

국립국어원 2023-01-28

발간등록번호
11-1371028-000958-01

2023년 대화 요약 말뭉치 연구 분석

연구책임자
함 영 균

제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구용역 계약에 따라 ‘2023년 대화 요약 말뭉치 연구 분석’에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업 기간: 2023년 5월 2일 ~ 2023년 11월 30일

2023년 11월 30일

연구책임자: 함영균(쥬테디썸)

연구 기관: 쥬테디썸, 연세대학교 산학협력단

연구책임자: 함영균

공동연구원: 김한샘

보조연구원: 정용빈, 이이슬, 오유진

박재완, 이지인

김유진, 강조은

<국문 요약>

2023년 대화 요약 말뭉치 연구 분석

본 사업은 일상 대화 및 회의록 요약에 대한 대규모 말뭉치 구축을 통해 국어 자원의 가치를 높이고 인공지능 기술 개발, 활용을 위한 국내 언어 정보 표준화 및 참고 기반 자료로 활용될 수 있는 데이터 세트 구축을 목표로 한다. 일상 대화 요약은 소주제별 중요 표현 추출 및 연결(문장, 화행 유형 주석, 중심어 선택, 문장 관계 주석 등)을 넘어 화자별 핵심 의견 요약문과 소주제별 추상 요약문, 전체 요약문을 작성한다. 회의록 요약은 주제 분할과 중요 표현 추출(중심어 선택, 화행 유형 주석)을 통해 추상 요약문(세부 요약문, 주제별 요약문 및 전체 요약문)을 작성한다. 이러한 요약 분석을 통해 일상 대화 및 논쟁적 대화 요약 모델의 다양한 평가 지표를 제공하고, 기술 발전을 위한 기반 자료로 활용할 것으로 기대한다.

본 사업에서는 일상 대화 요약 말뭉치 구축을 위한 일상 대화 요약 분석 지침을 신규 수립하고 “국립국어원 일상 대화 말뭉치 2020, 2021” 분석을 통해 선별한 원시 말뭉치 3,000건을 대상으로 요약 말뭉치를 구축한다. 또한 “국립국어원 국회 회의록 말뭉치 2021” 중 300건을 대상으로 핵심 정보를 추출하여 요약문을 작성하고 이를 기반으로 “2022 회의록 요약 분석”(국립국어원, 2022) 사업에서 구축한 “논쟁적 대화 요약 말뭉치 구축 지침”을 개선 및 보완하고 실제 적용 방안을 검토한다. 본 사업의 주요 목표와 성과는 다음과 같다.

▶ 일상 대화 요약을 위한 지침 신규 수립

일상 대화 요약 분석 지침은 구어체 대화를 문어체 대화로 요약하는 것뿐만 아니라 대화 속 화자별 핵심 의견 요약문과 소주제별 요약문을 작성한 다음 이를 바탕으로 전체 요약문을 작성하는 상향식 요약을 통해 깊이 있는 정보를 담은 요약문을 작성하는 지침을 포함하고 있다. 또 소주제별 중심어 작성 및 중요 표현 추출(문장 주석), 발화 정보 작성(화행 유형 주석)을 통해 발화의 다양한 정보를 구축할 수 있는 세부 지침을 마련하였다. 본 사업에서는 많은 작업자가 일관성 있는 요약문을 작성할 수 있도록 상세한 설명과 예시를 포함하는 지침을 작성하여 향후 대규모 일상 대화 요약 분석을 위한 표준화 및 기반 자료가 될 수 있는 기틀을 확보하였다.

▶ 논쟁적 대화 요약을 위한 지침 고도화

논쟁적 대화 요약 말뭉치 구축 지침(회의록 요약 말뭉치 연구 분석 보고서, 국립국어

원(2021, 2022))은 구어체 대화를 문어체로 요약하는 것 외에도 논쟁적 대화의 심층 정보를 담은 추상적 요약문을 상향식으로 작성하는 대지침과 국회 회의록에 적용될 수 있는 세부 지침을 포함하고 있다. 본 사업에서는 추가적인 요약 말뭉치 구축과 교차 검증을 통해 분석 지침을 고도화하여 향후 말뭉치 요약 연구에 활용할 수 있는 기반 자료를 마련하였다.

▶ 3,000건의 일상 대화 요약 말뭉치 및 300건의 회의록 요약 말뭉치 구축

본 사업에서는 개발된 지침을 활용하여 일상 대화 말뭉치 3,000건을 소주제별로 구분하고 중심어 작성, 중요 표현 추출 및 화자별, 소주제별 요약문을 작성한 후 이를 토대로 전체 요약문을 작성하였다. 국회 소위원회 회의록 역시 300건을 소주제별로 구분하고 중심어 작성 및 소주제별, 전체 요약문을 작성하였다. 본 사업을 통해 구축된 데이터는 일상 대화 및 논쟁적 대화 요약 모델의 다양한 평가 지표를 제공할 것으로 기대된다.

주요어: 요약, 일상 대화 요약, 일상 대화 요약 말뭉치 구축, 회의록 요약, 회의록 요약 말뭉치 구축

<Abstract>

Analysis of the Conversation Summary Corpus Research, 2023

This project is dedicated to the construction of a substantial corpus of NIKL Korean Dialogue Corpus 2020, 2021 and NIKL Korean Parliamentary Corpus 2021 summaries, to enhance the value of Korean language resources and establish a dataset that can be utilized as a reference for the standardization of domestic language information, thereby aiding in the development and application of artificial intelligence technologies. The NIKL Korean Dialogue Corpus 2020, 2021 summaries surpass the mere extraction and linkage of important expressions by topic (including sentence, speech act type annotation, central word selection, sentence relationship annotation, etc.), to generate comprehensive summaries encapsulating key opinions by speaker, abstract summaries by topic, and overall summaries, which can serve as foundational material for various technological developments. Summarising NIKL Korean Parliamentary Corpus 2021 involves topic segmentation and extraction of relevant expressions, including the identification of key terms and speech act classifications, culminating in abstract summaries comprising detailed breakdowns, topic-specific summaries, and overarching summaries. The resultant summaries will provide a basis for diverse technological advances.

This project aims to establish new guidelines for analyzing NIKL Korean Dialogue Corpus 2020, 2021 summaries to construct a NIKL Korean Dialogue Corpus 2020, 2021 summary corpus. A summary corpus will be built based on 3,000 raw transcripts selected through the analysis of the 'NIKL Korean Dialogue Corpus 2020, 2021' Key information will be extracted from 300 transcripts of the NIKL Korean Parliamentary Corpus 2021 to create summaries. Based on this, the Guidelines for the Construction of Controversial Conversation Summary Corpus established in the Summary Analysis of 2022 Minutes (National Institute of Korean Language, 2022) project will be refined and augmented, and practical applications will be examined. The principal goals and achievements of this project are as follows:

- Establishment of New Guidelines for NIKL Korean Dialogue Corpus 2020, 2021 Summaries:

The guidelines encompass a comprehensive approach for creating in-depth summaries through bottom-up summarization. This entails not only converting colloquial

conversations into written form but also compiling speaker-specific key opinion summaries and thematic summaries, which are then integrated into an overall summary. Additionally, the guidelines include detailed protocols for constructing diverse information about the speech, such as thematic keyword creation, important expression extraction (sentence annotation), and speech act type annotation. These guidelines, complete with detailed explanations and examples, aim to enable multiple individuals to produce consistent summaries and lay the groundwork for future large-scale NIKL Korean Dialogue Corpus 2020, 2021 summary analysis.

- Enhancement of Guidelines for Summarizing Contentious Conversations:

The guidelines for constructing a summary corpus of contentious conversations (Meeting Minutes Summary Corpus Research and Analysis Report, National Institute of the Korean Language (2021, 2022)) include high-level protocols for creating abstract summaries containing in-depth information of contentious conversations through a bottom-up approach, as well as detailed guidelines applicable to parliamentary minutes. This project has refined the analysis guidelines through the construction and cross-validation of additional summarization corpora, thereby preparing foundational material for future corpus summarization research.

- Construction of a Corpus of 3,000 NIKL Korean Dialogue Corpus 2020, 2021 Summaries and 300 NIKL Korean Parliamentary Corpus 2021 Summaries:

Utilizing the developed guidelines, the project successfully categorized 3,000 NIKL Korean Dialogue Corpus 2020, 2021 transcripts by theme, identified keywords, extracted important expressions, and created speaker-specific and topic-specific summaries. These were then synthesized into comprehensive overall summaries. Similarly, 300 parliamentary subcommittee NIKL Korean Parliamentary Corpus 2021 were categorized by topic, with keywords identified and summarized both by topic and in an overall summary. The data compiled through this project is expected to provide various evaluation indicators for models summarizing both NIKL Korean Dialogue Corpus 2020, 2021 and argumentative conversations.

Key Words: Summarization, Daily Conversation Summarization, Daily Conversation Summary Corpus Construction, Meeting Minutes Summarization, Meeting Minutes Summary Corpus Construction.

차 례

제1장 사업 개요

1.1. 사업 목적	1
1.2. 사업 수행 범위	2
1.3. 사업 수행 절차	3

제2장 사업 수행

2.1. 일상 대화 요약 분석 말뭉치 구축	5
2.1.1 일상 대화 요약 분석 말뭉치 작업 개요	5
2.1.2 일상 대화 요약 분석 말뭉치 선정 및 특징	8
2.1.3 일상 대화 요약 분석 말뭉치 구축 분석 지침	21
2.1.4 일상 대화 요약 분석 말뭉치 구축 작업자 선발 및 교육	22
2.1.5 일상 대화 요약 분석 말뭉치 구축 플랫폼 개발	25
2.1.6 일상 대화 요약 분석 말뭉치 구축 예시	39
2.2. 회의록 요약 분석 말뭉치 구축	44
2.2.1 회의록 요약 분석 말뭉치 작업 개요	44
2.2.2 회의록 요약 분석 말뭉치 선정 및 특징	46
2.2.3 회의록 요약 분석 말뭉치 구축 작업자 선발 및 교육	53
2.2.4 회의록 요약 분석 말뭉치 구축 플랫폼 개발	55
2.2.5 회의록 요약 분석 말뭉치 구축 분석 지침	63

차 례

제3장 사업 수행 결과

3.1. 일상 대화 요약 분석 말뭉치 구축 결과	65
3.1.1 일상 대화 요약 분석 말뭉치 통계	65
3.1.2 일상 대화 요약 분석 말뭉치 검수 결과	70
3.1.3 일상 대화 요약 분석 말뭉치 형식	80
3.2. 회의록 요약 분석 말뭉치 구축 결과	88
3.2.1 회의록 요약 분석 말뭉치 통계	88
3.2.2 회의록 요약 분석 말뭉치 검수 결과	90
3.2.3 회의록 요약 분석 말뭉치 형식	99
3.3. 기타(검수 프로그램 개발)	109

제4장 결론 및 추후 전망

4.1. 프로젝트 요약	111
4.2. 일상 대화 요약 분석 말뭉치 구축 방향 제언	112
4.3. 향후 발전 가능성	117

부록

- [붙임 1] 일상 대화 요약 분석 말뭉치 구축 분석 지침
- [붙임 2] 회의록 요약 분석 말뭉치 구축 분석 지침

표 차례

<표 1> 과업 범위 및 세부 과업 내용	2
<표 2> 일상 대화 요약 분석 작업 순서	6
<표 3> 일상 대화 원시 말뭉치 통계	8
<표 4> 일상 대화 요약 분석 말뭉치에서 제외한 세부 주제	8
<표 5> 일상 대화 요약 분석 말뭉치에서 선정한 세부 주제	9
<표 6> 일상 대화 요약 분석 말뭉치 세부 주제별 상위 분류	10
<표 7> 일상 대화 원시 말뭉치 오류 데이터 현황표	14
<표 8> 일상 대화 원시 말뭉치의 JSON 구조	17
<표 9> 일상 대화 원시 말뭉치 JSON 형식의 예	18
<표 10> 일상 대화 요약 분석 작업자 개요	22
<표 11> 일상 대화 요약 작업자 교육 일정	23
<표 12> 대화 요약 분석 말뭉치 구축 플랫폼 개요	26
<표 13> 일상 대화 요약 분석의 문장 주석 기준	30
<표 14> 일상 대화 요약 분석의 화행 유형 주석 기준	32
<표 15> 일상 대화 요약 분석의 중심 문장 간 관계 주석 기준	33
<표 16> 회의록 요약 작업 개요	45

표 차례

<표 17> 분석 대상 회의록 목록	46
<표 18> 회의록 원시 말뭉치 JSON의 구조	50
<표 19> 회의록 원시 말뭉치 JSON 형식의 예	51
<표 20> 회의록 요약 분석 작업자 개요	53
<표 21> 회의록 요약 작업자 교육 일정	53
<표 22> 회의록의 세부 요약문 유형	57
<표 23> 일상 대화 요약 분석 말뭉치의 문장 주석 비율	65
<표 24> 일상 대화 요약 분석 말뭉치의 중심 문장 수	66
<표 25> 일상 대화 요약 분석 말뭉치의 중심 문장에 연결된 뒷받침 문장 수 ...	66
<표 26> 일상 대화 요약 분석 말뭉치의 화행 유형 주석 비율	67
<표 27> 일상 대화 요약 분석 말뭉치의 중심 문장 간 관계 주석 비율 ..	67
<표 28> 일상 대화 요약 분석 말뭉치의 중심어 비율	68
<표 29> 일상 대화 요약 분석 말뭉치의 문서별, 소주제별 중심어 수 ...	68
<표 30> 일상 대화 요약 분석 말뭉치의 중심어 길이	69
<표 31> 일상 대화 요약 분석 말뭉치의 결정 사항 수	69
<표 32> 일상 대화의 중심어 선택 오류 유형	70

표 차례

<표 33> 일상 대화의 문장 및 화행 유형 주석 오류 유형	71
<표 34> 일상 대화의 요약문 작성 오류 유형	74
<표 35> 일상 대화 작업 지침 외 참고 사항 예시	78
<표 36> 일상 대화 요약문 작성 시 주의 사항 예시(주어)	78
<표 37> 일상 대화 사업 종료 후 추가 보완 작업 일정	79
<표 38> 일상 대화 요약 분석 말뭉치의 메타 데이터 구조	80
<표 39> 일상 대화 요약 분석 말뭉치의 메타 데이터 정보 예시	81
<표 40> 일상 대화 요약 분석 말뭉치의 발화 정보 구조	82
<표 41> 일상 대화 요약 분석 말뭉치의 발화 정보 예시	83
<표 42> 일상 대화 요약 분석 말뭉치의 요약 정보 구조	85
<표 43> 일상 대화 요약 분석 말뭉치의 요약 정보 예시	86
<표 44> 회의록 요약 분석 말뭉치 관련 통계 비교	88
<표 45> 회의록 요약 분석 말뭉치 통계	88
<표 46> 회의록 원시 말뭉치에 대한 요약문의 비율	89
<표 47> 회의록의 표현 및 띄어쓰기 오류 유형	90
<표 48> 회의록의 작성 누락 및 길이 오류 유형	93

표 차례

<표 49> 회의록 지침 외에 참고할 오류 유형 및 예시	96
<표 50> 회의록 교차 검수 및 보완 일정	97
<표 51> 회의록 요약 분석 말뭉치의 메타 데이터 구조	99
<표 52> 회의록 요약 분석 말뭉치의 메타 데이터 예시	101
<표 53> 회의록 요약 분석 말뭉치의 발화 정보 구조	102
<표 54> 회의록 요약 분석 말뭉치의 발화 정보 예시	103
<표 55> 회의록 요약 분석 말뭉치의 요약 정보 구조	105
<표 56> 회의록 요약 분석 말뭉치의 요약 정보 예시	107

그림 차례

[그림 1] 대화 요약 말뭉치 연구 분석 사업 추진 체계	3
[그림 2] 일상 대화 원시 말뭉치의 자문자답 오류 데이터 예시	12
[그림 3] 일상 대화 원시 말뭉치의 발화자 ID 오류 데이터 예시	13
[그림 4] 일상 대화 원시 말뭉치의 세부 주제(Topic) 오류 데이터 예시	14
[그림 5] 테디썸의 오류 데이터 대응 과정	15
[그림 6] 국립국어원의 오류 데이터 대응 과정	15
[그림 7] 테디썸의 오류 데이터 관리 예시	16
[그림 8] 일상 대화 요약 작업자 온라인 교육 화면	24
[그림 9] 대화 요약 분석 말뭉치 구축 플랫폼 화면	25
[그림 10] 일상 대화의 소주제 분할을 위해 발화를 선택하는 기능	27
[그림 11] 일상 대화의 소주제 분할 작업 시 검수자의 피드백을 받을 수 있는 기능	28
[그림 12] 일상 대화의 중심어 경고 알림창 기능	29
[그림 13] 일상 대화의 요약 분석을 위한 중심어 작성 기능	29
[그림 14] 일상 대화의 문장 유형 경고 알림창	30

그림 차례

[그림 15] 일상 대화의 요약 분석을 위한 문장 주석 기능 1	30
[그림 16] 일상 대화의 요약 분석을 위한 문장 주석 기능 2	30
[그림 17] 일상 대화의 요약 분석을 위한 뒷받침-중심 문장 연결 기능	31
[그림 18] 일상 대화의 뒷받침-중심 문장 연결 경고 알림창	31
[그림 19] 일상 대화의 요약 분석을 위한 화행 유형 주석 기능	32
[그림 20] 일상 대화의 화행 유형 경고 알림창	33
[그림 21] 일상 대화의 요약 분석을 위한 중심 문장 간 관계 주석 기능 ..	33
[그림 22] 일상 대화의 중심 문장 간 관계 주석 경고 알림창	34
[그림 23] 일상 대화의 요약 분석을 위한 소주제별 요약문 작성 기능 ..	34
[그림 24] 일상 대화의 요약 분석을 위한 전체 요약문 작성 기능	35
[그림 25] 일상 대화의 검수 노트(note) 기능	36
[그림 26] 일상 대화 요약 작업자의 작업 현황 확인 기능	37
[그림 27] 일상 대화 요약 관리자의 작업 현황 확인 기능	37
[그림 28] 일상 대화 요약 분석 플랫폼의 질의응답 화면	38
[그림 29] 일상 대화 요약의 소주제 분할 예시	39

그림 차례

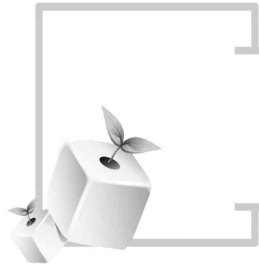
[그림 30] 일상 대화 요약의 중심어 선택, 문장 및 화행 유형 주석 예시 ...	40
[그림 31] 일상 대화 요약의 중심 문장 간 관계 주석 예시	41
[그림 32] 일상 대화 요약의 소주제별 요약문 작성 예시	42
[그림 33] 일상 대화 요약의 전체 요약문 작성 예시	43
[그림 34] 회의록 요약 작업자 온라인 교육 화면	54
[그림 35] 회의록의 주제 분할을 위해 일부 발화를 선택하는 기능 ..	55
[그림 36] 회의록의 주제(쟁점)에 해당하는 안건을 선택하는 기능 ..	56
[그림 37] 회의록의 주제(쟁점) 분할 기능	56
[그림 38] 회의록의 세부 요약문 작성 예시	57
[그림 39] 회의록의 세부 요약문 작성 기능	58
[그림 40] 상향식 대표 요약문 작성의 도식도	59
[그림 41] 회의록의 대표 요약문 작성 기능	59
[그림 42] 회의록의 검수 의견 작성 및 데이터 수정 기능	60
[그림 43] 회의록 요약 작업자의 작업 현황 확인 기능	61
[그림 44] 회의록 요약 관리자의 작업 현황 확인 기능	61
[그림 45] 회의록 요약 분석 플랫폼의 질의응답 화면	62

그림 차례

[그림 46] 일상 대화의 중심어 선택 오류	71
[그림 47] 일상 대화에서 중심 문장을 과하게 짧게 주석한 경우	72
[그림 48] 일상 대화에서 문장 주석을 잘못된 경우	73
[그림 49] 일상 대화에서 화자별 요약의 내용을 대화 요약에 그대로 사용한 경우	75
[그림 50] 일상 대화에서 화자별 요약의 내용을 반대로 작성한 경우 ...	75
[그림 51] 일상 대화 요약문에서 중심 문장의 내용이 누락된 경우	76
[그림 52] 일상 대화 요약문에서 비문 및 내용 누락이 발생한 경우	76
[그림 53] 일상 대화에서 소주제별 요약문에 비해 전체 요약문의 내용이 짧은 경우	77
[그림 54] 회의록 요약문의 작성 오류 예시(구어체)	91
[그림 55] 회의록 요약문의 작성 오류 예시(표현 오류)	91
[그림 56] 회의록 요약문의 작성 오류 예시(띄어쓰기)	92
[그림 57] 회의록 요약문의 작성 오류 예시(오타)	92
[그림 58] 회의록 요약문의 작성 오류 예시(비문)	94

그림 차례

[그림 59] 회의록 요약문의 작성 오류 예시(길이)	94
[그림 60] 회의록 요약문의 작성 오류 예시(작성 누락)	95
[그림 61] 대화 요약 검수 프로그램 화면	109
[그림 62] 대화 요약 검수 프로그램 화면(비공백 검색 예시)	110
[그림 63] 사업 개요(일상 대화 요약)	111
[그림 64] 사업 개요(회의록 요약)	111
[그림 65] 일상 대화 원시 말뭉치 중 하나의 소주제 안에 발화자가 한 명인 데이터 예시 1	113
[그림 66] 일상 대화 원시 말뭉치 중 하나의 소주제 안에 발화자가 한 명인 데이터 예시 2	113
[그림 67] 일상 대화 원시 말뭉치 중 빈칸 개입이 많은 데이터 예시 ..	114
[그림 68] 일상 대화 원시 말뭉치 중 발화가 너무 짧게 구분되어 있는 데이터 예시	115
[그림 69] 일상 대화 원시 말뭉치의 특징에 기반한 제안 사항	116
[그림 70] 일상 대화 요약 플랫폼 기능 개선 항목	117
[그림 71] 회의록 요약 플랫폼 기능 개선 항목	118



제 1 장

사업 개요



1.1 사업 목적

일상 대화 말뭉치는 사람들이 일상 속에서 겪는 다양한 상황과 주제에 대한 대화 말뭉치이며, 논쟁적 대화 말뭉치는 다양한 분야에서 아이디어를 교환하고, 계획을 수립하거나 정보를 공유하는 데 사용되는 보편적이고 목적 지향적인 대화 말뭉치이다. 이러한 대화 말뭉치는 음성인식 등의 기술을 바탕으로 전사(transcription)하여 보관되고 있다. 일상 대화 및 논쟁적 대화 말뭉치 연구 분석을 통해 실제 대화에서 발생할 수 있는 다양한 언어적 패턴과 대화의 유형, 흐름 등을 파악하고 다양한 응용 기술 개발을 위한 기반 자료로 활용할 수 있을 것이다.

