), 블루트(BLEURT)는 생성형 문제에 대한 딥러닝 기반의 평가 지표이다.

분류	내용	예시	비고
입력	대화	{"speaker": 1, "utterance": "우리 그때 처음에 나 데리고 가 줬던 초밥집 기억나?"}, {"speaker": 2, "utterance": "Company-name 상암동에 있는 데?"}, {"speaker": 1, "utterance": "어 나 진짜 깜짝 놀랐잖아. 나는"}, {"speaker": 2, "utterance": "그때까지 초밥은 그냥 회전 초밥집? 너가 데리고 가 줘서 오마카세 처음 알았어"}, {"speaker": 1, "utterance": "그때 나는 애는 초밥이 뭔지 모르나 보더라는 생각이 들어서 데리고 갔지"}, {"speaker": 2, "utterance": "무시하는 거야? "}, ...	문자열
	주제	초밥집	문자열
출력	요약	"이 대화에서 화자 1과 2는 초밥집에 대해 이야기했습니다. 1은 2에게 자신을 데리고 가 줬던 초밥집에서 오마카세를 처음 알게 되었고 신세계였다고 이야기했습니다. 그리고 음식은 배 채우려고 먹는 게 아니라 맛있는 음식을 먹어야 한다고 생각했다고 이야기했습니다(...)"	문자열
평가	ROUGE, BERTScore, BLEURT의 평균		

표 2 일상 대화 요약 과제 입력과 출력의 예

### 3. 자료 형식

데이터 세트는 제이슨(Java Script Object Notation) 형식으로 제공되며 각 제이슨(JSON)은 대화, 주제, 요약 정보를 포함한다. 표 3은 데이터 형식의 예시이다. 주어진 훈련 데이터와 시험 데이터는 동일한 제이슨(JSON) 형식으로 제공되며 시험 데이터의 경우에는 각 텍스트에 대한 출력(output) 항목이 없다. 참가 팀은 모델의 출력 결과를 추가하여 제출한다.

※ 훈련 데이터와 제출용 데이터의 형태 및 구성 요소는 동일하다.

항목	내용
훈련용 데이터 형식의 예	<pre>[   {     "id": "nikluge-2024-일상대화의주제별요약-train-000001",     "input": {       "conversation": [         {"speaker": 1, "utterance": "우리 그때 처음에 나 데리고 가 줬던 초밥집 기억나?"},         {"speaker": 2, "utterance": "Company-name 상암동에 있는 데?"},         {"speaker": 1, "utterance": "어 나 진짜 깜짝 놀랐잖아. 나는"},         {"speaker": 2, "utterance": "그때까지 초밥은 그냥 회전 초밥집? 너가 데리고 가 줘서 오마카세 처음 알았어."},         {"speaker": 1, "utterance": "그때 나는 애는 초밥이 뭔지 모르나 보더라는 생각이 들어서 데리고 갔지"},         {"speaker": 2, "utterance": "무시하는 거야? "},         ...       ],       "subject_keyword": "초밥집"     },     "output": "이 대화에서 화자 1과 2는 초밥집에 대해 이야기했습니다. 1은 2에게 자신을 데리고 가 줬던 초밥집에서 오마카세를 처음 알게 되었고 신세계였다고 이야기했습니다. 그리고 음식은 배 채우려고 먹는 게 아니라 맛있는 음식을 먹어야 한다고 생각했다고 이야기했습니다(...)"   },   ... ]</pre>
평가용 데이터	<pre>[   {     - 아이디(id)와 입력(input), 그리고 출력(output)으로 구성</pre>