본 사업의 목적은 크게 두 가지로 나뉜다. 첫 번째, 일상 대화 요약에 위한 방법론 및 지침을 신규 수립하여 “국립국어원 일상 대화 말뭉치 2020, 2021”에 대한 요약 분석 말뭉치 3,000건을 구축하는 것이다. 이를 위해 대화 내 소주제별 구간 분할을 통해 말뭉치를 구조화하고 소주제별 중심어 작성, 중요 표현 추출 및 연결(문장, 화행 유형 주석 등)을 통해 대화 정보를 분석함으로써 핵심 발화를 선정하여 화자별 요약 및 소주제별 요약문을 작성한다. 작성한 소주제별 요약문을 토대로 전체 요약문을 작성하는 상향식 요약 방법을 활용하고 이를 위한 요약 분석 지침을 마련한다. 또한 요약 분석 말뭉치의 일관성 및 정확성 확보를 위한 검증 방안을 수립하고 최종 분석 말뭉치 납품 이후의 보완 체계를 수립 및 제안한다.

두 번째, 2023 국회 회의록 요약 분석을 통해 기존에 개발된 ‘논쟁적 대화 요약 분석 말뭉치 구축 지침’을 고도화하고 이를 토대로 국회 소위원회 회의록 300건을 분석하는 것이다. 본 사업은 “회의록 요약 분석 2021, 2022”의 연속 사업으로 대화 내 주제별 구간 분할과 주제별 중심어 작성, 핵심 발화 선정, 세부 요약문과 주제별 작성을 수행한 후 전체 요약문을 작성하는 상향식 요약문 작성 방식을 채택하였으며 국회 회의록 300건에 대한 요약 분석 데이터를 추가 구축하였다.

일상 대화 요약 분석 말뭉치 3,000건과 국회 회의록 요약 분석 말뭉치 300건을 분석·배포함으로써 국내 인공지능 관련 연구 및 기술 개발을 위한 언어 정보 표준화 및 기반 자료로 활용되기를 기대한다.

1.2 사업 수행 범위

본 사업에서 수행한 과업의 범위는 다음과 같다.

과업	목표	세부 과업 내용
일상 대화	일상 대화 요약 분석 지침 수립 및 일상 대화 말뭉치 요약 분석(3,000건)	<ul style="list-style-type: none"> - '일상 대화 요약 분석 말뭉치 구축 지침' 신규 수립 - 국립국어원이 제공하는 일상 대화 원시 말뭉치에 대한 요약 분석 말뭉치 3,000건 구축 - 일상 대화 소주제 분할 및 분석, 소주제별 핵심 키워드 선택 - 문장 주석(중심 문장, 뒷받침 문장, 잉여적 문장) 및 화행 유형 주석(질문, 정보 제공, 반응 등) - 중심 문장 간 관계 주석 및 소주제별 요약문 작성(대화 요약, 화자별 요약, 결정 사항) - 일상 대화 전체에 대한 최종 요약문 작성(소주제별 요약문의 '대화 요약' 내용을 기계적으로 통합) - 구축된 요약 분석 말뭉치에 대한 일관성 및 정확성 검증 방안 및 검증 체계 수립 및 제안(교차 검증 수행)
국회 회의록	논쟁적 대화 요약 분석 지침 고도화 및 국회 회의록 말뭉치 요약 분석(300건)	<ul style="list-style-type: none"> - '논쟁적 대화 요약 분석 말뭉치 구축 지침' 고도화 - 국립국어원이 제공하는 국회 소위원회 회의록에 대한 요약 분석 말뭉치 300건 구축 - 회의록 주제 분할 및 분석, 주제별 핵심 키워드 선택 - 주제별 요약문 작성(세부 요약문, 주제별 요약문), 요약문 작성 시 참조한 발화 명시(주요 문장 추출) - 회의록 전체에 대한 최종 요약문 작성 - 구축된 요약 분석 말뭉치에 대한 일관성 및 정확성 검증 방안 및 검증 체계 수립 및 제안(교차 검증 수행)

표 1 과업 범위 및 세부 과업 내용

1.3 사업 수행 절차

일상 대화 요약 말뭉치 분석은 일관성 있는 요약 분석 말뭉치 구축을 위한 ‘일상 대화 요약 분석 지침’을 신규 수립하고 이를 토대로 일상 대화 요약 분석 말뭉치 3,000건을 구축한다. 논쟁적 대화 요약 말뭉치 분석은 ‘논쟁적 대화 요약 분석 말뭉치 구축 지침(회의록 요약 말뭉치 연구 분석 보고서, 국립국어원(2021, 2022))’을 세밀화, 고도화하고 이를 바탕으로 회의록 요약 분석 말뭉치 300건을 구축한다.

일관성 있는 데이터를 보다 효율적으로 구축하기 위해 요약 분석에 특화된 데이터 구축 플랫폼을 개발·개선하고 해당 플랫폼 내에서 소주제 분할, 중심어 선택, 소주제별 요약문, 전체 요약문 작성 등의 분석 작업과 지침 교육 및 검수를 수행하였다. 일상 대화 및 회의록 요약 분석은 ‘일상 대화 요약 분석 지침’과 ‘논쟁적 대화 요약 분석 말뭉치 구축 지침’을 준수하였으며 모든 과정은 검수자 및 관리자의 주기적인 OT와 교육, 모니터링을 통해 이루어졌다.

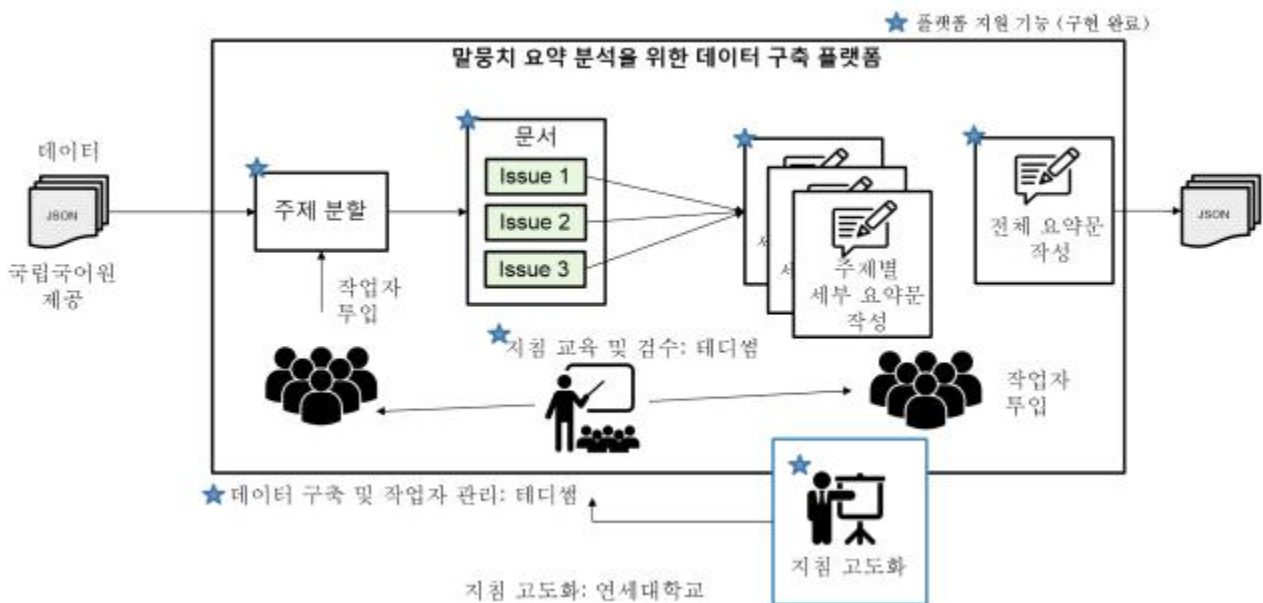
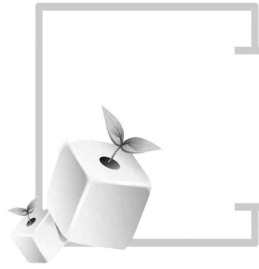


그림 1 대화 요약 말뭉치 연구 분석 사업 추진 체계



제 2 장

사업 수행



2.1 일상 대화 요약 말뭉치 구축

2.1.1 일상 대화 요약 분석 말뭉치 작업 개요

□ 일상 대화 요약 분석 말뭉치 작업 개요

일상 대화 요약 작업은 크게 소주제 분할과 요약 분석으로 구분된다. 소주제 분할은 대화 내 소주제에 따라 발화를 분류하는 작업으로 소주제 분할이 완료되면 분할된 소주제를 기준으로 발화 주석, 요약문 작성 등 요약 분석을 수행한다. 일상 대화 요약 분석 말뭉치는 다음과 같은 요소로 구성된다.

- 1) 소주제 분할 : 대화 주제에 따른 발화 분류 수행
- 2) 요약 분석

가) 발화 주석(문장 주석)

- (1) 발화 종류(문장 주석) : 중심 문장, 뒷받침 문장, 잉여적 문장 중 택 1

* 문장 주석을 한 다음 반드시 1개의 화행을 주석해야 한다.

- ① 중심 문장 : 소주제와 관련된 핵심적인 주장이나 의견, 질문이 포함된 발화(들)
 - 중심 문장에 적절한 중심어가 있는 경우 단어 또는 명사구로 작성하되 반드시 발화에서 그대로 가져와(복사-붙여넣기) 작성한다.
- ② 뒷받침 문장 : 중심 문장에 대한 뒷받침/부연 설명 발화
 - 선택한 뒷받침 문장이 어떤 중심 문장에 대한 뒷받침 문장인지 반드시 선택해야 한다.
- ③ 잉여적 문장 : 중심 문장, 뒷받침 문장이 아닌 모든 발화(들)

- (2) 화행 정보 주석 : 질문, 요구(요청), 정보 제공, 반응, 감정 표현, 기타 중 택 1

* 모든 문장에 반드시 1개의 화행을 주석해야 한다.

- ① 질문 : 예/아니오 질문, 의문사 포함 질문 등
- ② 요구(요청) : 요청, 명령, 제안 등등
- ③ 정보 제공 : 주장, 정정, 추측, 소망, 의지 등등

- ④ 반응 : 질문에 대한 답변, 동의 표현, 비동의 표현, 수락, 거부, 회피 등
- ⑤ 감정 표현 : 인사, 칭찬, 비판, 감사, 사과, 비난, 감탄사 등
- ⑥ 기타 : 위의 화행 구분에 들어가지 않는 그 외 모든 표현

나) 중심 문장 간 관계 주석 : 지지, 반박, 무관 중 택 1

(1) 중심 문장 간 관계 주석 요소

- ① 지지 관계 : 중심 문장 간 관계가 긍정적일 때 주석
(지지, 선호, 인정, 공감 등)
- ② 반박 관계 : 중심 문장 간 관계가 부정적이며 상충할 때 주석
(반박, 거절 등)
- ③ 무관 관계 : 중심 문장 간 관계가 성립하지 않는 경우일 때 주석

다) 주제별 요약문 작성 : 중심 문장, 뒷받침 문장 등의 내용을 중심으로 주제에 대한 요약문 작성

라) 전체 요약문 작성 : 주제별 요약문의 ‘대화 요약’ 내용을 바탕으로 전체 요약문 작성

□ 일상 대화 요약 분석 말뭉치 작업 과정

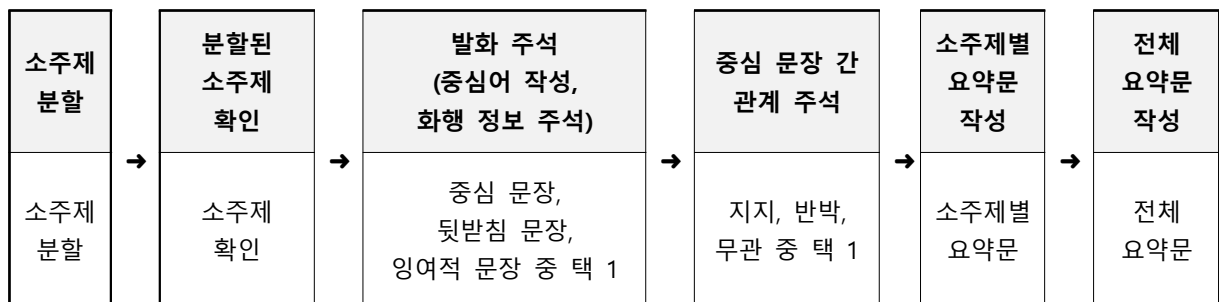


표 2 일상 대화 요약 분석 작업 순서

- 1) 소주제 분할 : 소주제에 따른 발화 분류 수행
- 2) 요약 분석
 - 가) 발화 주석(문장 주석, 화행 정보 주석)

- 발화 주석 : 중심 문장, 뒷받침 문장, 잉여적 문장 중 택 1
- 중심어 작성 : 중심어는 소주제를 잘 나타내는 핵심 단어 또는 명사구로 소주제별로 1개 이상 작성해야 하며 반드시 중심 문장 내에서 그대로 가져와 작성(중심어는 최대 3어절의 복합 명사(구)까지 허용)
- 화행 정보 주석 : 질문, 요구(요청), 정보 제공, 반응, 감정 표현, 기타 중 택 1
- 중심 문장-뒷받침 문장 연결 : 뒷받침 문장의 경우 어떤 중심 문장에 대한 뒷받침 문장인지 선택
- 나) 중심 문장 간 관계 주석 : 지지, 반박, 무관 중 택 1
- 다) 소주제별 요약문 작성 : 중심 문장, 뒷받침 문장으로 주석한 발화 내용을 바탕으로 대화 요약, 화자별 요약, 결정 사항으로 구분하여 요약문 작성
- 라) 전체 요약문 작성 : 소주제별 요약문의 대화 요약 내용을 바탕으로 전체 요약문 작성

2.1.2 일상 대화 요약 분석 말뭉치 선정 및 특징

□ 데이터 선정

일상 대화 요약 분석 말뭉치는 ‘국립국어원 일상 대화 말뭉치 2020, 2021’ 중에서 문서 3,000건을 선정하였고, 선정한 3,000건의 문서는 평균 1,500어절, 발화 수 380개 내외 분량이다.

항목	2020, 2021 일상 대화 원시 말뭉치
문서당 발화자 수 평균	2.09
문서당 발화 수 평균	379.40
문서당 어절 수 평균	1538.74
발화 평균 길이(글자 수)	33.91
발화 평균 길이(어절 수)	4.09

표 3 일상 대화 원시 말뭉치 통계

일상 대화 요약 분석 말뭉치에서 제외한 세부 주제는 논쟁적(협력적) 대화와 주제별 파일 수가 소수인 경우이다. 제외한 세부 주제는 다음과 같다.

*2021 협력적 대화의 데이터 수에는 ‘3, 4인 대화’ 데이터가 포함됨

제외한 세부 주제 및 기타(2020)	데이터 수	제외한 협력적 대화(2021)	데이터 수*	3, 4인 대화
임대료 폭등, 임대차보호법	2	AI의 직업 대체	131	41
무상복지, 복지관, 노인, 경쟁, 일자리	2	청소년에게 인터 넷·스마트폰이 미치는 영향	119	14
실수, 스포츠정신, 평창올림픽, 문제점	2	안락사·존엄사 법제화	113	19
치매, 예방, 보험	1	비대면 생활이 미치는 영향	112	8
디자인, 콜라보, 캐릭터, 주문제작	1	공공 공간의 CCTV 설치	102	19
감정, 복지, 지원금, 폭행, 법	1	가짜 뉴스에 대한 징벌적 손해배상	96	27

연말 선물, 케이크, 취향, 크리스마스, 선물	1	지역 내 기피시설 설치	94	26
기타	405	원자력 발전소의 존폐	90	34
합계	415	합계	857	188

표 4 일상 대화 요약 분석 말뭉치에서 제외한 세부 주제

또 ‘국립국어원 일상 대화 말뭉치 2020, 2021’의 세부 주제 15개를 모두 포괄할 수 있는 다섯 가지 상위 분류를 제안하였으며 세부 주제별 데이터가 균형 있게 포함될 수 있도록 선정하였다. 보다 다양한 요약 분석 말뭉치 구축을 위해 150건 이상의 3, 4인 대화를 포함하여 세부 주제별로 각 100건씩 선정(2020, 2021년 말뭉치 각 1,500건씩 선정)하였으며, 본 사업에서 제안한 상위 분류와 세부 주제는 다음과 같다.

2020 세부 주제	전체 데이터 수	2021 세부 주제	2인 대화 데이터 수	3, 4인 대화 데이터 수
여행지(국내/해외)	100	우정	67	33
회사/학교	100	쇼핑	62	38
영화	100	음악	60	40
아르바이트	100	대중교통	88	12
연애/결혼	100	관혼상제	83	17
먹거리	100	취직	88	12
반려동물	100	회사/학교	89	11
계절/날씨	100	휴가	97	3
가족	100	먹거리	98	2
스포츠/레저	100	반려동물	90	10
방송/연예	100	건강/다이어트	95	5
건강/다이어트	100	방송/연예	87	13
꿈(목표)	100	스포츠/레저	81	19
성격	100	가족	87	13
선물	100	경제/재테크	96	4
합계	1,500	합계	1,268	232

표 5 일상 대화 요약 분석 말뭉치에서 선정한 세부 주제

상위분류	세부 주제	연도	2인 대화 데이터 수	3, 4인 대화 데이터 수	분류별 데이터 총합
여가 및 엔터테인먼트	여행지 (국내/해외)	2020	100	0	900
	휴가	2021	97	3	
	영화	2020	100	0	
	스포츠/레저	2020	100	0	
	스포츠/레저	2021	81	19	
	쇼핑	2021	62	38	
	음악	2021	60	40	
	방송/연예	2020	100	0	
	방송/연예	2021	87	13	
일상생활	계절/날씨	2020	100	0	700
	대중교통	2021	88	12	
	관혼상제	2021	83	17	
	선물	2020	100	0	
	성격	2020	100	0	
	반려동물	2020	100	0	
	반려동물	2021	90	10	
음식 및 건강	먹거리	2020	100	0	400
	먹거리	2021	98	2	
	건강/다이어트	2020	100	0	
	건강/다이어트	2021	95	5	
진로 및 경제	회사/학교	2020	100	0	600
	회사/학교	2021	89	11	
	취직	2021	88	12	
	아르바이트	2020	100	0	
	경제/재테크	2021	96	4	
	꿈(목표)	2020	100	0	
인간관계	가족	2020	100	0	400

	가족	2021	87	13	
	우정	2021	67	33	
	연애/결혼	2020	100	0	
합계			2,768	232	3,000

표 6 일상 대화 요약 분석 말뭉치 세부 주제별 상위 분류

□ 일상 대화 말뭉치의 특징

일상 대화 말뭉치는 위에서 제시한 세부 주제별로 15분 분량의 대화가 발화 순서에 따라 전사된 문서이다. 일상 대화는 한 명의 저자에 의해 주제별로 기술되는 뉴스, 도서 등의 문어 데이터나 특정 주제에 대해 동의, 비동의 등 여러 의견이 등장하는 논쟁적 대화와는 성질이 다르다. 복잡한 논변 구조로 이루어진 논쟁적 대화와 다르게 대부분이 일상 속 다양한 상황과 주제에 관한 자연스러운 대화로 이루어져 있기 때문에 질문, 응답, 감정 표현 등의 여러 발화 유형이 포함되어 있다. 또한 발화자의 발화 습관이나 방식에 따라 문장 구조가 좌우되고 일상적인 언어와 관용구가 포함되어 있기 때문에 대화의 흐름과 의미를 잘 이해할 수 있어야 한다.

또 찬성, 반대와 같은 의견이 나올 수 있는 논쟁적 대화가 아니라 취향이나 고민 등의 일상적 주제로 이루어져 있어 ‘결정 사항’이 거의 나오지 않는다는 특징이 있다.

한편 원시 말뭉치 구축 방식에 따른 특징으로는 과도한 빈칸 개입, 지나치게 짧게 끊어진 발화, 의미 없는 감탄사만 분리된 발화 등이 있다. 이에 대해서는 마지막 장에서 구체적으로 알아보고 보다 나은 데이터 구축을 위해 일상 대화 말뭉치 구축 방향에 대해서도 제안하고자 한다.

□ 일상 대화 원시 말뭉치의 오류 데이터 유형

일상 대화 요약 분석 수행 중 원시 말뭉치에 일부 오류가 있음을 확인하였고 이를 수정하며 요약 작업 및 검수를 수행하였다. 발견한 오류는 크게 세 가지로 구분할 수 있으며 오류 데이터는 국립국어원의 확인 및 수정을 통해 파일을 교체한 후 작업을 수행하였다.

1) 자문자답

원시 말뭉치의 오류 중 자문자답은 한 명의 발화자가 질문을 한 후 상대 발화자가 답변하지 않고 질문을 한 발화자가 답변을 하는 것처럼 작성된 오류이다. 이는 상대 발화자의 ID가 질문을 한 발화자의 ID로 잘못 기입되어 있는 경우이다.

```
{
  "id": "SDRW200000290.1.1.529",
  "form": "학교 체육복 자주 입고 다니셨어요?",
  "original_form": "학교 체육복 자주 입고 다니셨어요?",
  "speaker_id": "SD2000408",
  "start": 864.55205,
  "end": 869.20302,
  "note": ""
},
{
  "id": "SDRW200000290.1.1.530",
  "form": "2예. 저는 체육복",
  "original_form": "2예. 저는 체육복",
  "speaker_id": "SD2000408",
  "start": 869.21307,
  "end": 872.03307,
  "note": ""
},
{
  "id": "SDRW200000290.1.1.531",
  "form": "자주 입고 다녔어요.",
  "original_form": "자주 입고 다녔어요.",
  "speaker_id": "SD2000408",
  "start": 872.04304,
  "end": 873.04203,
  "note": ""
},
}
```

```
{
  "id": "SDRW200000218.1.1.111",
  "form": "몇 년 연애하고 결혼하고 싶어?",
  "original_form": "몇 년 연애하고 결혼하고 싶어?",
  "speaker_id": "SD2000312",
  "start": 307.77706,
  "end": 311.14602,
  "note": ""
},
{
  "id": "SDRW200000218.1.1.112",
  "form": "나는",
  "original_form": "나는",
  "speaker_id": "SD2000312",
  "start": 311.15606,
  "end": 312.61801,
  "note": ""
},
{
  "id": "SDRW200000218.1.1.113",
  "form": "일단 모르는 상태에서 결혼을 하면",
  "original_form": "일단 모르는 상태에서 결혼을 하면",
  "speaker_id": "SD2000312",
  "start": 312.62806,
  "end": 316.13905,
  "note": ""
},
{
  "id": "SDRW200000218.1.1.114",
  "form": "조금 실수할 수 있는",
  "original_form": "조금 실수할 수 있는",
}
```

```
{
  "id": "SDRW2000001626.1.1.1",
  "form": "지금 대학교 1학년이라고 하셨는데",
  "original_form": "지금 대학교 일 학년이라고 하셨는데",
  "speaker_id": "SD2001968",
  "start": 14.26008,
  "end": 17.88301,
  "note": ""
},
{
  "id": "SDRW2000001626.1.1.2",
  "form": "혹시 학교는 어느 대학교 다니고 있어요?지금",
  "original_form": "혹시 학교는 어느 대학교 다니고 있어요?지금",
  "speaker_id": "SD2001969",
  "start": 17.89304,
  "end": 20.88801,
  "note": ""
},
{
  "id": "SDRW2000001626.1.1.3",
  "form": "우송대학교",
  "original_form": "우송대학교",
  "speaker_id": "SD2001969",
  "start": 20.89803,
  "end": 22.21305,
  "note": ""
}
```

대화의 맥락상
‘혹시 학교는 어느 대학교 다니고 있어요?지금’
이 발화의 발화자는 SD2001968로 보이지만
SD2001969로 작성되어 있음

그림 2 일상 대화 원시 말뭉치의 자문자답 오류 데이터 예시

2) 발화자 ID 오류

자문자답 오류와 비슷한 경우로, 발화자 ID가 잘못 기입되어 있는 오류이다. 아래 예시는 문서 앞부분에서는 포장마차에 가봤다고 말한 발화자의 ID(SD2100023)가 뒷부분에서는 포장마차에 가보지 못했다고 말하는 발화자의 ID(SD2100023)로 기입되어 있는 경우이다. 해당 부분은 국립국어원에서 음성 파일을 검토하여 오류임을 확인하고, 포장마차에 못 가봤다고 한 발화의 발화자 ID를 상대 발화자의 ID(SD2100024)로 수정하였다.

```
{
  "id": "SDRW2100000024.1.1.407",
  "form": "뭐 어.",
  "original_form": "뭐 어.",
  "speaker_id": "SD2100023",
  "start": "704.75900",
  "end": "706.21669",
  "note": "발화검침"
},
{
  "id": "SDRW2100000024.1.1.408",
  "form": "뭐지?",
  "original_form": "뭐지?",
  "speaker_id": "SD2100024",
  "start": "707.22903",
  "end": "709.66899",
  "note": ""
},
{
  "id": "SDRW2100000024.1.1.409",
  "form": "아 나 가봤지롱.",
  "original_form": "아 나 가봤지롱.",
  "speaker_id": "SD2100023",
  "start": "707.60746",
  "end": "709.12669",
  "note": "발화검침"
},
{
  "id": "SDRW2100000024.1.1.410",
  "form": "나 서면에 가 봤지.",
  "original_form": "나 서면에 가 봤지.",
  "speaker_id": "SD2100023",
  "start": "709.77285",
  "end": "711.47902",
  "note": ""
},
{
  "id": "SDRW2100000024.1.1.422",
  "form": "어 그래 서면 서면",
  "original_form": "어 그래 서면 서면",
  "speaker_id": "SD2100023",
  "start": "728.22902",
  "end": "729.61897",
  "note": "발화검침"
},
{
  "id": "SDRW2100000024.1.1.423",
  "form": "어어어 거기",
  "original_form": "어어어 거기",
  "speaker_id": "SD2100024",
  "start": "728.74508",
  "end": "730.80588",
  "note": "발화검침"
},
{
  "id": "SDRW2100000024.1.1.424",
  "form": "나 그 한 번 포장마차를 안 가봤어 가지고",
  "original_form": "나 거 한 번 포장마차를 안 가봤어 가지고",
  "speaker_id": "SD2100023",
  "start": "729.61897",
  "end": "732.13550",
  "note": "발화검침"
},
{
  "id": "SDRW2100000024.1.1.425",
  "form": "너무 너무 가 보고 싶네.",
  "original_form": "너무 너무 가 보고 싶네.",
  "speaker_id": "SD2100024",
  "start": "730.80588",
  "end": "733.60902",
  "note": "발화검침"
}
```

그림 3 일상 대화 원시 말뭉치의 발화자 ID 오류 데이터 예시

3) 세부 주제(Topic) 오류

제시된 세부 주제(Topic)와 상관없는 발화가 포함된 데이터 오류로 국립국어원의 확인을 통해 세부 주제를 수정 또는 삭제 후 진행하였다. 아래 예시는 ‘쇼핑>휴대폰과 케이스 얘기’가 세부 주제인데 먹거리에 대한 발화가 포함되어 있는 경우이다. 일상 대화 말뭉치는 화자들에게 세부 주제를 선정하여 대화를 나누게 하지만, 화자들이 자유롭게 대화를 나눈 결과물을 전사한 자료이기 때문에 세부 주제에 해당하지 않는 내용이 포함될 수 있다.

```

{
  "document": [
    {
      "id": "SDRW2100003810.1.1",
      "metadata": {
        "title": "4인 일상 대화",
        "author": "개인 발화자",
        "publisher": "개인 발화 녹음",
        "date": "20220121",
        "topic": "쇼핑 > 휴대폰과 케이스 얘기",
        "speaker": [
          {
            "id": "SDRW2100003810.1.1.34",
            "form": "어릴 때는 세 개 1000원이었는데",
            "original_form": "어릴 때는 세 개 천 원이었는데",
            "speaker_id": "SD2102074",
            "start": "76.15000",
            "end": "78.38000",
            "note": ""
          },
          {
            "id": "SDRW2100003810.1.1.35",
            "form": "이제 두 개로 줄어서 너무 슬퍼.",
            "original_form": "이제 두 개로 줄어서 너무 슬퍼.",
            "speaker_id": "SD2102074",
            "start": "78.56000",
            "end": "80.43000",
            "note": ""
          },
          {
            "id": "SDRW2100003810.1.1.36",
            "form": "아 근데 아까 거기 1943 앞에 호떡 팔 거든.",
            "original_form": "아 근데 아까 거기 일구사삼 앞에 호떡 팔 거든.",
            "speaker_id": "SD2102072",
            "start": "82.19000",
            "end": "85.15000",
            "note": ""
          },
          {
            "id": "SDRW2100003810.1.1.37",
            "form": "그게 진짜 맛있으니까 한번 먹어보고",
            "original_form": "그게 진짜 맛있으니까 한번 먹어보고",
            "speaker_id": "SD2102072",
            "start": "85.15000",
            "end": "87.20000",
            "note": ""
          }
        ]
      }
    },
    {
      "id": "SDRW2100003810.1.1.28",
      "form": "아까 name2이가 봉어빵 사 줬잖아.",
      "original_form": "아까 &name2&이가 봉어빵 사 줬잖아.",
      "speaker_id": "SD2102074",
      "start": "62.90000",
      "end": "65.98000",
      "note": ""
    },
    {
      "id": "SDRW2100003810.1.1.29",
      "form": "그 너무 맛있었고",
      "original_form": "그 너무 맛있었고",
      "speaker_id": "SD2102074",
      "start": "66.36000",
      "end": "67.73000",
      "note": ""
    }
  ]
}

```

그림 4 일상 대화 원시 말뭉치의 세부 주제(Topic) 오류 데이터 예시

□ 일상 대화 원시 말뭉치의 오류 데이터 통계

일상 대화 원시 말뭉치 3,000건을 분석하며 발견한 오류 데이터의 비율은 2.13%로 64건의 원시 말뭉치를 수정 및 교체하였다.

구분	데이터	확인 및 수정 기간
확인한 데이터	약 3,000건	8월~11월
확인 요청한 데이터	73건	
오류 데이터	64건	
비율	2.13%	

표 7 일상 대화 원시 말뭉치 오류 데이터 현황표

□ 일상 대화 원시 말뭉치의 오류 데이터 대응 방법

작업자와 검수자, 관리자가 작업을 수행하며 발견한 오류 데이터는 국립국어원에 전달해 확인을 요청하였다. 이후 국립국어원으로부터 수정된 데이터와 검토 의견을 받아 수정한 파일로 교체 후 분석 작업을 진행하였다. 테디썸과 국립국어원의 대응 과정을 정리하면 다음과 같다.

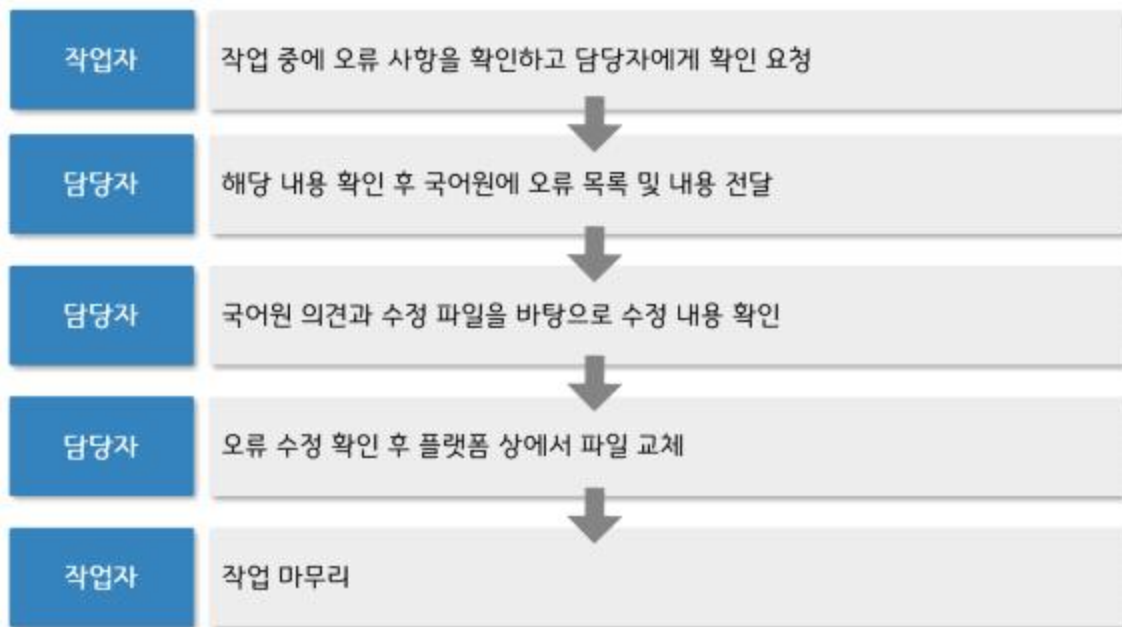


그림 5 테디썸의 오류 데이터 대응 과정

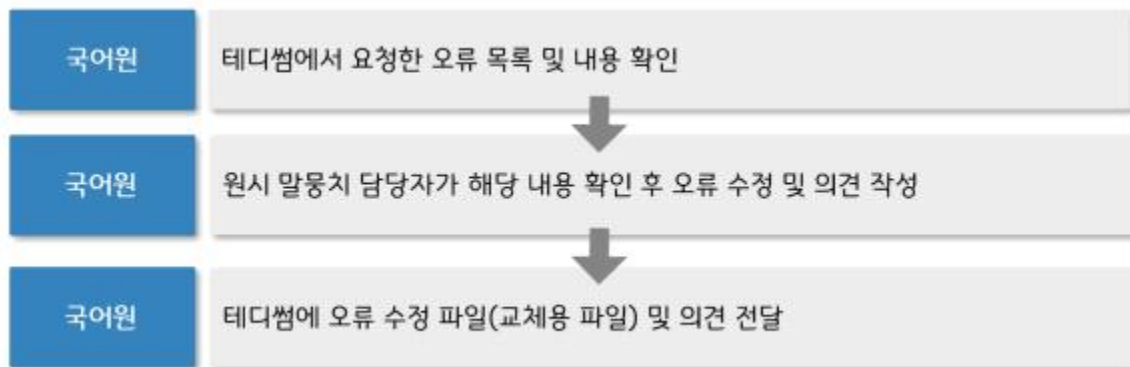


그림 6 국립국어원의 오류 데이터 대응 과정

사업 담당자와 테디썸 내부 인력이 오류 내용과 수정 사항 및 국립국어원의 검토 의견을 바탕으로 교차 확인하여 오류 파일을 교체한 후 작업을 수행하였다. 오류 데이터 관리 예시는 아래와 같다.

순번	파일 목록	오류 내용	확인요청	현황	참고 발화 번호	관리자 확인
1	SDRW2000000076	자문자답	8/16	교체 완료	-	○
2	SDRW2000000290	자문자답	8/28	교체 완료	529~	○
3	SDRW2000000295	토픽-내용 불일치	8/30	교체 완료		○
4	SDRW2000000296	토픽-내용 불일치	8/30	교체 완료		○
5	SDRW2000001008	자문자답	9/22	교체 완료	170~201	○
6	SDRW2000001212	자문자답	9/22	확인 요망 재확인 요청 완료 (10/12) 확인 요망(10/16) 교체 완료	198~	재확인 필요.. 198번 관련 자문자답 수정이 안된것같은 ○
7	SDRW2000000648	토픽-내용 불일치	9/22	교체 완료	스포츠/레저 > 미국, 싱가포르, 서울, 가족	○
8	SDRW2000001706	화자바뀜	10/10	교체 완료	발화 202~232 경찰대를 가고싶은 화자는 SD2002084인데 경찰대 질문을 다른 발화자에게 하고 있음 (228번 발화에서 084 화자가 다른 화자에게 경찰대 가기위한 준비과정을 질문함)	○
9	SDRW2100000018	토픽오류	10/16	교체 완료	아침운동권유라고 토픽이 설정되어 있으나 건강을 위한 음식섭취, 영양제 등의 대화가 주가로 있어 토픽 변경 필요	○
10	SDRW2100000024	화자바뀜	10/30	교체 완료	406 발화부터 시작되는 포장마차 관련 대화에서 SD2100023이 포장마차를 가보았다고 했는데 424 발화에서 가보지 않았다고 하고 있으므로 발화자가바뀐 것으로 보임	○

그림 7 테디섬의 오류 데이터 관리 예시

□ 원시 말뭉치 데이터 형식

일상 대화 원시 말뭉치는 국립국어원에서 제공한 “국립국어원 일상 대화 말뭉치 2020, 2021”을 활용하였다. 표 8은 일상 대화 원시 말뭉치의 JSON 구조를 설명한 것이다.

1수준	2 수준	3 수준	4 수준	설명
id				* 말뭉치 파일 ID
metadata				* 파일의 메타 정보
	title			* 원시 말뭉치: 국립국어원 [말뭉치 유형 구분] [파일 ID] 예) 국립국어원 구어 말뭉치 SDRW2100000002
	creator			* 생성자: 국립국어원
	distributor			* 배포자: 국립국어원
	year			* 말뭉치 구축 연도 예) 2020, 2021
	category			* 분류 예) 구어 > 사적 대화 > 일상대화
	annotation_level			* 분석 층위 예) 원시
	sampling			* 샘플링 방식 (본문 전체/ 부분 추출 - 임의 추출 / 부분 추출 - 특정 부분 추출 등)
document				* 대화 정보
	id			* 대화 ID * '원시 말뭉치 파일 ID.파일 내 문서 번호'로 구성 - SBRW2100000001.1
	metadata			* 대화의 메타 정보
		title		* 대화 제목 - 2인 일상 대화
		author		* 저작(권)자 - 개인 발화자
		publisher		* 발행자 - 개인 발화 녹음
		date		* 녹음일자 * 형식: YYYYMMDD - 20210805
		topic		* 주제 - 먹거리 > 저녁 음식 선택
		speaker		* 발화자 정보
			id	* 발화자 ID - SD2100001
			age	* 나이 - 20대
			occupation	* 직업
			sex	* 성별: 남성/여성
			birthplace	* 출생지
			principal_residence	* 주 성장지
			current_residence	* 현 거주지
			education	* [구어-일상 대화] 최종 학력: 초졸 이하, 중졸, 고졸, 대졸, 대재, 대학원 이상
		setting		* 환경 정보

			relation	* 화자 간 관계: 친구, 부부...
	utterance			* 발화 정보
		id		* 발화 ID
		form		* 발화 내용 정제 형태, 철자 전사
		original_form		* 원문 표기된 그대로의 형태, 발음 전사
		speaker_id		* 발화자 ID
		start		* [구어-일상 대화] 발화 시작 시간
		end		* [구어-일상 대화] 발화 종료 시간
		note		* 전사자 기타 메모

표 8 일상 대화 원시 말뭉치의 JSON 구조

아래 예시는 본 사업에서 실제로 분석한 문서의 원시 말뭉치 일부로 표 9에서 데이터 형식을 확인할 수 있다.

```
{
  "id": "SDRW2100000024",
  "metadata": {
    "title": "국립국어원 구어 말뭉치 SDRW2100000024",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2021",
    "category": "구어 > 사적대화 > 일상 대화",
    "annotation_level": [
      "원시"
    ],
    "sampling": "본문 전체"
  },
  "document": [
    {
      "id": "SDRW2100000024.1",
      "metadata": {
        "title": "2인 일상 대화",
        "author": "개인 발화자",
        "publisher": "개인 발화 녹음",
        "date": "20210810",
        "topic": "휴가 > 휴가지 선택",
        "speaker": [
          {
            "id": "SD2100023",
```

```

        "age": "40대",
        "occupation": "사무 종사자",
        "sex": "여성",
        "birthplace": "서울",
        "principal_residence": "서울",
        "current_residence": "서울",
        "education": "대졸"
    },
    {
        "id": "SD2100024",
        "age": "30대",
        "occupation": "주부",
        "sex": "여성",
        "birthplace": "서울",
        "principal_residence": "서울",
        "current_residence": "서울",
        "education": "대졸"
    }
],
"setting": {
    "relation": "형 제/자매"
}
},
"utterance": [
    {
        "id": "SDRW2100000024.1.1.1",
        "form": "지난번에 파주 갔다 왔잖아 언니.",
        "original_form": "지난번에 파주 갔다 왔잖아 언니.",
        "speaker_id": "SD2100024",
        "start": "0.98000",
        "end": "3.33900",
        "note": ""
    },
    {
        "id": "SDRW2100000024.1.1.2",
        "form": "어.",
        "original_form": "어.",
        "speaker_id": "SD2100023",
        "start": "2.78900",

```

```

        "end": "3.24900",
        "note": "발화접침"
    },
    {
        "id": "SDRW2100000024.1.1.3",
        "form": "그런 테가 난 가고 싶어.",
        "original_form": "그런 테가 난 가고 싶어.",
        "speaker_id": "SD2100024",
        "start": "3.98900",
        "end": "5.69000",
        "note": ""
    },
    {
        "id": "SDRW2100000024.1.1.4",
        "form": "그러니까 그냥 독채로 있고 대신에 풀 빌라가 있어야 돼.",
        "original_form": "그니까 그냥 독채로 있고 대신에 풀 빌라가 있어야 돼.",
        "speaker_id": "SD2100024",
        "start": "6.37000",
        "end": "9.72600",
        "note": ""
    },
    {
        "id": "SDRW2100000024.1.1.5",
        "form": "너무 좋지.",
        "original_form": "너무 좋지.",
        "speaker_id": "SD2100023",
        "start": "6.50000",
        "end": "7.44000",
        "note": "발화접침"
    },

```

표 9 일상 대화 원시 말뭉치 JSON 형식의 예

2.1.3 일상 대화 요약 분석 말뭉치 구축 분석 지침

본 사업에서는 효율적이고 일관성 있는 데이터 구축을 위한 ‘일상 대화 요약 분석 지침’을 신규 수립하였다. 해당 지침은 많은 작업자가 지침의 내용을 쉽게 이해하고 데이터 구축 플랫폼에 빠르게 적응할 수 있도록 상세한 설명과 실제 작업 예시를 포함하고 있다. 지침의 상세한 내용은 ‘[붙임 1] 일상 대화 요약 분석 말뭉치 구축 분석 지침’에서 확인할 수 있다.

2.1.4 일상 대화 요약 분석 말뭉치 구축 작업자 선발 및 교육

□ 작업자 선발

일상 대화 요약은 작업의 난도가 높고 분량이 많다. 이를 위해 인공지능 학습용 데이터 구축 사업에 대해 충분히 이해하고 있는 경력자를 기준으로 채용 후 요약과 관련한 사전 테스트 및 면접을 통해 적합한 인원을 선발하였다. 작업 인력의 개요는 다음과 같다.

구분	작업자 구분	전공	이력	비고
소주제 분할	작업자 1	융합예술과	이미지 데이터 라벨링 음성 데이터 태깅 논문 표그림 매칭태깅 작업	2022 회의록 요약 작업자
	작업자 2	교육리더십과 학교변화전공	경기도교육연구원 프리랜서 (통계분석, 교정 및 교열) M2E미디어웍스 프리랜서 (데이터 검수 및 업데이트)	
	작업자 3	경영학과	줌 인터넷, 공간정보산업진흥원, 두비랩, 스피링크	
소주제 분할 및 요약	작업자 4	회계학과	이미지 데이터 구축 업무	신규 작업자
	작업자 5	생명과학, 경영학	대전 카이스트 내 스타트업 데이터 구축 업무	
	작업자 6	-	클라우드웍스 교육 수료 라벨링 실습	
요약	작업자 7	-	법률데이터레이블링 한국어강의 음성 정제 및 전사 작업	
	작업자 8	불어불문학	AI 빅데이터 구축사업 참여 의료데이터 구축사업 음성데이터 녹음 가공 및 검수 등	
	작업자 9	전자공학과	데이터라벨링 이미지, 텍스트 다양한 경력 보유	
	작업자 10	문화인류학, 신문방송학	음성데이터 평가 프로젝트 (검수 및 수정), 대화문 오류 유형 분류, 영문서류 ocr 경험 유	
	작업자 11	영어영문학과	인공지능 학습용 챗봇 대화문	

			생성 및 검수, 요약문 작성 AI 음성데이터 검수 등 경력 다수	
	작업자 12	-	한국어 혐오 라벨링 & 한국어 혐오 공격, 챗봇QA, MTS 시제성라벨링 (확실성, 불확실성 & 과거, 현재, 미래), 영상라벨링	

표 10 일상 대화 요약 분석 작업자 개요

□ 작업자 교육

작업자들의 작업 교육을 위해 사업 담당자가 실제데이터를 바탕으로 샘플 데이터를 작성하여 데이터 구축 플랫폼에서 언제든지 해당 내용을 확인하면서 작업할 수 있도록 하였다. 또한 각 과제를 수행하기 전에 작업별(소주제 분할, 요약), 등급별(검수자, 작업자) 교육을 진행하였으며 작업에 어려움을 느끼는 작업자의 경우 개별적으로 소통하여 작업 지침 내용을 이해할 수 있도록 하였다. 중간에 작업자가 변경이 된 경우에도 바로 작업에 적응할 수 있도록 별도의 교육을 진행하였다.