형식의 예 (제출 전)	<pre> {id": "nikluge-2024-일상대화의주제별요약-train-000001", "input": {   "conversation": [     {"speaker": 1, "utterance": "우리 그때 처음에 나 데리고 가 줬던 초밥집 기억나?"},     {"speaker": 2, "utterance": "Company-name 상암동에 있는 데?"},     {"speaker": 1, "utterance": "어 나 진짜 깜짝 놀랐잖아. 나는"},     {"speaker": 2, "utterance": "그때까지 초밥은 그냥 회전 초밥집? 너가 데리고 가 줘서 오마카세 처음 알았어."},     {"speaker": 1, "utterance": "그때 나는 애는 초밥이 뭔지 모르나 보다라는 생각이 들어서 데리고 갔지"},     {"speaker": 2, "utterance": "무시하는 거야? "},     ...   ],   "subject_keyword": "초밥집" } }, ... ] </pre>
	<ul style="list-style-type: none"> <li>- 학습용 데이터와 동일한 형태</li> <li>- "output" 키와 값을 제거한 데이터</li> </ul>
제출 데이터 형식의 예	<pre> [ {   {     "id": "nikluge-2024-일상대화의주제별요약-train-000001",     "input": {       "conversation": [         {"speaker": 1, "utterance": "우리 그때 처음에 나 데리고 가 줬던 초밥집 기억나?"},         {"speaker": 2, "utterance": "Company-name 상암동에 있는 데?"},         {"speaker": 1, "utterance": "어 나 진짜 깜짝 놀랐잖아. 나는"},         {"speaker": 2, "utterance": "그때까지 초밥은 그냥 회전 초밥집? 너가 데리고 가 줘서 오마카세 처음 알았어."},         {"speaker": 1, "utterance": "그때 나는 애는 초밥이 뭔지 모르나 보다라는 생각이 들어서 데리고 갔지"},         {"speaker": 2, "utterance": "무시하는 거야? "},         ...       ],       "subject_keyword": "초밥집"     },     "output": "이 대화에서 화자 1과 2는 초밥집에 대해 이야기했습니다. 1은 2에게 자신을 데리고 가 줬던 초밥집에서 오마카세를 처음 알게 되었고 신세계였다고 이야기했습니다. 그리고 음식은 배 채우려고 먹는 게 아니라 맛있는 음식을 먹어야 한다고 생각했다고 이야기했습니다(...)"   },   ... ] </pre>
	<ul style="list-style-type: none"> <li>- 평가용 데이터에 "output"키와 값을 생성.</li> </ul>

표 3 데이터 형식의 예

표 4는 데이터의 규모이다. 데이터 세트는 훈련(train), 검증(validation), 시험(test) 데이터로 무작위 분할하여 제공한다.

	훈련	검증	시험
대화 수	506	102	408

표 4 데이터 규모



#### 4. 기준 모델(한국어 언어 모델 블로섬 Bllossom 기반)

이 대회 기준 모델인 블로섬(Blossom)은 허깅페이스(huggingface)에 공개되어 있고, 과제 수행 시 기준 모델의 학습과 평가를 재현하기 위한 코드는 깃허브(github)에서 확인할 수 있다.

- [기준 모델] <https://huggingface.co/MLP-KTLim/llama-3-Korean-Blossom-8B>
- [코드] [https://github.com/teddysum/Korean\\_DCS\\_2024](https://github.com/teddysum/Korean_DCS_2024)

#### 5. 대회 참가 규정

- 인공지능의 한국어 능력 평가에 관심 있는 누구나 팀을 구성하여 참가할 수 있다.
- 팀 구성원은 국립국어원 언어정보나눔터 회원이어야 한다.
- 팀장은 팀원의 전자 우편 주소(회원 가입된 전자 우편 주소)를 입력하여 팀을 구성할 수 있으며, 모든 팀원이 참가 신청에 동의하면 팀 구성이 완료된다.
- 참가자는 접수 기간 중 여러 개의 팀에 참여할 수 있으며, 팀 구성 인원 제한은 없다. 다만, 접수 기간 종료 후 정성 평가 대상이 되는 팀이 여러 개라면 참가자는 과제별 하나의 팀을 소속 팀으로 결정하여야 한다.
- 참가 팀은 과제를 해결한 결과를 정해진 양식에 맞추어 제출한다(1일 5개까지 제출 가능).
- 라이선스에 문제가 없는 모델을 개발하여야 한다.
- 외부 API를 통해 호출하는 모델(OpenAI API 등)은 제출할 수 없다.
- 개발한 모델은 RTX 4090 24GB 1개에서 구동 가능하여야 한다.
- 외부에 공개된 데이터는 사용이 불가하며 과제 데이터(말뭉치)를 초거대언어모델(LLM) 등을 활용해 변형 및 증강하여 사용할 수 없다. 모델 입력을 위한 형식 및 형태 변형만이 허용된다.
- 순위표 기준 정성 평가 신청서를 제출한 상위 5개 팀을 정성 평가 및 발표 평가 대상으로 선정한다. (해당 팀에게는 별도 안내 예정)
- 최종 수상자는 정량 평가, 정성 평가, 발표 평가를 종합하여 선정한다.

#### 6. 시스템 사용 방법

- 인공지능(AI)말평 → 알립니다 → 사용 안내 메뉴 내 동영상 참고

#### 7. 문의

- 진행 중 과제 → 일상 대화 요약 클릭 → 문의(들여가기 후 작성 가능)