데이터 가공 방법을 알기 쉽게 설명한 작업 지침 문서를 데이터 구축 플랫폼에서 언제든지 확인할 수 있도록 다운로드 기능을 추가하였고 작업 지침이나 내용에 대해 학습이 어려울 경우 참고할 수 있도록 작업 예시를 제공하였다. 작업을 수행하면서 지침이나 예시로 판단하기 어려울 때는 실시간으로 질문을 작성하고 답변받을 수 있도록 플랫폼 내 질의응답 창구와 단체 SNS를 운영하였다. 질의응답의 내용을 모든 작업자가 확인할 수 있도록 공개 및 공지하였으며 개별 면담과 데이터 분석 결과에 대한 지속적인 피드백을 수행하였다. 사업 기간 내 수행한 작업자 교육 일정은 다음과 같다.

일정	작업 교육 내용	비고
7/3	소주제 분할 OT	전원 참가
8/1	요약 OT	전원 참가
8/4	중간투입 작업자(황*주) 요약 OT	*작업자 변경 곽*미 -> 황*주
8/16	중간점검 OT(작업량 점검 등)	전원 참가
10/16	중간점검 OT(작업량 점검 등)	전원 참가

11/13	마지막 OT (작업량 점검 등)	전원 참가
-------	----------------------	-------

표 11 일상 대화 요약 작업자 교육 일정

작업자 교육은 그림 8과 같이 주로 비대면으로 이루어졌으며 작업 시작 전 진행한 작업 지침 교육을 제외하면 대부분 작업 관련 질의응답, 자주 발견되는 오류나 효율적인 작업 방법 등에 대한 교육을 진행하였다.

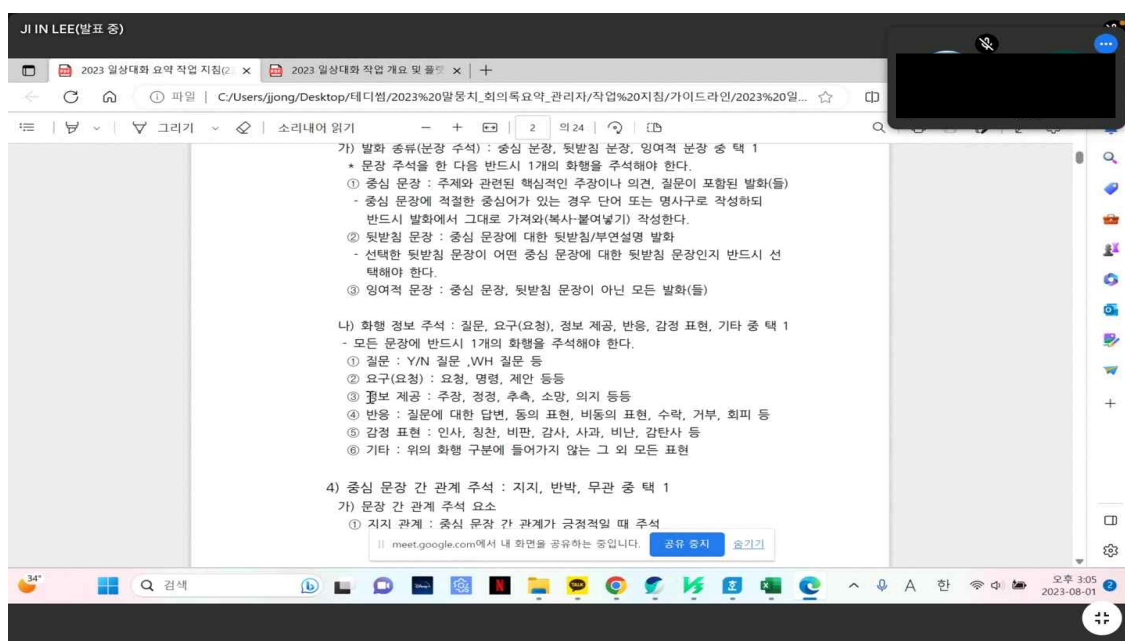


그림 8 일상 대화 요약 작업자 온라인 교육 화면

2.1.5 일상 대화 요약 분석 말뭉치 구축 플랫폼 개발

본 사업팀은 일상 대화뿐만 아니라 회의록 요약 분석 말뭉치를 함께 구축 및 관리하기 위해 하나의 플랫폼에서 데이터 구축, 검수 등을 수행할 수 있는 일체형 시스템을 개발하였다.

관리자가 등급별(작업자, 검수자 등) 계정을 등록하고 작업별 권한을 부여하면 그림 9와 같이 해당 계정에 접근이 허용된 작업의 목록이 뜨는 것을 확인할 수 있다.



그림 9 대화 요약 분석 말뭉치 구축 플랫폼 화면

표 12에서 확인할 수 있듯 대화 요약 분석 말뭉치 구축 플랫폼은 하나의 플랫폼 내에서 여러 사업 및 작업을 동시에 관리할 수 있도록 개발하였다. 작업자가 작업을 완료하면 검수자가 확인 후 검수 의견을 작성하여 작업자에게 전달하고, 검수 의견을 바탕으로 수정한다. 수정된 데이터를 한 번 더 검수하여 '최종 완료' 처리를 하게 되는 모든 과정을 하나의 플랫폼에서 수행할 수 있다. 또한 해당 플랫폼에서는 작업별, 작업자별 작업 현황을 관리하고 작업별 작업 지침도 내려받을 수 있으며 작업과 관련한 질의응답을 실시간으로 주고받을 수 있다.



표 12 대화 요약 분석 말뭉치 구축 플랫폼 개요

□ 소주제 분할 기능

본 사업에는 일상 대화 문서 전체에 대한 요약문 이외에도 각 소주제에 대한 소주제별 요약문을 작성하는 작업이 포함되어 있다. 따라서 소주제에 따라 데이터를 분할하는 작업이 선행되어야 한다. 이를 위해 본 사업에서는 기존의 요약 분석 플랫폼에 일상 대화의 발화를 소주제에 맞게 선택하여 분할할 수 있는 기능을 추가 적용하였다.

그림 10은 소주제에 맞는 발화를 선택하는 기능, 그림 11은 소주제 분할 후 수정이 필요한 경우 검수자의 의견을 받아볼 수 있는 노트(note) 기능에 대한 화면이다.

- 원문 전체를 읽어본 후 하나의 주제로 묶을 수 있는 발화 선택
 - 2개 이상의 발화 선택시 Shift 키를 이용하여 첫 발화와 마지막 발화로 묶어서 선택 가능

이 화면은 '작업 예시'로 검수 note에 구분한 주제를 작성하여 작업자가 참고할 수 있도록 함

그림 10 일상 대화의 소주제 분할을 위해 발화를 선택하는 기능



그림 11 일상 대화의 소주제 분할 작업 시 검수자의 피드백을 받을 수 있는 기능

□ 중요 표현 추출 및 연결 기능

일상 대화 요약 분석의 중요 표현 추출 및 연결 기능은 총 네 가지로 중심어 작성, 문장 주석, 화행 유형 주석, 중심 문장 간 관계 주석으로 구분할 수 있다. 각 기능에 대한 상세한 내용은 다음과 같다.

1) 중심어 작성

중심어는 발화자의 발화에서 소주제를 가장 잘 설명할 수 있는 단어(핵심 어휘) 혹은 명사구(최대 3어절)로 소주제별로 1개 이상 작성하여야 한다. 이를 위해 소주제별 중심어가 1개 미만인 경우 다음 작업으로 넘어가지 않도록 그림 12와 같이 경고 알림창을 띄우는 기능을 추가 적용하였다. 또 중심어는 요약자의 임의적 판단에 의한 수정은 허용하지 않으며, 반드시 실제 발화(중심 문장)에서 그대로 추출하여 입력할 수 있는 단어 혹은 명사구만 허용하기 때문에 그림 13과 같이 '유효성 검증' 기능을 통해 원문의 내용과 다른 중심어를 작성하는 경우 작성 및 저장이 되지 않도록 하였다.



그림 12 일상 대화의 중심어 경고 알림창 기능



그림 13 일상 대화의 요약 분석을 위한 중심어 작성 기능

2) 문장 주석

일상 대화 원시 말뭉치의 발화를 소주제와 관련된 핵심 발화와 그렇지 않은 발화로 구분하여 분석하기 위해 표 13의 세 가지 유형으로 나누어 주석하도록 하였다.

또 문장 주석이 누락되는 일이 없도록 문장 주석을 하지 않은 경우 그림 14와 같이 경고 알림창이 뜨는 기능을 추가하였다. 그림 15, 16은 동일한 유형으로 주석할 발화를 선택한 후 기준에 맞게 문장을 주석하는 기능을 보여주는 화면이다.

문장 주석	내용
중심 문장	소주제와 관련된 핵심적인 주장이나 의견, 질문이 포함된 발화(들)
뒷받침 문장	중심 문장에 대한 뒷받침/부연설명 발화
잉여적 문장	중심 문장, 뒷받침 문장이 아닌 모든 발화(들)

표 13 일상 대화 요약 분석의 문장 주석 기준



그림 14 일상 대화의 문장 유형 경고 알림창



그림 15 일상 대화의 요약 분석을 위한 문장 주석 기능 1



그림 16 일상 대화의 요약 분석을 위한 문장 주석 기능 2

뒷받침 문장으로 주석하는 경우 반드시 어떤 중심 문장에 대한 뒷받침 문장(부연 설명)인지를 선택하여야 한다. 이를 위해 뒷받침 문장으로 주석 후 중심 문장 번호를 입력하지 않거나 그림 17과 같이 '유효성 검증'을 하지 않는 경우 그림 18과 같은 경고 알림창이 뜨고 다음 작업으로 넘어가지 않는 기능을 추가하였다.



그림 17 일상 대화의 요약 분석을 위한 뒷받침-중심 문장 연결 기능

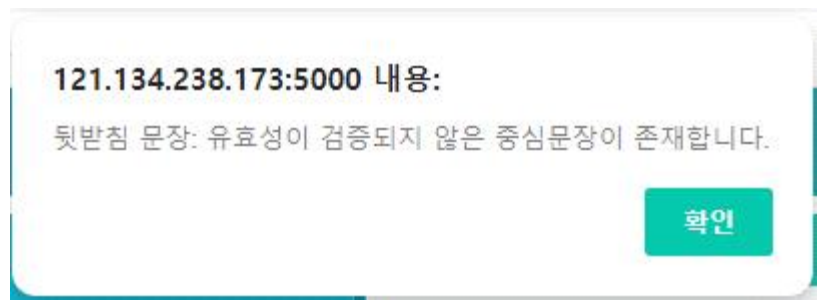


그림 18 일상 대화의 뒷받침-중심 문장 연결 경고 알림창

3) 화행 유형 주석

문장 주석을 한 후에는 반드시 화행 유형을 주석하도록 되어 있으며 화행 유형 주석 기준은 표 14와 같이 여섯 가지로 나뉜다. 화행 유형 주석 누락 방지를 위해 화행 유형을 주석하지 않고 다음 작업으로 넘어가려고 하는 경우 그림 20과 같은 경고 알림창이 뜨고 저장이 되지 않는 기능을 적용하였다.

화행 유형	내용
질문	예/아니오 질문, 의문사 포함 질문 등
요구(요청)	요청, 명령, 제안 등
정보 제공	주장, 정정, 추측, 소망, 의지 등
반응	질문에 대한 답변, 동의 표현, 비동의 표현, 수락, 거부, 회피 등
감정 표현	인사, 칭찬, 비판, 감사, 사과, 비난, 감탄사 등
기타	위의 화행 구분에 들어가지 않는 그 외 모든 표현

표 14 일상 대화 요약 분석의 화행 유형 주석 기준



그림 19 일상 대화의 요약 분석을 위한 화행 유형 주석 기능



그림 20 일상 대화의 화행 유형 경고 알림창

4) 중심 문장 간 관계 주석 기능

문장 및 화행 유형 주석 후에는 그림 21과 같이 중심 문장 간의 관계를 주석하여야 한다. 관계 주석은 지지, 반박, 무관으로 나뉘며 취향이나 고민에 대한 내용이 많고 예/아니오 질문보다 주관적인 답변을 해야 하는 질문이 많은 일상 대화의 특성상 지지, 반박보다는 무관의 비율이 압도적으로 높다. 중심 문장 간 관계 주석 역시 누락 방지를 위해 관계 주석 누락 시에는 그림 22와 같은 경고 알림창이 뜨는 기능을 적용하였다.

관계 구분	내용
지지	지지, 선호, 인정, 공감 등
반박	반박, 거절 등
무관	관계가 성립하지 않는 경우

표 15 일상 대화 요약 분석의 중심 문장 간 관계 주석 기준



그림 21 일상 대화의 요약 분석을 위한 중심 문장 간 관계 주석 기능

121.134.238.173:5000 내용:

관계가 입력되지 않은 문장이 존재합니다.

확인

그림 22 일상 대화의 중심 문장 간 관계 주석 경고 알림창

□ 소주제별 요약문 작성 기능

소주제별 발화를 소주제와 관련된 핵심 문장(중심 문장)과 그에 대한 부연 설명(뒷받침 문장) 그리고 소주제와 관련이 없거나 불필요한 문장(잉여적 문장)으로 구분한 다음에는 중심 문장과 뒷받침 문장을 바탕으로 화자별 요약문과 대화 요약문(소주제별 요약문)을 작성하여야 한다. 효율적인 데이터 구축을 위해 그림 23과 같이 주석한 문장을 색으로 구분하여 표시하였으며 이를 토대로 화자별 요약문과 대화 요약문을 작성하였다.

또한 작업자가 작업을 완료한 후 수정이 필요하다고 판단되는 경우 검수자의 의견을 볼 수 있는 노트(note) 기능도 적용되어 있다.

teddysum | 2023대화요약 데이터 구축 플랫폼

작업 작업자집 Q&A 관리 피드백

주제별 요약문 작성 1 / 3

※원문의 정보를 최대한 포괄하여 작성하는 것이 목표이므로 대화의 내용을 충분히 담아낼 것

발화자	발화
SD2000079	normal 화가 보였는지 모르겠는데 화가 일하 전후 영웅레스로온 그날 열 차에서에서 노래를 듣기전후, 근데 열차 안에 영웅레스로온 노래를 시작했어. 모래 그날 화장은 개목의 모래로 화장을 적용한 것이지 모래로.

대화 요약

화자별 요약
SD2000079

화자별 요약
SD2000080

• 요약문 작성

- 대화 요약 : 대화에서 드러나는 주제와 그 주제에 대한 화자별 의견, 결정 사항을 명료하게 작성

- 화자별 요약 : 주제에 대한 화자별 의견을 근거 등을 포함하여 구체적으로 작성

- 결정 사항 : 대화를 통해 나온 결정 사항이 있는 경우 작성

*분홍색 바탕 : 중심 문장

노란색 바탕 : 뒷받침 문장

회색 바탕 : 잉여적 문장

본문 요약 관련 주석

작업자집

(다음) 발화 주석

그림 23 일상 대화의 요약 분석을 위한 소주제별 요약문 작성 기능

☐ 전체 요약문 작성 기능

소주제별 요약문 내의 ‘대화 요약’ 내용을 기계적으로 통합한 후 중복되거나 불필요한 내용은 삭제하고 문장을 다듬어 전체 요약문을 작성하는 상향식 요약문 작성 방식을 채택하였다. 본 사업에서는 데이터 분석 지침을 준수하여 요약문을 작성할 수 있도록 그림 24와 같이 좌측 화면에서 소주제별 요약문을 확인하면서 우측 화면에서 전체 요약문을 작성할 수 있는 기능을 제공하였다.

[illegible]

그림 24 일상 대화의 요약 분석을 위한 전체 요약문 작성 기능

☐ 검수 및 수정 기능

작업자가 작업 완료 파일을 제출하면 검수자가 분석 데이터의 내용을 확인한 후 수정이 필요 없는 경우에는 ‘최종 완료’ 처리한다. 수정이 필요한 경우에는 그림 25와 같이 검수 노트(note) 기능을 활용하여 어떤 부분에 왜 수정이 필요한지 작업자에게 알려줄 수 있도록 내용을 작성하여 ‘수정 필요’ 처리를 하면 작업자가 노트(note) 내용을 확인한 후 수정한다.

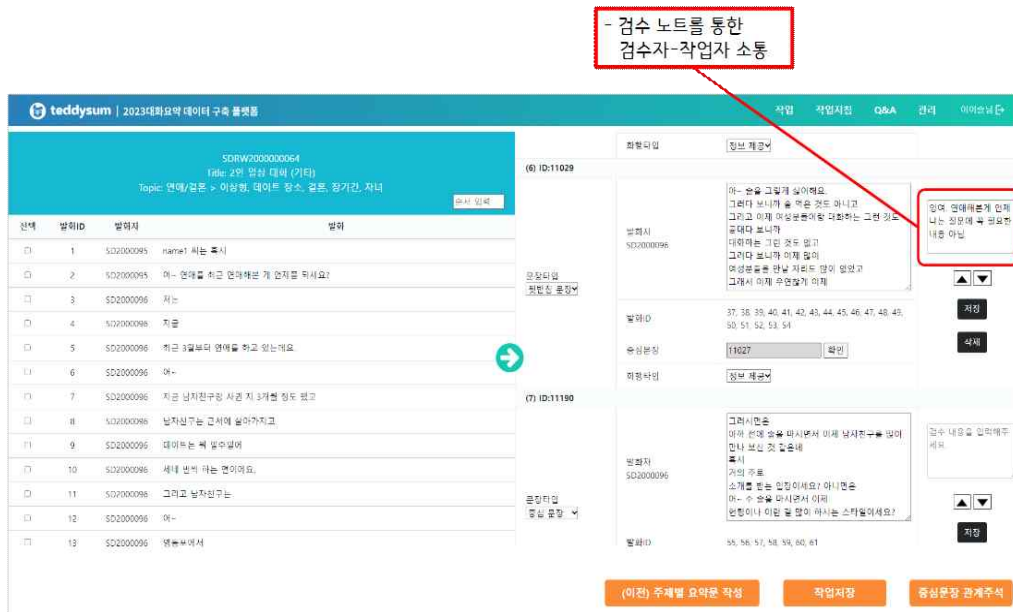
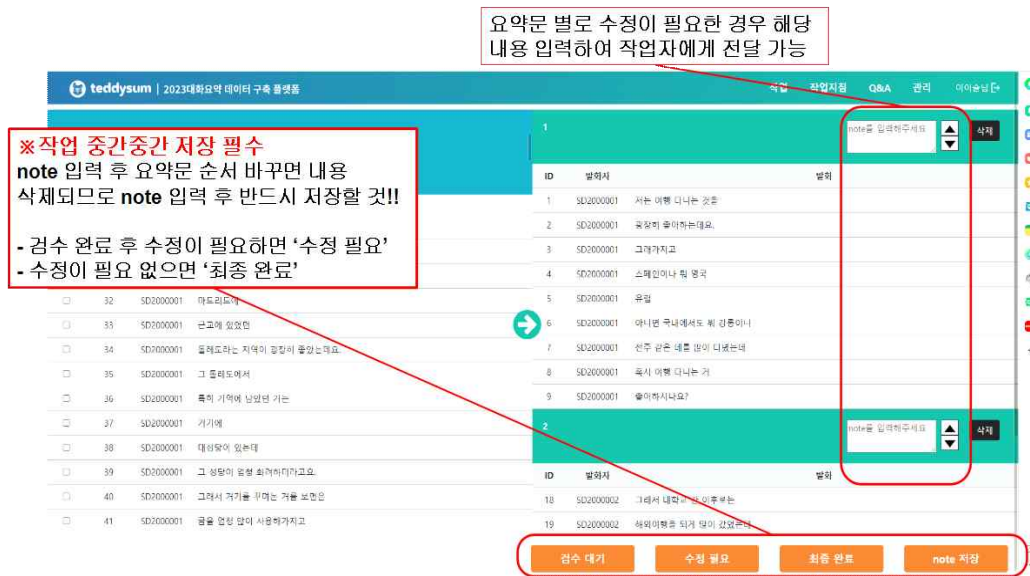


그림 25 일상 대화의 검수 노트(note) 기능

□ 관리 기능

본 사업팀은 제한된 일정과 예산에 따라 작업자들의 일정과 작업 현황을 관리할 수 있는 도구가 필요하다고 판단하였다. 따라서 개인별, 기간별 작업 현황과 주간, 전체 목표치를 확인하고 잔여 작업량을 파악할 수 있는 모니터링 페이지를 개발하였다. 이를 통해 작업자는 다른 작업자들을 포함한 본인의 작업 현황을 손쉽게 파악하여 작업 계획을 수립·수행할 수 있고 관리자는 전체 목표치와 개별 작업 현황을 확인함으로써 효율적인 관리가 가능하다.



그림 26 일상 대화 요약 작업자의 작업 현황 확인 기능



그림 27 일상 대화 요약 관리자의 작업 현황 확인 기능

□ 질의응답 기능

일상 대화 데이터 구축 플랫폼 내에 질의응답 창구를 운용하여 데이터 구축 과정에서 발생하는 질문을 플랫폼 내에서 실시간으로 문의하고 답변을 확인할 수 있는 기능을 개발하였다. 또한 작성되는 모든 질문과 답변을 공개함으로써 모든 작업자가 해당 내용을 확인하여 작업에 참고할 수 있도록 하였다.

2.1.6 일상 대화 요약 분석 말뭉치 구축 예시

일상 대화 요약은 일상 대화 요약 분석 지침을 바탕으로 이루어지며 소주제 분할, 중심표현 추출 및 연결(중심어 작성, 문장 주석, 화행 유형 주석, 중심 문장 간 관계 주석), 소주제별 요약문, 전체 요약문 작성을 수행한다. 그림 29~31은 요약 분석 지침을 토대로 작성한 일상 대화 요약 분석 말뭉치의 예시이다.

□ 소주제 분할

요약 작업을 수행하기 전 그림 29와 같이 발화의 내용에 따라 소주제 분할 작업을 수행한다. 아래 예시는 공통적으로 ‘취업’이라는 주제로 이야기하고 있지만 발화 내용에 따라 ‘간호사 취업 이후 부서 배정에 대한 실망과 적응’이라는 주제와 ‘오빠의 취업에 대한 걱정과 방향성 조언’이라는 소주제로 분할할 수 있다.

발화ID	발화	
SD2102304	name1이네	<div><간호사 취업 이후 부서 배정에 대한 실망과 적응>이라는 주제로 분할</div> <div><오빠의 취업에 대한 걱정과 방향성 조언>이라는 주제로 분할</div>
	간호사 취업했는데	
	만족하니?	
SD2102305	나는 원래 내가 제일 가고 싶었던 병원에 가서	
	만족은 하는데	
	막상 돼 보니까 병원이 시설이	
	좀 마음에 안 들고	
	좀 아쉽기도 하고	
(중략)		
SD2102305	무조건 일단은 다니기는 일단 오래 다닐 거고	
	뭐 어떤 부서를 뭐 받더라도 일단 열심히 공부해서 적응해 나가려고 노력해야지.	
SD2102304	오빠는 걱정이다 지금.	
	뭘 해야 할지도 모르겠고 희망도 없는 거 같고.	
	그래서 나는 걱정이 오빠다.	
	오빠는	
	경영학과 나와서 뭘 할지	
	어디로 갈지.	
	아니 너는 뭘 오빠야 어떤 거 어떻게 했으면 좋겠어?	
	뭐 어디에 어떻게 들어갔으면 좋을 거 같고 뭐	
	어떻게 했으면 좋겠는지 한번 생각해 줄래?	

그림 29 일상 대화 요약의 소주제 분할 예시

□ 중심어 선택, 발화 주석(문장 및 화행 유형 주석)

소주제 분할이 완료되면 소주제별로 분류된 발화에 대해 문장 주석(중심 문장, 뒷받침 문장, 잉여적 문장)과 화행 정보 주석(질문, 반응, 정보 제공 등)을 수행한다. 또 소주제에 맞는 중심어(그림 30의 ‘간호사 취업’)를 1개 이상 작성하고 뒷받침 문장으로 주석한 경우에는 어떤 중심 문장에 대한 뒷받침 문장(부연 설명)인지를 작성한다.

발화ID	발화	문장 주석	화행 정보	중심 문장 연결
SD2102304	간호사 취업했는데 만족하냐?	중심 문장1	질문	
SD2102305	나는 원래 내가 제일 가고 싶었던 병원에 돼서 만족은 하는데 막상 돼 보니까 병원이 시설이 좀 마음에 안 들고 좀 아쉽기도 하고	중심 문장2	반응	
	(중략)			
	여기는 그것이 안 되고 이제 환원으로 막 부수 부서를 막 배치한다는 거야. 그래서 나는 그것이 되게 실망스럽고 그래서 그것 때문에 내가 지금 고민을 좀 했었어.	뒷받침 문장	정보 제공	중심 문장 2에 대한 뒷받침 문장
	고민돼야 고민해도 지금 취업이 되어 있는데 부서 때문에 다른 쪽으로 뭐 갈 수도 있는 것도 아니고 이제 일단 적응을 좀 해 보다가 다른 쪽으로 한번 알아보든지 아니면 뭐 공무원 준비를 해 보든지 어떻게 해야 될 거 같고	뒷받침 문장	정보 제공	중심 문장 3에 대한 뒷받침 문장
		잉여적 문장	요구	
SD2102304				

그림 30 일상 대화 요약의 중심어 선택, 문장 및 화행 유형 주석 예시

□ 중심 문장 간 관계 주석

발화 주석 작업이 완료되면 그림 31과 같이 ‘중심 문장’으로 주석한 문장들의 관계를 ‘지지’, ‘반박’, ‘무관’ 중 하나로 주석한다.

별화ID	발화	문장 주석	중심 문장 관계
SD2102304	간호사 쉬엄했는데	중심 문장1	
	만족하니?		
SD2102305	나는 원래 내가 제일 가고 싶었던 병원에 돼서	중심 문장2	무관
	만족은 하는데		
	막상 돼 보니까 병원이 시설이		
	좀 마음에 안 들고		
	좀 아쉽기도 하고		
SD2102304	어 1년은 내가 볼 때는 좀	중심 문장3	무관
	버터 줬으면 좋겠는데 어떻게		
	어		
	네가 좀 찰할 수 있는지 한번 다른 길도 알아보고 인터넷도 좀 알아봤으면 좋겠어.		

그림 31 일상 대화 요약의 중심 문장 간 관계 주석 예시

□ 소주제별 요약문 작성

발화 주석과 중심 문장 간 관계 주석이 끝나면 소주제별 요약문을 작성한다. 소주제별 요약문은 ‘대화 요약’과 ‘화자별 요약’, ‘결정 사항’으로 구분되며 ‘대화 요약’에서의 화자별 요약 내용보다 ‘화자별 요약’의 내용을 더 길고 구체적으로 작성하여야 한다.

- 주제 : 간호사 취업 이후 부서 배정에 대한 실망과 적응

• ‘대화요약’의 화자별 요약보다 ‘화자별 요약’의 내용을 더 길고 구체적으로 작성

구분	내용
대화 요약	화자 SD2102304와 SD2102305는 SD2102305의 간호사 취업 이후 부서 배정에 대해 이야기했습니다. 부서가 랜덤으로 배정되어 실망했다는 SD2102305에게 SD2102304는 여러 가능성을 열어두되 1년 정도는 버티며 적응해볼 것을 조언했고 SD2102305 역시 이에 동의하며 열심히 적응해보겠다고 이야기했습니다.
SD2102304 요약	SD2102304는 이미 취업을 하기도 했고, 원하는 부서로의 이동도 어려우니 1년은 건디며 근무할 필요가 있다고 이야기하면서도 우선 노력해보다가 힘들면 공무원 시험이나 다른 길은 없는지 찾아볼 것을 제안했습니다.
SD2102305 요약	SD2102305는 취업한 병원의 시설이나 부서 랜덤 배정에 대해 실망스럽고 아쉬운 마음이 들었으며 원하는 부서로의 이동이 어려울 것 같아 걱정이지만, 어느 부서로 가더라도 1년은 다녀보면서 공부하면서 적응하겠다고 다짐했습니다.
결정 사항	(결정사항 없음)

- 주제 : 오빠의 취업에 대한 걱정과 방향성 조언

구분	내용
대화 요약	화자 SD2102304와 SD2102305는 현재 취업을 하지 못한 오빠에 대해 이야기했습니다. SD2102304가 오빠의 미취업 상태에 대해 걱정하며 어떤 방향으로 취업을 준비하면 좋을지 SD2102305에게 물었고, SD2102305는 취업에 필요한 것들을 이야기하며 무엇보다 자신이 무엇을 하고 싶은지 아는 것이 가장 중요하다고 조언했습니다.
SD2102304 요약	SD2102304는 오빠가 취업을 하지 않고 있는 점에 대해 고민하고 걱정하는 한편 취업을 위해 무엇이 필요할지 SD2102305에게 조언을 구했습니다.
SD2102305 요약	SD2102305는 토익 등 기본적인 자격증 취득을 제안하며 취업을 위해 필요한 방법들을 추천했고, 가장 중요한 것은 자기 자신에 대해 알고, 어떤 회사에 들어가고 싶은지 아는 것이라고 주장하며 이에 대해 오빠에게 직접 묻는 것이 좋겠다고 답변했습니다.
결정 사항	(결정사항 없음)

그림 32 일상 대화 요약의 소주제별 요약문 작성 예시

□ 전체 요약문 작성

소주제별 요약문 작성이 끝나면 일상 대화 문서 전체에 대한 전체 요약문을 작성한다. 전체 요약문은 소주제별 요약문 내의 ‘대화 요약’ 내용만을 기계적으로 통합한 후, 중복되거나 불필요한 부분을 삭제하고 자연스러운 문장이 되도록 수정한다.


구분	내용
주제1의 대화요약	화자 SD2102304와 SD2102305는 SD2102305의 간호사 취업 이후 부서 배정에 대해 이야기했습니다. 부서가 랜덤으로 배정되어 실망했다는 SD2102305에게 SD2102304는 여러 가능성을 열어두되 1년 정도는 버티며 적응해볼 것을 조언했고 SD2102305 역시 이에 동의하며 열심히 적응해보겠다고 이야기했습니다.
주제2의 대화요약	화자 SD2102304와 SD2102305는 현재 취업을 하지 못한 오빠에 대해 이야기했습니다. SD2102304가 오빠의 미취업 상태에 대해 걱정하며 어떤 방향으로 취업을 준비하면 좋을지 SD2102305에게 물었고, SD2102305는 취업에 필요한 것들을 이야기하며 무엇보다 자신이 무엇을 하고 싶은지 아는 것이 가장 중요하다고 조언했습니다.
 <div style="border: 1px dashed black; padding: 5px; display: inline-block;"> <ul style="list-style-type: none"> • 기계적 통합 후 중복내용, 불필요한 부분은 삭제하고 부자연스러운 부분이 없도록 수정하여 작성 </div>	
구분	내용
최종 요약문	이 대화에서는 간호사 취업 후의 고민과 오빠의 취업 문제에 대해 이야기했습니다. SD2102305와 SD2102304는 는 취업 이후 부서 랜덤 배정에 실망한 것에 대해 적응하고 버티볼 필요성에 대한 이야기를 나누었습니다. 또한 화자 SD2102304는 현재 취업을 하지 못한 오빠에 대해 걱정하며 조언을 구했으며 이에 대해 화자 SD2102305는 무엇보다 자신이 무얼 하고 싶은지 아는 것이 중요하다는 의견을 공유했습니다.

그림 33 일상 대화 요약의 전체 요약문 작성 예시

2.2 회의록 요약 분석 말뭉치 구축

2.2.1 회의록 요약 분석 말뭉치 작업 개요

□ 회의록 요약 분석 말뭉치 작업 개요

회의록 요약 작업은 크게 소주제 분할과 요약 분석으로 구분된다. 소주제 분할은 대화 내 소주제에 따라 발화를 분류하는 작업으로 소주제 분할이 완료되면 분할된 소주제를 기준으로 발화 주석, 요약문 작성 등 요약 분석을 수행한다. 회의록 요약 분석 말뭉치는 다음과 같은 요소로 구성된다.

- 1) 주제 분할 : 쟁점(법안 등)에 따른 발화 분류 수행
- 2) 요약 분석

가) 구분

- (1) 전체 요약문 : 하나의 회의록 전체에 대한 요약문
- (2) 주제(Topic) : 주제와 연관이 있는 심사된 안건
- (3) 중심어(Issue keyword) : 주제를 포괄적으로 나타낼 수 있는 단어 혹은 명사구
- (4) 발화 : 주제에 대해 논의한 발화

나) 발화 종류

- (1) 주제별 요약문(Abstract) : 논쟁적 대화 전체를 포괄하는 요약문
 - * 주제별로 반드시 1개의 주제별 요약문(Abstract)을 작성
- (2) 문제 제기(Problems/Issues) : 논쟁적 대화에서 논의의 대상이 되는 쟁점/문제/이슈
 - * 발화의 대부분을 차지함
- (3) 결정 사항(Decisions) : 논쟁적 대화에서 도출된 결정 사항
- (4) 후속 조치(Actions) : 논쟁적 대화를 통해 결정된 후속 조치
 - * 회의록에 따라 결정 사항이나 후속 조치가 없는 경우에는 결정 사항(Decisions), 후속 조치(Actions)는 작성하지 않음

□ 회의록 요약 분석 말뭉치 작업 과정

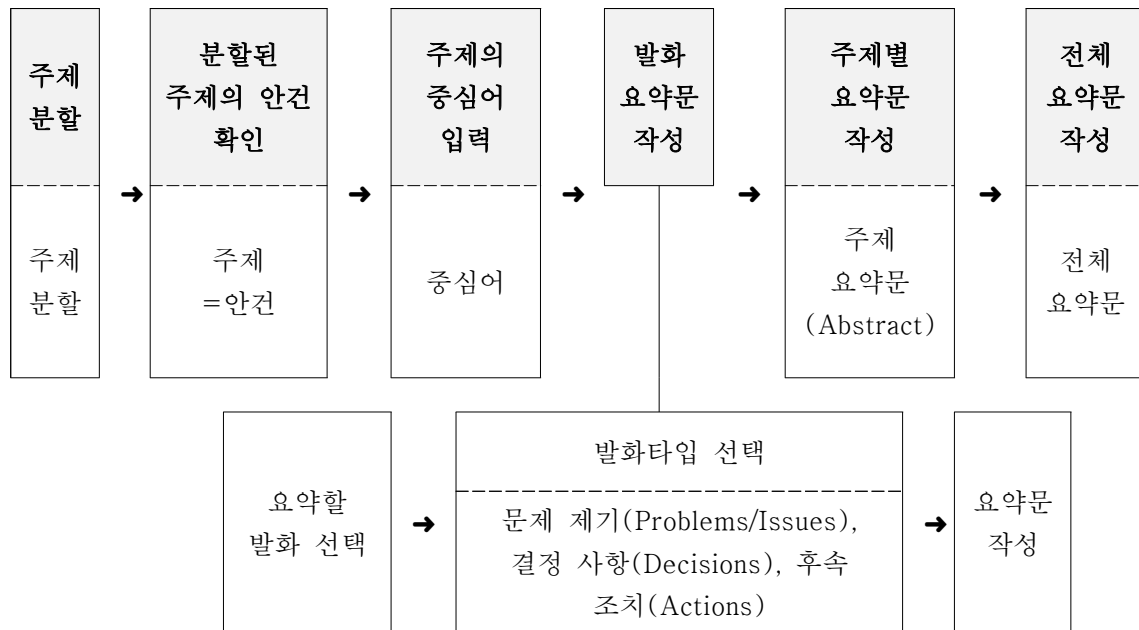


표 16 회의록 요약 작업 개요

- 1) 주제 분할 : 주제(안건 등)에 따른 발화 분류 수행
- 2) 요약 분석

가) 분할된 주제의 안건 확인

나) 중심어 작성 : 주제를 포괄적으로 나타낼 수 있는 중심어(단어 혹은 명사구)를 ‘복사-붙여넣기’ 하여 입력

다) 요약문 작성 : 주제 내에서 발화에 대한 요약문 작성

(1) 요약할 발화 선택

(2) 발화타입 선택 : 문제 제기(Problems/Issues), 결정 사항(Decisions), 후속 조치(Actions) 중 택 1

(3) 주제별 요약문(Abstract) 작성

라) 전체 요약문 작성 : 주제별 요약문을 참고하여 회의록 전체를 아우르는 요약문을 작성

2.2.2 회의록 요약 분석 말뭉치 선정 및 특징

□ 데이터 선정

국회 회의록 요약 분석 말뭉치는 ‘국립국어원 국회 회의록 말뭉치 2021’ 중 2021, 2022년에 구축한 요약문의 회의록 원문을 제외하고 평균 8,000어절, 500개 문장 내외 분량의 문서 300건을 분석 대상으로 한다. 분석 대상 회의록 목록은 표 17과 같다.

- 데이터 목록

회의록	빈도수
財政經濟委員會會議錄 (재정경제위원회회의록)	32
文化觀光委員會會議錄 (문화관광위원회회의록)	23
環境勞動委員會會議錄 (환경노동위원회회의록)	14
保健福祉委員會會議錄 (보건복지위원회회의록)	14
法制司法委員會會議錄 (법제사법위원회회의록)	14
政務委員會會議錄 (정부위원회회의록)	12
企劃財政委員會會議錄 (기획재정위원회회의록)	10
行政自治委員會會議錄 (행정자치위원회회의록)	9
教育委員會會議錄 (교육위원회회의록)	8
文化體育觀光放送通信委員會會議錄 (문화체육관광방송통신위원회회의록)	8
豫算決算特別委員會會議錄 (예산결산특별위원회회의록)	7
統一外交通商委員會會議錄 (통일외교통상위원회회의록)	7
國防委員會會議錄 (국방위원회회의록)	7

行政安全委員會會議錄 (행정안전위원회회의록)	6
司法制度改革特別委員會會議錄 (사법제도개혁특별위원회회의록)	6
교육문화체육관광위원회회의록	6
환경노동위원회회의록	6
헌법개정특별위원회회의록	6
國會運營委員會會議錄 (국회운영위원회회의록)	5
政治改革特別委員會會議錄 (정치개혁특별위원회회의록)	5
政治關係法特別委員會會議錄 (정치관계법특별위원회회의록)	5
기획재정위원회회의록	5
4차산업혁명특별위원회회의록	5
產業資源委員會會議錄 (산업자원위원회회의록)	4
建設交通委員會會議錄 (건설교통위원회회의록)	4
外交通商統一委員會會議錄 (외교통상통일위원회회의록)	4
농림축산식품해양수산위원회회의록	4
법제사법위원회회의록	4
科學技術情報通信委員會會議錄 (과학기술정보통신위원회회의록)	3
農林海洋水產委員會會議錄 (농림해양수산위원회회의록)	3
保健福祉家族委員會會議錄 (보건복지가족위원회회의록)	3
國土海洋委員會會議錄 (국토해양위원회회의록)	3
農林水產食品委員會會議錄 (농림수산식품위원회회의록)	3
보건복지위원회회의록	3
국토교통위원회회의록	3
창조경제활성화특별위원회회의록	3

국방위원회회의록	3
國會改革特別委員會會議錄 (국회개혁특별위원회회의록)	2
情報委員會會議錄 (정보위원회회의록)	2
女性家族委員會會議錄 (여성가족위원회회의록)	2
教育科學技術委員會會議錄 (교육과학기술위원회회의록)	2
知識經濟委員會會議錄 (지식경제위원회회의록)	2
예산결산특별위원회회의록	2
방송공정성특별위원회회의록	2
사법제도개혁특별위원회회의록	2
정치개혁특별위원회회의록	2
國際競技大會支援特別委員會會議錄 (국제경기대회지원특별위원회회의록)	1
國際競技大會 (국제경기대회)	1
氣候變化對策特別委員會會議錄 (기후변화대책특별위원회회의록)	1
일자리만들기特別委員會會議錄 (일자리만들기특별위원회회의록)	1
貯蓄銀行非理疑惑真相糾明을 위한國政調査特別 委員會會議錄 (저축은행비리의혹진상규명을위한국정조사특별 위원회회의록)	1
교육과학기술위원회회의록	1
지식경제위원회회의록	1
산업통상자원위원회회의록	1
군인권개선및병영문화혁신특별위원회회의록	1
공무원연금개혁특별위원회회의록	1
외교통일위원회회의록	1
안전행정위원회회의록	1

정무위원회회의록	1
과학기술정보방송통신위원회회의록	1
헌법개정및정치개혁특별위원회회의록	1

표 17 분석 대상 회의록 목록

□ 국회 회의록의 특징

국회 소위원회 회의는 소위원장의 진행 발언을 시작으로 전문위원의 보고, 정부 부처의 입장, 의원들의 질의응답, 주장 등을 주고받으며 쟁점에 대해 논의한 후 소위원장의 가결 유무 및 폐회 발언으로 마무리되는 것이 일반적이다. 이렇듯 국회 소위원회 회의록 데이터는 일반적인 뉴스나 도서 데이터와는 구조적인 차이가 있다. 일반적인 데이터의 요약 작업에서는 단일 문서의 핵심 내용을 간결하게 요약하는 것이 대부분이지만, 논쟁적 대화의 대표적인 사례인 국회 회의록은 시간순으로 발화된 내용을 담고 있기 때문에 각 쟁점을 단순히 몇 문장으로 축약하기 어려운 구조를 갖고 있다.

국회 회의록은 다양한 의견이 교차되고 발언 의도에 따라 의미가 달라지며 쟁점마다 여러 주장이 나타난다. 또한 법안 분석, 질의응답 등 복잡한 논변 구조를 갖고 있기 때문에 이에 대한 충분한 이해가 필요하다. 따라서 국회 회의록 요약 분석 말뭉치 구축을 위해서는 이러한 구조와 목적을 충분히 고려해야 하며 각 쟁점의 핵심 내용뿐만 아니라 문제 제기, 결정 사항, 후속 조치 등 다양한 요소를 포함하는 요약문을 구축할 필요가 있다.

□ 원시 말뭉치 데이터 형식

회의록 원시 말뭉치는 “국립국어원 국회 회의록 말뭉치 2021”을 사용하였다. 아래 표 18은 회의록 원시 말뭉치의 JSON 구조를 나타낸 것으로, 본 사업에서 실제로 분석한 문서의 일부인 원시 말뭉치의 데이터 예시 내용을 표 19에서 확인할 수 있다.

1수준	2 수준	3 수준	4 수준	설명
id				* 말뭉치 파일 ID * 고유 ID로 중복이 없어야 함. - SBRW2100000001
metadata				* 파일의 메타 정보
	title			* 원시 말뭉치: 국립국어원 [말뭉치 유형 구분] [파일 ID] 예) 국립국어원 국회 회의록 원시 말뭉치 SBRW2100000001
	creator			* 생성자: 국립국어원
	distributor			* 배포자: 국립국어원
	year			* 말뭉치 구축 연도 - 2021
	category			* 분류 - 회의록 > 국회소위원회 > 보건복지위원회회의록
	annotation_level			* 분석 층위 - 원시
	sampling			* 샘플링 방식 - 본문 전체
document				* 문서 정보
	id			* 문서 ID * '원시 말뭉치 파일 ID.파일 내 문서 번호'로 구성 - SBRW2100000001.1
	metadata			* 문서의 메타 정보
		title		* 문서 제목 - 제33회국회(임시회) 보건복지위원회회의록(법안심사소위원회) 제5호
		author		* 작성자, 게시자 등 저작(권)자 - 국회사무처
		publisher		* 출판사, 신문사, 방송사 등 - 국회사무처
		date		* 작성일시, 게시일시 * 형식: YYYYMMDD - 20150706
		topic		* 주제 - 회의록 말뭉치: <심사된 안건> 삽입 1. 2018년도 예산안(계속) > 가. 여성가족부 소관 / 2. 2018년도 기금운용계획안(계속) > 가. 양성평등기금, 나. 청소년육성기금
		speaker		* 발화자 정보 - 회의록 말뭉치: 회의 참여자 정보
			id	* 발화자 ID - 김상희
			occupation	* 직업 - 회의록에 기재된 직위/직책: 소위원장
	utterance			* 구어(준구어, 메신저 포함), 메신저 발화 정보
		id		* 발화 ID * '원시 말뭉치 파일 ID.1.1.발화 번호'로 구성 - SBRW2100000001.1.1.1 (document 내 첫 번째 발화) - SBRW2100000001.1.1.2 (document 내 두 번째 발화)
		form		* 발화 내용 정제 형태, 철자 전사 - 회의록: 정제 후(한자 변환, 서기 메모 삭제 등)
		original_form		* 원문 표기된 그대로의 형태, 발음 전사 - 회의록: 정제 전
		speaker_id		* 발화자 ID, 등장 인물 - 회의록: document > metadata > speaker > id 기입
		note		* 전사자 노트, 준구어의 경우 발화 외 지문 등

표 18 회의록 원시 말뭉치 JSON의 구조

```

{
  "id": "SBRW2100000009",
  "metadata": {
    "title": "국립국어원 국회 회의록 원시 말뭉치 SBRW2100000009",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2021",
    "category": "회의록 > 국회소위원회 > 국토교통위원회회의록",
    "annotation_level": ["원시"],
    "sampling": "본문 전체"
  },
  "document": {
    "id": "SBRW2100000009.1",
    "metadata": {
      "title": "제322회국회(임시회)국토교통위원회회의록(철도산업발전소위원회)제1호",
      "author": "국회사무처",
      "publisher": "국회사무처",
      "date": "20140207",
      "topic": "1. 정책자문협의체 구성의 건",
      "speaker": [
        {
          "id": "강석호",
          "occupation": "소위원장"
        },
        {
          "id": "김경욱",
          "occupation": "국토교통부철도국장"
        },
        {
          "id": "김영훈",
          "occupation": "철도노조지도위원"
        }
      ]
    },
    "utterance": [
      {
        "id": "SBRW2100000009.1.1.1",
        "form": "의석을 정돈해 주시기 바랍니다.",

```

```

        "original_form": "의석을 정돈해 주시기 바랍니다.",
        "speaker_id": "강석호",
        "note": "(10시20분 개의)"
    },
    {
        "id": "SBRW2100000009.1.1.2",
        "form": "성원이 되었으므로 제322회 국회(임시회) 국토교통위원회 제1차 철도산업발전소위원회를 개의합니다.",
        "original_form": "성원이 되었으므로 제322회 국회(임시회) 국토교통위원회 제1차 철도산업발전소위원회를 개의합니다.",
        "speaker_id": "강석호",
        "note": ""
    },
    {
        "id": "SBRW2100000009.1.1.3",
        "form": "바쁘신 와중에도 회의에 참석해 주신 여러 위원님들과 박종흠 교통물류실장과 또 김영래 철도공사 부사장을 비롯한 관계관 여러분께 감사의 말씀을 드립니다.",
        "original_form": "바쁘신 와중에도 회의에 참석해 주신 여러 위원님들과 박종흠 교통물류실장과 또 김영래 철도공사 부사장을 비롯한 관계관 여러분께 감사의 말씀을 드립니다.",
        "speaker_id": "강석호",
        "note": ""
    },
    {
        "id": "SBRW2100000009.1.1.4",
        "form": "오늘은 먼저 정책자문협의체 구성안을 의결한 다음에 국토교통부로부터 고속철도 운영 경쟁체제 도입에 따른 효율화 방안에 대해서 상세한 보고와 또 추가로 제출된 자료를 포함한 철도산업 발전방안에 대해서 여러 위원님들의 질의답변 시간을 갖는 순서로 하겠습니다.",
        "original_form": "오늘은 먼저 정책자문협의체 구성안을 의결한 다음에 국토교통부로부터 고속철도 운영 경쟁체제 도입에 따른 효율화 방안에 대해서 상세한 보고와 또 추가로 제출된 자료를 포함한 철도산업 발전방안에 대해서 여러 위원님들의 질의답변 시간을 갖는 순서로 하겠습니다.",
        "speaker_id": "강석호",
        "note": ""
    },
    {

```

표 19 회의록 원시 말뭉치 JSON 형식의 예

2.2.3 회의록 요약 분석 말뭉치 구축 작업자 선발 및 교육

□ 작업자 선발

회의록 요약 작업은 회의록 요약 분석 지침의 내용이 복잡하고 작업 분량도 많아 난도가 높은 편이다. 양질의 데이터 구축을 위해 2022 회의록 요약 분석 사업에 참여했던 작업자 중 성실하고 우수한 작업 능력을 발휘했던 작업자를 선별하였다. 작업 인력 개요는 표 20과 같다.

구분	작업자 구분	전공	이력	비고
주제분할 및 요약	작업자 1	일본어학	2022 회의록 요약 작업	2022 회의록 요약 작업자
요약	작업자 2	융합예술과	이미지 데이터 라벨링 음성 데이터 태깅 논문 표그림 매칭태깅 작업	
	작업자 3	교육리더십과 학교변화전공	경기도교육연구원 프리랜서 (통계분석, 교정 및 교열) M2E미디어웍스 프리랜서 (데이터 검수 및 업데이트)	
	작업자 4	경영학과	줌 인터넷, 공간정보산업진흥원, 누비랩, 스피링크	

표 20 회의록 요약 분석 작업자 개요

□ 작업자 교육

작업자 교육은 각 과제 수행 전 집단 OT를 통해 이루어졌으며 주제 분할과 회의록 요약에 대한 교육으로 나누어 진행하였다. 또한 실시간 온라인 질의응답 및 지속적인 피드백, 개별 면담을 통해 데이터 구축을 수행하였으며 작업자 교육 일정은 아래와 같다.

일정	국회 회의록	비고
7/3	주제분할 OT	전원 참가
8/1	요약 OT	전원 참가
8/16	중간점검 OT(작업량 점검 등)	전원 참가

10/16	중간점검 OT(작업량 점검 등)	전원 참가
11/13	마지막 OT (교차 검수 및 수정 안내 등)	전원 참가

표 21 회의록 요약 작업자 교육 일정

작업자 교육은 작업 지침 교육을 제외하면 자주 발견되는 오류, 검수 의견 등에 대한 수정 교육으로 진행되었으며 그림 34와 같이 주로 비대면으로 이루어졌다.

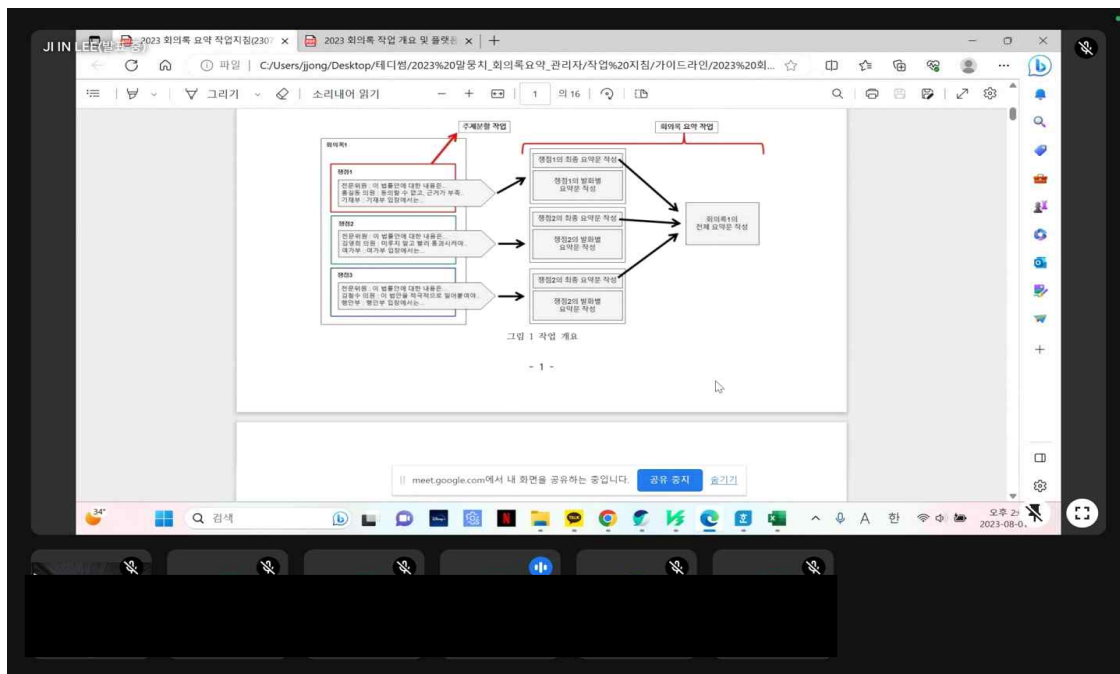




그림 34 회의록 요약 작업자 온라인 교육 화면

2.2.4 회의록 요약 분석 말뭉치 구축 플랫폼 개발

□ 주제 분할 기능

일반적인 요약 분석 말뭉치와는 다르게 회의록 요약 분석 말뭉치는 다양한 안건과 구조를 포함하고 있기 때문에 안건별로 쟁점을 요약하는 작업이 필요하다. 이를 위해 회의록 내용을 주제별로 나눌 수 있는 플랫폼을 개발함으로써 발화를 주제에 맞게 분할하고 안건과 연결하는 작업을 효율적으로 수행할 수 있다.

 | 회의록 요약 분석을 위한 데이터 구축 플랫폼

 **회의록 요약 분석 플랫폼** **작업현황** **작업** **작업지침** **Q&A**

Topic 중 선택

선택	발화ID	발화자	발화
<input type="checkbox"/>	1	김희정	의석을 정돈해 주시기 바랍니다.
<input checked="" type="checkbox"/>	2	김희정	성원이 되었으므로 제316회 국회(임시회) 제1차 법안심사소위원회를 개의하겠습니다.
<input checked="" type="checkbox"/>	3	김희정	의사일정에 들어가기 전에 법안심사소위원회 위원개선과 관련하여 말씀드리겠습니다.
<input checked="" type="checkbox"/>	4	김희정	여야 간사님들의 법안심사소위 위원 사보임 요청에 의하여 새누리당에서는 서상기 위원님이 사임하시고 대신에 박인숙 위원님이 보임되었습니다. 그리고 민주당 박혜자 위원님과 진보정의당 정진후 위원님이 사임하시고 대신에 이용섭 위원님과 윤관석 위원님이 보임되었다는 말씀을 드립니다.
<input checked="" type="checkbox"/>	5	김희정	간단하게 인사말씀 하시지요.
<input type="checkbox"/>	6	박인숙	둘락날락하다가 완전히 들어왔습니다. 열심히 하겠습니다.
<input type="checkbox"/>	7	김희정	이용섭 위원님.
<input type="checkbox"/>	8	이용섭	좋은 성과 내도록 최선을 다하겠습니다.

그림 35 회의록의 주제 분할을 위해 일부 발화를 선택하는 기능

title을 입력해주세요.

Topic 중 선택

☐ 1. 사립학교법 일부개정법률안(민병주 의원 대표발의)
☒ 2. 교육공무직원의 채용 및 처우에 관한 법률안(유기홍 의원 대표발의)(계속)
☒ 3. 학교직원의 채용 및 근무에 관한 법률안(이예리사 의원 대표발의)
☒ 4. 초·중등교육법 일부개정법률안(김희정 의원 대표발의)
☐ 5. 초·중등교육법 일부개정법률안(정진후 의원 대표발의)(의안번호 2415)(계속)
☐ 6. 초·중등교육법 일부개정법률안(박성호 의원 대표발의)
☐ 7. 초·중등교육법 일부개정법률안(신계륜 의원 대표발의)
☐ 8. 초·중등교육법 일부개정법률안(김상희 의원 대표발의)(의안번호 4403)
☐ 9. 초·중등교육법 일부개정법률안(김태년 의원 대표발의)
☐ 10. 초·중등교육법 일부개정법률안(이군현 의원 대표발의)(계속)
☐ 11. 초·중등교육법 일부개정법률안(정부 제출)(계속)
☐ 12. 초·중등교육법 일부개정법률안(이재영 의원 대표발의)(계속)
☐ 13. 초·중등교육법 일부개정법률안(박인숙 의원 대표발의)(계속)
☐ 14. 초·중등교육법 일부개정법률안(서상기 의원 대표발의)(계속)
☐ 15. 초·중등교육법 일부개정법률안(민병주 의원 대표발의)(계속)

그림 36 회의록의 주제(쟁점)에 해당하는 안건을 선택하는 기능

teddysum | 회의록 요약 분석을 위한 데이터 구축 플랫폼

회의록 요약 분석 플랫폼 작업현황 작업 작업지침 Q&A

teddy/날

로그아웃

사립학교법 일부개정법률안 (민병주)

Topic 중 선택

Title: 위임 후 학자권 상환 특별법 일부개정법률안 (유은혜)
 Topic: 21. 위임 후 학자권 상환 특별법 일부개정법률안(유은혜 의원 대표발의)(계속)

Note를 입력해주세요

저장

번호	발의자	발의
18	김희정	출 전에 올림두원장이 지난달에 있어서 21일 위임 후 학자권 상환 특별법 일부개정법률안에 대해서 보고 해 주시기 바랍니다.
24	임진태	2학에서는 위임 후 학자권 상환 특별법의 적용을 받는 대상에 대한충성을 요청하는 내용입니다.
37	박병호	제안에 우선 많은 관심 부탁해 학자권 소요되는 그런 어려운 일이 있어 가지고 과제를이 수월하가가 어 렵습니다. 이 일 잘해해 주시기 바랍니다.
41	박병호	대학원에 포함하는 것은 대학원생은 연구개발이라든지 정책유역으로 지원하는 게 더 현명하이다. 장학금이나 이런 지원을 지원보다는 BK21이라든지 기타 연구지원사업을 통해서 지원하는 게 좋겠다는 그런 의 견이고요.
59	유은혜	그리고 제가 알기로 대학원생들에 대해서 지원 할수있는 것처럼 BK21 연구비, 인건비 지원 등이 혁신다고 받는다고.제가 학부생하고 대학원생하고 대충 신청을을 보니까 대학원생들이 더 높아요. 비슷해요.
66	유은혜	그리고 지금 연구비나 인건비 지원이 대학원생들에게 많이 지급되는 것처럼 지원을 하신지만 실제로 그런 혜택을 받는 학생들이 많다면 대학원들이 몰려야 하는데 대학원은 점점 더 줄고 있습니다. 대학원생들의 대를 수요는요.
69	김정남	지금 대학원이 학부화되는 현상이 과연 바람직인가. 그렇지 않아도 전국이 과대학 대학 전체으로 인해서 분상을 잃고 있고 지금 그 문제를 어떻게 해결할까 고민이 되는 거예요 대학원 전체화지도 어떻게 대응하되 모습으로 잘 공부해 보면 교육 인력들의 학자권 자원의 연구가 바람직한가 하는 생각이 듭니다. 그래서 이

그림 37 회의록의 주제(쟁점) 분할 기능

□ 주제(쟁점)별 요약문 작성 기능

분할된 주제별 발화들은 각 발화의 집합을 하나의 쟁점으로 간주하고 각 쟁점에 대한 요약문을 작성해야 하며 쟁점별 요약문은 표 22와 같은 세부 요약문을 포함한다. 또한 그림 38과 같이 주제에 대한 안건을 연결하고 주제를 나타내는 중심어를 작성한 후 2개 이상의 세부 요약문을 작성하게 된다. 이러한 작업을 효과적으로 수행하기 위하여 그림 39와 같이 데이터 요약 플랫폼을 개발·개선하였다.

세부 요약문의 유형	내용
문제 제기 (Problems/Issues)	논쟁적 대화에서 논의의 대상이 되는 쟁점/문제/이슈
결정 사항 (Decisions)	논쟁적 대화를 통해 도출된 결정 사항
후속 조치 (Actions)	논쟁적 대화를 통해 결정된 내용의 후속조치나 과업사항

표 22 회의록의 세부 요약문 유형

중심어	영산강·섬진강수계 물관리 및 주민지원 등에 관한 법률 일부개정 법률안	
안건	9. 영산강·섬진강수계 물관리 및 주민지원 등에 관한 법률 일부개정 법률안(우윤근 의원 대표발의)	
요약문	영산강·섬진강수계 물관리 및 주민지원 등에 관한 법률 일부개정 법률안에 대해 수계관리기금에 대한 지원 범위와 지역에 대한 이견으로 보류를 결정했다.	
	Problems/Issues	전문위원은 해당 영산강·섬진강수계 물관리 및 주민지원 등에 관한 법률 일부개정법률안에 대해 주요 골자 및 수정 의건과 보완해야 할 사항들을 개괄하였다.
	Problems/Issues	윤종수 환경부차관은 상·하류 간의 공존을 위한 수계관리기금으로 상류를 돕고 있는데, 여기에 하류도 지원한다고 한다면 해당 법의 취지에 어긋나므로 119쪽 부분에 대한 삭제 요청과 함께, 상수원 관리지역은 별도의 지원 방안을 마련해야 하므로 해당 법안에 대해 문제점을 지적하였다.
	Decisions	의사일정 제9항은 소위원회 계속 심사가 결정되어 보류되었다.

그림 38 회의록의 세부 요약문 작성 예시

① 키워드 입력

② 키워드를 추출한 발화ID 입력

③ 확인버튼 클릭

④ 요약할 발화 선택 (shift 버튼 사용 가능)

⑤ 화살표 버튼 클릭

⑥ 발화 타입 선택

-PI
-action
-decision
-abstract

⑦ 내용 요약

※abstract는 반드시 작성
abstract를 작성하지 않으면 넘어가지 않음. 반드시 쟁점별 abstract를 작성할 것!!

⑧ 요약 작업 반복 후 저장

⑨ 다음 버튼 클릭 (다음 토픽 요약 작업)

그림 39 회의록의 세부 요약문 작성 기능

□ 대표 요약문 작성 기능

그림 40에서 확인할 수 있듯이 각 세부 요약문을 통합하고 운문하여 다시쓰기(paraphrasing) 방식으로 주제별 요약문을 작성한 후 이를 다시 통합하고 중복된 내용을 제거하여 회의록 문서 전체에 대한 대표 요약문을 작성한다.

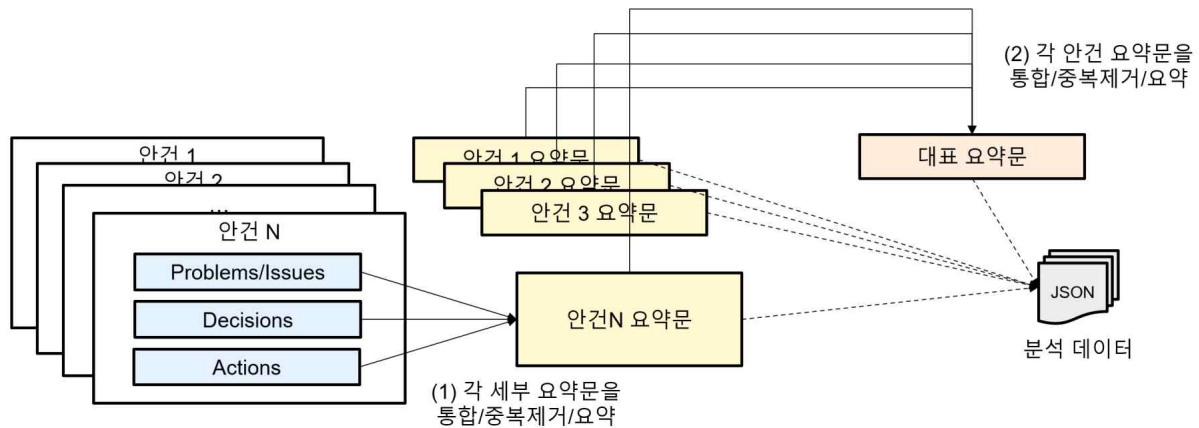


그림 40 상향식 대표 요약문 작성의 도식도

이러한 상향식 요약문 작업 방식을 준수하기 위해 그림 41과 같이 좌측에서 주제별 요약문 내용을 확인하면서 우측에서 대표 요약문을 작성할 수 있는 기능을 개발하였다.

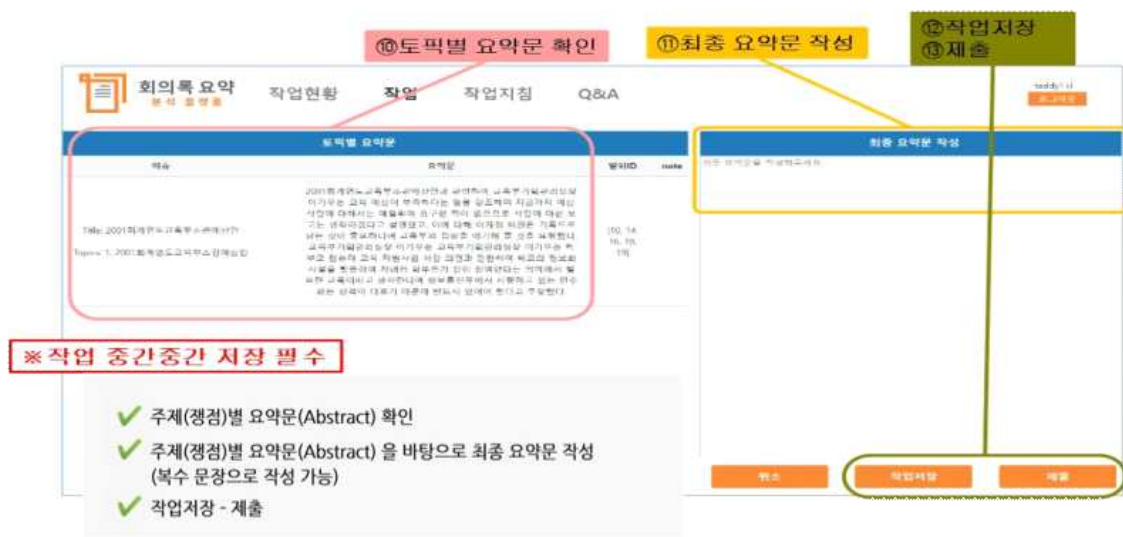


그림 41 회의록의 대표 요약문 작성 기능

□ 데이터 검수 및 수정 기능

작업 결과를 확인한 검수자가 작업자에게 어떤 부분을 왜 수정해야 하는지를 알려줌으로써 작업자가 데이터를 수정하고 다음 작업에도 참고할 수 있도록 하기 위해 그림 42와 같이 검수 노트(note) 기능을 적용하였다.

The interface displays a list of items for review. Each item has a checkbox, a text box for notes, and a '삭제' (Delete) button. The items are organized into sections: '수정 전 화면' (Before Edit) and '수정 후 화면' (After Edit). A red box highlights the '작업자의 수정 내용' (Worker's Edit Content) section.

수정 전 화면 (Before Edit):

- Item 5: Decisions. Note: 이회규 위원은 그렇다면 나중에 기술전문가들과 함께 논의를 해서 완벽하게 고치는 것이 좋겠다고 의견을 밝혔고, 소위원장 선생님은 이회규 위원과 의견이 통의하여 이 문제는 나중에 다시 하기로 하고 넘어가는 것으로 결정했다. (96, 97, 98, 99)
- Item 6: Decisions. Note: 이회규 위원은 주후에 기술전문가들과 함께 논의를 해서 완벽하게 고치는 것이 좋겠다고 의견을 밝혔으며, 소위원장 선생님은 이회규 위원과 의견이 통의하여 12항을 포함하여 이 문제는 나중에 다시 논의하기로 결정하였다. (96, 97, 98, 99)
- Item 7: Actions. Note: 원시일장 4항 국회법중개정법률안과 12항 국정감사법조사 기관장법률중개정법률안은 주후에 기술전문가들과 함께 논의를 해서 완벽하게 고치는 것이 좋겠다고 의견에 따라 나중에 다시 논의할 것이다. (94, 95, 96, 97, 98, 99)

수정 후 화면 (After Edit):

- Item 6: Decisions. Note: 이회규 위원은 주후에 기술전문가들과 함께 논의를 해서 완벽하게 고치는 것이 좋겠다고 의견을 밝혔으며, 소위원장 선생님은 이회규 위원과 의견이 통의하여 12항을 포함하여 이 문제는 나중에 다시 논의하기로 결정하였다. (96, 97, 98, 99)
- Item 7: Actions. Note: 원시일장 4항 국회법중개정법률안과 12항 국정감사법조사 기관장법률중개정법률안은 주후에 기술전문가들과 함께 논의를 해서 완벽하게 고치는 것이 좋겠다고 의견에 따라 나중에 다시 논의할 것이다. (94, 95, 96, 97, 98, 99)

작업자의 수정 내용 (Worker's Edit Content):

- Item 6: Decisions. Note: 이회규 위원은 주후에 기술전문가들과 함께 논의를 해서 완벽하게 고치는 것이 좋겠다고 의견을 밝혔으며, 소위원장 선생님은 이회규 위원과 의견이 통의하여 12항을 포함하여 이 문제는 나중에 다시 논의하기로 결정하였다. (96, 97, 98, 99)
- Item 7: Actions. Note: 원시일장 4항 국회법중개정법률안과 12항 국정감사법조사 기관장법률중개정법률안은 주후에 기술전문가들과 함께 논의를 해서 완벽하게 고치는 것이 좋겠다고 의견에 따라 나중에 다시 논의할 것이다. (94, 95, 96, 97, 98, 99)

검수자가 검수 note와 아래 수정내용을 확인한 후 이 요약문은 삭제처리함.

그림 42 회의록의 검수 의견 작성 및 데이터 수정 기능

□ 작업 현황 관리 기능

작업자들의 근태와 일정을 관리하기 위해 개인별 작업량 할당 기능과 모니터링 기능을 개발하였다. 실시간으로 주간 및 전체 목표치를 확인하고 잔여 작업량을 파악함으로써 진척 상황을 모니터링할 수 있다. 이를 통해 작업자들이 자발적으로 일정과 작업량을 지킬 수 있도록 유도하고 작업 목표치를 효율적으로 관리할 수 있다.

주간 작업현황

	미작업	작업 중	검수 대기	수정 필요	최종 완료
김지승	0	0	0	0	0
이지인	0	0	0	0	0
변종석	0	0	0	0	0
오유진2	0	0	0	0	0

그림 43 회의록 요약 작업자의 작업 현황 확인 기능

- 작업 현황 화면 제공(작업자/관리자)
- 개인별/기간별 작업량 할당
- 전체 및 주간 목표 작업량 설정
- 작업 진척률 및 일정 관리에 용이

teddysum 2023회의록요약 데이터 구축 플랫폼		작업현황					
		이전 주	미작업	작업 중	검수 대기	수정 필요	최종 완료
주제 분포 미작업	0	전체	0	0	0	0	0
주제 분포 작업 중	0	김지승	12	1 (1%)	5 (4%)	0	2 (4%)
주제 분포 검수 대기	0	박민석	2	9 (45%)	7 (35%)	0	2 (10%)
주제 분포 수정 필요	0	김지영	17	3 (18%)	0 (0%)	0 (0%)	0 (0%)
주제 분포 최종 완료	3000	황희우	18	2 (11%)	0 (0%)	0	0 (22%)
주제 분포 계	3000	전 기간	미작업	작업 중	검수 대기	수정 필요	최종 완료
대회 요약 미작업	2130	전체	0	0	0	0	0
대회 요약 작업 중	75	김지승	214	2 (1%)	39 (18%)	0	51 (24%)
대회 요약 검수 대기	452	박민석	206	11 (5%)	26 (13%)	0	65 (31%)
대회 요약 수정 필요	27	김지영	222	9 (4%)	65 (30%)	4 (2%)	9 (4%)
대회 요약 최종 완료	313	황희우	223	3 (1%)	61 (27%)	0	22 (10%)
대회 요약 계	3000	전 기간	미작업	작업 중	검수 대기	수정 필요	최종 완료

그림 44 회의록 요약 관리자의 작업 현황 확인 기능

□ 질의응답 기능

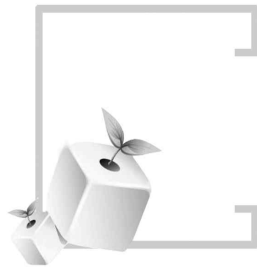
데이터 구축 플랫폼을 통해 실시간으로 질의응답을 주고받을 수 있는 기능을 개발하였다. 어떤 문서에서 어떤 부분이 궁금한지 구체적으로 질문을 작성할 수 있고 그에 대한 답변도 확인할 수 있다. 또한 모든 질의응답 내용이 공개되기 때문에 다른 작업자들 역시 해당 내용을 확인하여 작업에 참고할 수 있도록 하였다.



그림 45 회의록 요약 분석 플랫폼의 질의응답 화면

2.2.5 회의록 요약 분석 말뭉치 구축 분석 지침

본 사업에서는 사업을 수행하면서 회의록 요약 말뭉치 연구 분석 보고서(국립국어원 2021, 2022)를 통해 개발한 ‘논쟁적 대화 요약 분석 말뭉치 구축 지침’을 세밀화하고자 하였다. 이 지침은 무엇보다 작업자가 손쉽게 지침 내용을 이해하고 보다 빠르게 데이터 구축 플랫폼에 적응할 수 있도록 상세한 설명과 실제 예시를 바탕으로 작성한 지침이다. 해당 지침의 상세한 내용은 ‘[붙임 2] 회의록 요약 분석 말뭉치 구축 분석 지침’에서 확인할 수 있다.



제 3 장

사업 수행 결과



3.1 일상 대화 요약 분석 말뭉치 구축 결과

3.1.1 일상 대화 요약 분석 말뭉치 통계

다음은 일상 대화 요약 분석 말뭉치 구축 결과를 바탕으로 분석한 내용이다.

□ 문장 주석 비율

요약 분석 말뭉치를 대상으로 문장 주석 비율을 분석하였다. 표 23에서 확인할 수 있듯이, 일상 대화 원시 말뭉치에서는 주제와 관련이 없다고 판단되는 불필요한 문장(잉여적 문장)이 주제와 관련된 핵심 문장(중심 문장) 이상으로 존재한다. 이는 ‘웃음’이나 ‘기침’ 등 발화가 아닌 소리(빈칸으로 표시) 또는 ‘아’, ‘근데’와 같이 과도하게 짧게 끊어진 발화가 자주 등장하는 일상 대화 원시 말뭉치의 특성에 기반한다.

구분	수	비율
중심 문장	78,901	32.58%
뒷받침 문장	22,605	9.33%
잉여적 문장	140,668	58.08%
전체 문장 수	242,174	100%

표 23 일상 대화 요약 분석 말뭉치의 문장 주석 비율

□ 중심 문장 수

일상 대화 요약 분석 말뭉치는 하나의 문서에 평균적으로 1~3개의 소주제가 포함되어 있다. 요약 분석 말뭉치 3,000건의 소주제별 평균 중심 문장 수는 약 10개, 문서별 평균 중심 문장 수는 약 26개로 확인되었다. 소주제별 최저 중심 문장 수가 1개인 이유는 소주제에 포함된 내용이 무척 짧은 경우가 존재하기 때문이고, 최고 중심 문장 수가 100개를 넘는 이유는 빈칸 개입 등 다양한 이유로 하나로 묶어 주석하기 어려운 발화가 다수 존재하기 때문인 것으로 보인다.

구분	문서별 중심 문장 수	소주제별 중심 문장 수
최고	211	110
최저	6	1
평균	26.3	10.16

표 24 일상 대화 요약 분석 말뭉치의 중심 문장 수

□ 중심 문장에 연결된 뒷받침 문장 수

발화를 뒷받침 문장으로 주석하였다면 반드시 그 문장이 어떤 중심 문장에 대한 뒷받침 문장(부연 설명)인지를 선택하여야 한다. 또한 발화자의 발화 습관이나 상대 발화자의 발화 위치 등에 따라 뒷받침 문장의 수가 많아지는 경우도 있다.

구분	중심 문장에 연결된 뒷받침 문장 수
최고	8
최저	0
평균	1.1

표 25 일상 대화 요약 분석 말뭉치의 중심 문장에 연결된 뒷받침 문장 수

□ 화행 유형 주석 비율

화행 유형 주석은 발화가 소주제와 관련된 내용인지 아닌지를 구분(문장 주석)한 후 해당 발화가 어떤 화행 정보를 갖고 있는지를 주석하는 작업이다. 화행 유형은 아래 여섯 가지로 분류하였으며 상세 기준은 일상 대화 요약 분석 말뭉치 작업 지침을 통해 확인할 수 있다.

표 26을 보면 여섯 가지의 화행 유형 중 ‘정보 제공’이 가장 높고 ‘반응’이 두 번째, ‘질문’이 세 번째로 높게 나온 것을 알 수 있다. 이는 일상 대화가 일상적인 소주제에 관한 질문과 그에 대한 답변으로 구성되어 있기 때문이며 ‘반응’은 질문에 대한 즉각적이고 짧은 답변에, ‘정보 제공’은 질문에 대한 부가적이고 긴 답변에 해당한다.

구분	수	비율
질문	41,328	17.07%
반응	47,576	19.65%
정보 제공	142,756	58.95%
요구	1,871	0.77%
감정 표현	733	0.3%
기타	7,910	3.27%

표 26 일상 대화 요약 분석 말뭉치의 화행 유형 주석 비율

□ 중심 문장 간 관계 주석 비율

문장 주석 및 화행 유형 주석을 마치고 나면 중심 문장 간의 관계를 주석하여야 한다.

지지, 반박, 무관 중 무관의 비율이 무척 높은 것을 알 수 있는데 이는 일상 대화 원시 말뭉치의 특성에 기반한다. 국립국어원에서 제공한 일상 대화 원시 말뭉치는 일상 속 다양한 주제에 대해 두 사람 또는 서너 사람이 짧은 대화를 하는 데이터로 대체로 주제와 관련된 질문과 답변을 주고받는다. 일상 대화는 어떤 내용에 대해 동의, 비동의할 수 있는 논쟁적 주제보다는 취향이나 근황, 고민 등에 관한 일상적 주제로 구성되어 있기 때문에 ‘네’, ‘아니오’로 답변할 수 있는 질문보다는 서술형으로 답변할 수 있는 질문이 많을 수밖에 없다. 따라서 지지, 반박보다 무관의 비율이 압도적으로 높게 나올 수밖에 없다.

구분	수	비율
지지	2,193	3.08%
반박	322	0.45%
무관	68,621	96.46%
전체	71,136	100%

표 27 일상 대화 요약 분석 말뭉치의 중심 문장 간 관계 주석 비율

□ 중심어 수

요약 분석 말뭉치를 대상으로 중심어 수와 비율을 분석하였다. 중심어는 소주제를 포괄하는 단어 또는 명사구로 소주제별로 반드시 최소 1개 이상 작성해야 한다. 사업 초반 작업 시에는 불필요하게 너무 많은 중심어를 잡는 경우가 있어 수행 과정에서 이를 수정하였다.

구분	중심 문장	중심어	비율
수	78,901	11,446	14.51%

표 28 일상 대화 요약 분석 말뭉치의 중심어 비율

중심어는 소주제를 아우를 수 있는 범위의 단어(또는 명사구)를 작성해야 하며 적절한 중심어를 찾을 수 없는 경우 소주제보다 넓은 범위의 단어를 작성하도록 되어 있다. 문서별 중심어 수가 가장 많은 경우는 13개의 중심어가 작성되었는데 이는 문서에 평균보다 많은 소주제가 포함되어 있기 때문이다. 표 29를 통해 알 수 있듯 문서별 평균 중심어는 약 3개, 소주제별 평균 중심어는 약 1개 작성되었다.

구분	문서별 중심어 수	소주제별 중심어 수
최고	13	8
최저	1	1
평균	3.81	1.47

표 29 일상 대화 요약 분석 말뭉치의 문서별, 소주제별 중심어 수

□ 중심어 길이

중심어는 소주제를 아우를 수 있는 단어 또는 명사구로 최대 3어절까지 작성할 수 있다. 적합한 중심어가 없는 경우에는 어떤 단어가 의미하는 범위가 소주제보다 넓더라도 해당 단어를 사용하도록 하였다.

최고 및 최저 길이의 중심어 예시는 표 30에서 확인할 수 있다.

구분	중심어 길이 (띄어쓰기 포함 음절 수)	예시
최고	12	좋아하는 예능 프로그램
최저	1	꽃, 일, 옷
평균	2.9	

표 30 일상 대화 요약 분석 말뭉치의 중심어 길이

□ 소주제별 요약문 내 ‘결정 사항’

소주제별 요약문은 소주제에 대한 전반적인 내용을 담은 ‘대화 요약’과 소주제에 대한 화자별 의견인 ‘화자별 요약’ 그리고 ‘결정 사항’으로 구분된다. ‘결정 사항’은 대화를 통해 결정 사항이 발생한 경우에만 작성하는 선택적 요약문이다. 그러나 근황, 고민 등 일상적인 주제로 이루어진 일상 대화의 경우 결정 사항이 나오기 어렵기 때문에 문서 대부분 결정 사항이 작성되지 않는다.

문서별 결정 사항 수	내용 예시
43	SD2100007과 SD2100008은 태국의 방콕으로 여행을 가기로 합의한 뒤 세부 여행 계획을 추후 함께 정하기로 약속했습니다.
	SD2101850은 SD2101849에게 토익 교재 레벨 투를 빌려주기로 했습니다.

표 31 일상 대화 요약 분석 말뭉치의 결정 사항 수

3.1.2 일상 대화 요약 분석 말뭉치 검수 결과

□ 오류 유형 분석

3,000건의 일상 대화 요약 분석 말뭉치를 구축하며 작업자에 따라 다양한 오류가 발생하였고 검수자의 실시간 검수를 통해 검수 내용을 전달하고 수정하는 작업을 반복하였다. 이 과정에서 눈에 띄게 반복해서 발생하는 몇 가지 오류 유형을 분석하였다.

1) 중심어 선택 오류

중심어는 소주제를 포괄하는 단어 또는 명사구(최대 3어절)로 반드시 소주제별로 1개 이상 작성하여야 한다. 중심어 선택 오류에서는 표 32와 같이 소주제를 모두 아우르는 중심어가 아닌 세부 중심어를 과도하게 많이 잡은 경우, 명사구로 보기 어려운 중심어를 작성하는 경우 등이 자주 발생하였다.

예를 들어 먹거리가 소주제인 문서에서는 ‘떡볶이’, ‘순대’와 같이 음식의 종류를 중심어로 선택하기보다는 ‘음식’이나 ‘분식’, ‘양식’과 같이 소주제 자체를 아우를 수 있는 중심어를 작성해야 한다. 또 중심어를 작성하기 어렵다고 판단되는 문서에서는 ‘부동산에 투자’, ‘여행지에서 음식’과 같은 중심어를 잡기보다는 차라리 범위가 넓어지더라도 단어인 ‘투자’, ‘음식’으로 잡는 것이 마땅하다.

유형	내용	대응
중심어 과다 선택	‘음식’이 소주제인 경우 대화에 등장하는 모든 음식의 종류를 중심어로 잡는 경우	검수 노트(note) 작성 및 SNS 전체 공지
중심어 중복 선택	위의 중심 문장에서 이미 선택한 중심어를 아래의 중심 문장에서 반복적으로 선택	
명사구가 아닌 중심어 선택	‘부동산에 투자’, ‘여행지에서 음식’과 같이 명사구가 아닌 중심어를 작성	

표 32 일상 대화의 중심어 선택 오류 유형

다음 그림 46은 중심어를 잘못 선택한 예시로, 해당 문서에서 적절한 중심어는 ‘선물’인데 그보다 범위가 좁은 ‘보석’으로 중심어를 선택하여 검수자의 검수 노트(note)에 따라 수정하였다.

The screenshot shows a web application for document analysis. On the left, there is a list of messages with columns for selection, message ID, sender, and message content. On the right, there is a detailed view of a selected message. In this view, the 'center word' (중심어) is highlighted in a green box, and the word '보석' (Jewelry) is selected from a dropdown menu, which is also highlighted in a green box. A red arrow points from the list to the detailed view.

그림 46 일상 대화의 중심어 선택 오류

2) 문장 및 화행 유형 주석 오류

문장 및 화행 유형 주석 오류는 표 33과 같이 나누어 주석해야 하는 발화를 하나로 묶어 주석했거나, 적합하지 않은 문장 또는 화행 유형으로 주석한 경우로 나뉜다.

유형	내용	대응
발화 분할 오류	문장 내 정보 제공, 질문이 함께 포함되어 주석된 오류	검수 노트(note) 작성 및 SNS 전체 공지
문장 주석 선택 오류 1	의미 없는 발화를 '중심 문장'으로 주석	
문장 주석 선택 오류 2	질문에 대한 답변으로 보기 어려운 발화를 '뒷받침 문장'으로 주석	
화행 유형 주석 오류	질문에 대한 대답을 주석하는 경우 반응/정보제공을 혼동하는 오류	

표 33 일상 대화의 문장 및 화행 유형 주석 오류 유형

하나의 발화를 중심 문장으로 주석하였으나 그림 47과 같이 해당 발화에서 아무 정보도 얻을 수 없는 경우가 있다. 이 경우 ‘잉여적 문장’으로 주석하거나 앞뒤에 다른 발화를 묶어 주석하여야 한다.

SDRW200000607
Title: 2인 일상 대화 (기타)
Topic: 계절/날씨 > 봄, 여름, PC방, 음식

순서 입력

<input type="checkbox"/>	217	SD2000741	시 시골에 인제 시골에 벼농사 이런 걸 많이 하다 보니까
<input type="checkbox"/>	218	SD2000741	인제 그~
<input type="checkbox"/>	219	SD2000741	수확하는 느낌도 있고 그다음에
<input type="checkbox"/>	220	SD2000741	회사 다니면서는 인제
<input type="checkbox"/>	221	SD2000741	올 한 해도 인제 거의 끝나가는구나 그런 어떤
<input type="checkbox"/>	222	SD2000741	유종의 미를 거두어가는 느낌도 있고
<input type="checkbox"/>	223	SD2000741	그런 느낌이 조금 강했던 거 같아요.
<input type="checkbox"/>	224	SD2000741	겨울은
<input type="checkbox"/>	225	SD2000741	한겨울에 요즘에 느끼는 게 좀 느낌이 뭐냐면
<input type="checkbox"/>	226	SD2000741	올
<input type="checkbox"/>	227	SD2000741	이~ 옛날에는 정말 눈이 진짜 많이 왔어요. 그래서
<input type="checkbox"/>	228	SD2000741	그~ 기억이 x는데 모는데
<input type="checkbox"/>	229	SD2000741	어렸을 때
<input type="checkbox"/>	230	SD2000741	연탄이 있었을 때 연탄에다 눈 이렇게 굴러가지고 눈사람도 많이 만들고

(21) ID:32625

발화자 SD2000741

발화ID 224

중심어 확인

화행타입 반응

문장타입 중심 문장

겨울은

검수 내용을 입력해주세요.

▲ ▼

저장 삭제

(22) ID:32626

발화자 SD2000741

발화ID 225, 226, 227, 228

문장타입 잉여적 문장

한겨울에 요즘에 느끼는 게 좀 느낌이 뭐냐면 올 이~ 옛날에는 정말 눈이 진짜 많이 왔어요. 그래서 그~ 기억이 x는데 모는데

검수 내용을 입력해주세요.

▲ ▼

저장

그림 47 일상 대화에서 중심 문장을 과하게 짧게 주석한 경우

다음 그림 48은 적합하지 않은 문장 유형으로 주석한 예이다. 첫 번째는 휴학을 하면 좋은 점에 대한 답변을 중심 문장으로 주석해야 하는데, 부연 설명으로 보이는 발화를 중심 문장으로 주석하였고 검수자의 피드백에 따라 수정하였다.

두 번째는 경기를 볼 때 가장 중요하게 생각하는 부분이 무엇인지 물었는데 그에 대한 답변이 아닌, 선수에 대해 설명한 발화를 뒷받침 문장으로 주석하여 잉여적 문장으로 수정하였다.

문장타입 중심 문장 ▼	발화자 SD2000406	혹시 후학을 직접 하시게 됐는데 후학의 좋은 점 이 있을까요?
	발화ID	28, 29, 30, 31, 32, 33
	중심어	<input type="text"/> 확인
	화형타입	질문 ▼
(5) ID:18301		
문장타입 중심 문장 ▼	발화자 SD2000408	저는 일단 1년 후학을 했는데 반년은 오전에 평일 동안 알바를 하면서 오후에는 이제 자격증 정월이랑 트릭 같은 거 공부를 하면서 지냈고
	발화ID	34, 35, 36, 37, 38, 39, 40, 41
	중심어	<input type="text"/> 확인
	화형타입	정보 제공 ▼

(3) ID:2162		
문장타입 뒷받침 문장 ▼	발화자 SD2000090	그래서 좋아하는 에프시 서울 선수 중에 지금은 은퇴를 했지만 좀 김진규라는 선수가 있었는데 특히 좀 애중에 선수긴 하지만 특히 그 선수가 개인적으로 상대 공격수를 특히 어떻게 막는지 개인적으로 인상 깊게 봤던 편이고 또 수비형 미드필더라는 포지션이 그 행당
	발화ID	389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399
	중심문장	<input type="text" value="35265"/> 확인

그림 48 일상 대화에서 문장 주석을 잘못된 경우

3) 요약문 작성 오류

요약문은 원문의 구어체를 그대로 사용하거나 ‘화자별 요약’의 내용을 그대로 ‘대화 요약’에 사용하여서는 안 된다. 또한 ‘잉여적 문장’으로 주석한 발화의 내용은 요약문에 작성하지 않는다.

요약문 작성 오류 유형은 표 34와 같이 원문을 그대로 차용해 구어체로 작성하거나 주석한 중심 문장에 비해 요약문이 빈약한 경우, 잉여적 문장을 요약문에 참고하여 작성한 경우 등이 있다.

유형	내용	대응
구어체 또는 비문	원문을 그대로 가져와 구어체가 포함되어 있거나 부자연스러운 문장	검수 노트(note) 작성 및 SNS 전체 공지
빈약한 요약문	중심 문장에 비해 요약문이 빈약한 경우	
요약문 내용 누락	하나의 세부 주제 안에서 여러 소주제가 등장하는데 중심 문장에 포함된 내용인데도 요약문에서 누락된 경우	
임의로 줄 인 요약문	전체 요약문은 소주제별 요약문의 '대화 요약' 부분을 자연스럽게 잇도록 되어있는데, '대화 요약' 내용보 다 '전체 요약문'을 굳이 짧게 작성한 경우	
잉여적 문장 참고	요약문 작성 시 중심, 뒷받침 문장만 참고하도록 되어있는데, 잉여적 문장의 내용을 요약문에 작성한 경우	
요약문 중복	'대화 요약'과 '화자별 요약'의 내용을 똑같이 작성	
화자 오류	화자별 요약문의 내용을 반대로 작성한 경우	
대화 요약 내용 중복	대화 요약문 작성 시 같은 내용이 중복되는 경우	

표 34 일상 대화의 요약문 작성 오류 유형

다음은 '화자별 요약'의 내용을 그대로 복사-붙여넣기 하여 '대화 요약'에 사용한 경우이다. 각 요약문은 동일한 표현 사용을 지양하므로 동의어나 다른 표현을 사용하여야 하고 '대화 요약'의 화자별 의견은 '화자별 요약'의 내용보다 간결하게 작성해야 하기 때문에 검수자의 피드백에 따라 수정하였다.

주제별 요약문 작성 1 / 2		※원문의 정보를 최대한 포함하여 작성하는 것이 목표이므로 대화의 내용을 충분히 담아 낼 것	
발화자	발화	대화 요약	화자별 요약
SD2000544	나는 요즘 솔직히 스포츠에 관한 걸 안 보기도 했고 어~ 좋아했던 거 좋아했던 거라기보다는 내가 했던 거는 솔직히 하키가 제일 좋아했던 거 같아. 하키를 하기도 했고 특히 필드하키를 했었어. 근데 필드하키는 내가 축구하는 거랑 되게 많이 비슷해가지고 되게 하기도 쉽고 뭐 너도 알기 쉬운 거야.	이 대화에서는 하키와 축구에 대해 이야기했습니다. SD2000544는 필드하키에서 미드필더와 수비로 뛰었으며, 비 인기종목이라 관심이 없다고 이야기했습니다. 또 축구와 많이 비슷하다고 이야기했고 축구 골키퍼의 역할, 부상에 대해 물었습니다. SD2000545는 골키퍼는 책임감이 필요한 포지션이고 몸싸움이 잦아 고의적인 부상, 사고로 생기는 부상으로 나뉜다고 이야기했습니다.	화자별 요약 SD2000544
SD2000544	너는 어떻게 생각을 해? 하키에 대해서	SD2000544는 필드하키에서 미드필더와 수비로 뛰었으며, 비 인기종목이라 관심이 없다고 이야기했습니다. 또 축구와 많이 비슷하다고 이야기했고 축구 골키퍼의 역할, 부상에 대해 물었습니다.	화자별 요약 SD2000545
SD2000545	아~ 근데 그전에 궁금한 게 있는데 어~ 축구랑 하키랑 비슷하잖아. 나도 하키를 본 적이 있어. 근데 진짜로 어~ 뭐랄까? 거기 있잖아. 그~ 규칙에서 그~ 규칙 같은 부분도 다 비슷하냐? 예를 들어 뭐 축구면은 뭐 코너리도 있고 뭐 프리리도 있고 한데 하키도 그러냐? 어. 맞아.	SD2000545는 골키퍼는 책임감이 필요한 포지션이고 몸싸움이 잦아 고의적인 부상, 사고로 생기는 부상으로 나뉜다고 이야기했습니다.	화자별 요약 SD2000545

그림 49 일상 대화에서 화자별 요약의 내용을 대화 요약에 그대로 사용한 경우

다음 그림 50은 대화 요약 작성 시 화자별 요약문의 내용을 반대로 작성한 경우의 예시로 화자별 의견이 올바르게 작성되도록 수정하였다.

주제별 요약문 작성 3 / 3		※원문의 정보를 최대한 포함하여 작성하는 것이 목표이므로 대화의 내용을 충분히 담아 낼 것	
발화자	발화	대화 요약	화자별 요약
SD2000715	어~ 그리고 저는 사실 과딩 처음 다녔던 대학 처음 다녔던 대학교에서 만 해 보고 두 번째에 다녔던 대학교에서는 과 씨씨가 돼가지고 안 했거든요?	다음 대화에서 화자 SD2000714와 SD2000715는 과 소개팅을 해본 경험에 대해 이야기를 나눕니다. SD2000714는 과 소개팅을 해 본 경험이 한 번 있었고 SD2000715는 총 2번의 과 소개팅 경험이 있다고 답했는데 두 번 모두 만족스럽지 않았던 경험이었다고 답했습니다. 또한 SD2000715는 과 소개팅이 일반적인 소개팅보다는 조금 더 편하고 부담이 적은 느낌이라고 설명했습니다.	화자별 요약 SD2000715
SD2000715	과딩 하셨나요?	SD2000715는 화자가 다녔던 첫 번째 대학교에서 과 소개팅을 한 번 해본 경험이 있다고 답했습니다.	화자별 요약 SD2000714
SD2000714	네. 제도 두 번 정도 해봤었는데 먼저 처음에 했을 때는 제가 1학년 때였어요. 1학년 때 해 봤는데 그때는 이제 뭐 모르고 갔죠. 저희	SD2000714는 대학교 재학 시절 총 2번의 과 소개팅 경험이 있다고 답했습니다. 첫 번째 과 소개팅에서는 상대방이 마음에 들지 않아서 금방 모임을 파한 후 함께 동행했던 친구들과 밤 늦게 까지 노는 등 소개팅과는 거리가 먼 경험이었었고 두 번째 경험 또한 이성 친구를 사귀지 못하고 실패했다고 말했습니다. 또한 화자는 과 소개팅은 일반적인 소개팅과는 다르게 부담 없이 재밌게 시간을 보낼 수 있는 분위기라서 더 편하다고 답했습니다.	화자별 요약 SD2000714
SD2000714	저희는 나를 이제 기숙사 살았거든요. 근데 기숙사에 예프 포라고 그~ 있었어요 그때 당시 이제 꽃보다남자 할 텐데 기숙사에 예프 포라고 해가지고 한 명은 진짜 송승헌 닮은 분 있었었고 어~ 제가 있었었고 오다듬에 그~ 부활의 정동하 닮은 분 있었거든요. 달은 사람 있었었고		

그림 50 일상 대화에서 화자별 요약의 내용을 반대로 작성한 경우

다음은 소주제별 요약문의 ‘대화 요약’에서 중심 문장으로 주석한 일부 내용을 누락하고 작성한 예시로, 그림 51과 같이 검수자의 검수 노트(note)를 통해 피드백을 받아 수정하였다.

※원문의 정보를 최대한 포함하여 작성하는 것이 목표이므로 대화의 내용을 충분히 담아낼 것	
대화 요약	다음 대화에서 SD2000096와 SD2000095는 데이트에 대해 이야기를 나눴습니다. SD2000096는 문래동 또는 영등포에서 데이트를 즐기는 편이라고 했고 SD2000095는 결혼은 아직 생각이 없고 서른 중반이 적당하다고 이야기했습니다.
화자별 요약 SD2000096	SD2000096는 주로 영등포 또는 문래동에서 술을 마시거나 공연을 보는 등의 데이트를 즐긴다고 말했습니다.
화자별 요약 SD2000095	SD2000095는 데이트라고 하기에는 다소 애매한 데이트를 해본 경험은 있으나 연애 경험은 없기 때문에 아직 결혼을 해야 한다는 생각은 없다고 말했습니다. 또한 가벼운 만남을 원하고 있으며 미래에 결혼을 하게 된다면 서른 중반이 적당한 것 같다고 말했습니다.

그림 51 일상 대화 요약문에서 중심 문장의 내용이 누락된 경우

다음은 일상 대화 요약문에서 부자연스러운 문장이 포함되어 있고 주석한 중심 문장보다 요약문의 내용이 빈약한 경우를 보여준다. 그림 52와 같이 검수자의 상세한 피드백을 바탕으로 수정하였다.

대화 요약	다음 대화에서 화자 SD2000095와 SD2000096은 연애 스타일에 대한 이야기를 나누었습니다. 화자 SD2000096은 자주 데이트를 하는 편이고 장기 연애를 더 많았으며 소개 보다는 자연스러운 만남을 선호한다고 이야기했습니다.
화자별 요약 SD2000095	SD2000095는 최근 시작한 연애에 대해 질문했습니다.
화자별 요약 SD2000096	SD2000096은 올해 3월부터 연애를 시작했고 일주일에 세 번 이상 데이트를 하는 편이며 소개팅은 인위적으로 만남을 추천하는 느낌이라 좋아하지 않는다고 말했습니다. 또한 단기 연애 보다 장기 연애를 더 많이 했다고 이야기했습니다.

그림 52 일상 대화 요약문에서 비문 및 내용 누락이 발생한 경우

마지막 예시로는 소주제별 요약문의 ‘대화 요약’ 내용보다 전체 요약문의 내용이 과도하게 짧은 경우로 중복되거나 불필요한 내용이 아니라면 굳이 삭제할 필요가 없으므로 기계적인 통합 후 윤문하여 제출하도록 피드백한 사례이다.



그림 53 일상 대화에서 소주제별 요약문에 비해 전체 요약문의 내용이 짧은 경우

□ 보완 조치

본 사업팀에서는 양질의 데이터 구축을 위해 사업 종료 후 일정 기간 분석 데이터를 검토·보완하기로 하였다. 기준이 되는 작업 지침을 바탕으로 데이터 평가 및 수정을 진행하되 지침 내용 외에 자주 발견되는 오류와 교차 검수 시 받은 질문의 내용을 참고하여 전수 검사를 진행할 예정이다.

다음은 지침 내용 외에 자주 발견되는 오류와 작업자의 질문을 바탕으로 정리한 내용이다.

- 1) 원문의 명사를 요약문에 작성 시 오류가 아니면 원문의 띄어쓰기, 표현 그대로 사용
- 2) 고유명사가 원문에 잘못된 표현으로 기재되어 있는 경우 올바른 고유명사로 수정해서 작성
- 3) 원문의 표현이나 띄어쓰기가 오류인 경우 올바른 표현으로 변경하여 작성

원문	요약문 작성의 틀린 예	요약문 작성의 옳은 예
무한 v 도전	무한도전	무한 v 도전
~할수밖에 없어서(후략)	~할수밖에 없어서(후략)	~할 v 수밖에 없어서(후략)
에이형	A형	에이형
애 이형	애 이형	에이형 / A형
넷플 렉 스	넷플 렉 스	넷플릭스
야구선수 이대 오	야구선수 이대 오	야구선수 이대 호

표 35 일상 대화 작업 지침 외 참고 사항 예시

4) 화자(주어) 뒤에 붙는 조사 작성 주의

요약문 작성의 틀린 예	요약문 작성의 옳은 예
화자 SD200187 4은	화자 SD200187 4는
화자 SD200254 6는	화자 SD200187 6은
화자 SD200254 1는	화자 SD200187 1은

표 36 일상 대화 요약문 작성 시 주의 사항 예시(주어)

- 5) 발화자/문장 유형/화행 유형이 모두 같은 경우 하나로 묶어서 주석
 - 6) 잉여적 문장 내용을 요약문에 참고하지 않도록 주의
 - 7) 중심 문장으로 선택했음에도 임의적으로 요약문에 내용을 생략하지 않도록 주의
예) 대부분의 문장을 중심 문장으로 주석하고, 요약문은 중심 문장 중 50%만
참고하여 작성되는 경우
 - 8) 비문 주의
 - 9) 주술 호응이 맞는지 반드시 확인하고 목적어가 빠지지 않았는지 확인
 - 10) 과도하게 많은 내용을 한 문장에 쓰는 것은 지양
예) ~했고, ~했고, 했으며,~했습니다.
~했는데, ~는데, ~했습니다.
- *하나의 문장을 너무 길게 작성하는 경우 비문이 될 가능성이 높기 때문에 가급
적이면 문장은 주술 호응에 맞게 짧게 작성하되 '그리고, 또한, 한편, 또' 등과
같은 접속사를 사용하여 이어서 작성

본 사업팀에서 계획한 사업 종료 후 보완 일정은 표 37과 같다.

구분	내용		비고
기간	2023.12~2024.02		3개월 간 추가 작업
인원	12월	5명	테디썸 내부 인력 및 일상 대화 요약 우수 작업자
	1월	3명	
	2월	2명	
작업 내용	일상 대화 요약 데이터 3,000건에 대한 전수 검사 및 수정 작업		

표 37 일상 대화 요약 사업 종료 후 추가 보완 작업 일정

3.1.3 일상 대화 요약 분석 말뭉치 형식

□ 일상 대화 요약 분석 말뭉치의 메타 데이터

표 38은 일상 대화 요약 분석 말뭉치의 메타 데이터 구조를 나타낸다. 일상 대화 요약 분석 말뭉치는 3,000건의 문서를 1개의 파일로 저장한 후 납품하였으며 실제로 분석한 문서의 메타 데이터 예시를 표 39에서 확인할 수 있다.

1수준	2 수준	3 수준	4 수준	설명
id				* 말뭉치 파일 ID
metadata				* 파일의 메타 정보
	title			* 원시 말뭉치: 국립국어원 [말뭉치 유형 구분] [파일 ID] 예) 국립국어원 일상 대화 요약 분석 말뭉치 SDSC2300001001
	creator			* 생성자: 국립국어원
	distributor			* 배포자: 국립국어원
	year			* 말뭉치 구축 연도 예) 2023
	category			* 분류 예) 구어 > 사적 대화 > 일상대화
	annotation_level			* 분석 층위 예) 요약 분석
	sampling			* 샘플링 방식 (본문 전체/ 부분 추출 - 임의 추출 / 부분 추출 - 특정 부분 추출 등)
document				* 문서 정보
	id			* 문서 ID * '원시 말뭉치 파일 ID.파일 내 문서 번호'로 구성 - SDRW2100000001.1
	metadata			* 문서의 메타 정보
		title		* 문서 제목 - 2인 일상 대화
		author		* 작성자, 게시자 등 저작(권)자 - 개인 발화자
		publisher		- 개인 발화 녹음
		date		* 작성일시, 게시일시 * 형식: YYYYMMDD
		topic		* 주제 - 먹거리 > 저녁 음식 선택
		speaker		* 발화자 정보
			id	* 발화자 ID - SD2100001
			age	* 나이.
			occupation	* 직업
			sex	* 성별: 남성/여성
		setting		* 환경 정보
			relation	* 관계: [가족] 부부...
	utterance			* 구어(준구어, 메신저 포함), 메신저 발화 정보
		id		* 발화 ID
		form		* 발화 내용 정제 형태, 표기 전사
		speaker_id		* 발화자 ID

표 38 일상 대화 요약 분석 말뭉치의 메타 데이터 구조

```

"document": [
  {
    "id": "SDRW2100001413.1",
    "metadata": {
      "title": "2인 일상 대화",
      "author": "개인 발화자",
      "publisher": "개인 발화 녹음",
      "date": "20211119",
      "topic": "방송/연예 > 경쟁프로그램에 대한 생각, 흥미롭게 봤던 프로그램",
      "speaker": [
        {
          "id": "SD2100783",
          "age": "30대",
          "occupation": "기타",
          "sex": "여성"
        },
        {
          "id": "SD2100784",
          "age": "30대",
          "occupation": "기타",
          "sex": "여성"
        }
      ],
      "setting": {
        "relation": "친구"
      }
    },
    "utterance": [
      {
        "id": "SDRW2100001413.1.1.1",
        "form": "내가 또 계속 면접 얘기를 하게 되는데.",
        "speaker_id": "SD2100784"
      },
      {
        "id": "SDRW2100001413.1.1.2",
        "form": "어제 면접에서",
        "speaker_id": "SD2100784"
      }
    ]
  }
]

```

표 39 일상 대화 요약 분석 말뭉치의 메타 데이터 정보 예시

□ 발화 정보

각 문서의 발화에 대한 정보는 sentence에 모두 기록되어 있으며, 각 발화를 띄어쓰기로 토큰화한 word의 위치 정보 및 문자열을 모두 기록하여 납품되었다. 각 발화에 대한 id값은 원시 말뭉치의 발화 id와 같이 적용되었다.

2 수준	3 수준	4 수준	설명
discourse			* 담화 분석
	sentence_set		
		sentence_set_id	* 문장 세트에 대한 ID 부여 - 예) 1
		speaker_id	* 문장 세트의 발화자 ID 예) SD2000004
		utterance_id	* 문장 세트에 포함된 발화 ID - 예) SDRW2000000002.1.1.1
		type	* 문장 세트의 유형 - 예) 중심, 뒷받침, 잉여
		dialog_act	* 문장 세트의 화행 - 예) 질문, 요구, 정보 제공, 반응, 감정 표현, 기타
		target	* 뒷받침 문장으로 주석한 경우, 어떤 중심 문장에 대한 뒷받침 문장인지 작성 - 예) 5(중심 문장 세트 ID)
	relational_triple		
		head	* 문장 세트 ID - 관계의 소스 노드 (예: 지지하는 문장)
		tail	* 문장 세트 ID - 관계의 타겟 노드 (예: 지지받는 문장)
		relation	* 문장 관계 - 예: "지지", "반박", "무관"

표 40 일상 대화 요약 분석 말뭉치의 발화 정보 구조

```

"discourse": {
  "sentence_set": [
    {
      "sentence_set_id": "1",
      "speaker_id": "SD2100784",
      "utterance_id": [
        "SDRW2100001413.1.1.1",
        "SDRW2100001413.1.1.2",
        "SDRW2100001413.1.1.3",
        "SDRW2100001413.1.1.4",
        "SDRW2100001413.1.1.5",
        "SDRW2100001413.1.1.6",
        "SDRW2100001413.1.1.7",
        "SDRW2100001413.1.1.8",
        "SDRW2100001413.1.1.9",
        "SDRW2100001413.1.1.10",
        "SDRW2100001413.1.1.11",
        "SDRW2100001413.1.1.12",
        "SDRW2100001413.1.1.13"
      ],
      "type": "뒷받침 문장",
      "dialog_act": "정보 제공",
      "target": 3
    },
    {
      "sentence_set_id": "2",
      "speaker_id": "SD2100784",
      "utterance_id": [
        "SDRW2100001413.1.1.14",
        "SDRW2100001413.1.1.15",
        "SDRW2100001413.1.1.16"
      ],
      "type": "중심 문장",
      "dialog_act": "정보 제공"
    },
    {
      "sentence_set_id": "3",
      "speaker_id": "SD2100784",
      "utterance_id": [

```

(중략)	<pre> "SDRW2100001413.1.1.17"], "type": "잉여적 문장", "dialog_act": "질문" }, "relational_triple": [{ "head": 2, "tail": 7, "relation": "무관" }, </pre>
------	--

표 41 일상 대화 요약 분석 말뭉치의 발화 정보 예시

□ 요약 정보

일상 대화 문서에 대한 요약문은 전체 요약문과 주제별 요약문으로 구분된다.

전체 요약문은 일상 대화 문서 전체에 대한 요약문으로 문서에 포함된 소주제에 대한 ‘대화 요약’의 내용을 기계적으로 통합 후 중복되는 내용이나 불필요한 내용은 삭제하고 환언 및 윤문하여 작성한다.

주제별 요약문은 일상 대화 문서 내의 소주제별 요약문으로 해당 소주제에 대한 전반적인 요약문인 ‘대화 요약’과 소주제에 대한 화자별 의견인 ‘화자별 요약’, 대화를 통해 나온 결정 사항이 있는 경우 작성하는 ‘결정 사항’으로 구분된다.

일상 대화 요약 분석 말뭉치의 요약 정보 구조는 표 42와 같으며, 실제 분석한 말뭉치의 요약 정보 예시는 표 43에서 확인할 수 있다.

2 수준	3 수준	4 수준	5 수준	6 수준	설명
SC					* 대화 요약 분석 정보
	main_summary				* Topic별 요약문을 취합한 전체 요약문
	Topic				* 각 Topic별 주제 분할 정보
		utterance_group			
			sentence_set_id		* 문장 세트 ID - 예) 1
			keyword		* 해당 쟁점을 대표하는 중심어(keyword) - 예) "반려동물"
			utterance_id		* 해당 중심어를 포함하는 발화 ID - 예) SDRW2000000002.1.1.1
			begin		* 중심어의 문장 내 시작 위치 (UTF-8 문자 위치로 0부터 시작)
			end		* 중심어의 문장 내 끝 위치 (UTF-8 문자 위치로 0부터 시작)
		topic_summary			* 각 Topic별 요약문 정보
			conversation_summary		* Topic별 요약문 중 대화 요약 정보
			speaker_summary		* Topic별 요약문 중 화자별 요약 정보
				speaker_id	* 화자ID
				summary	* 화자별 요약
			decisions		* Topic별 요약문 중 결정 사항 정보

표 42 일상 대화 요약 분석 말뭉치의 요약 정보 구조

"SC": {

"main_summary": "두 사람은 서바이벌 프로그램, 서바이벌 드라마와 영화에 관한 대화를 했습니다. SD2100784는 스트리트 우먼 파이터에 불만인 부분도 있지만, 대중의 관심을 끌만큼 흥미로웠다고 말했습니다. 또한 우리 사회 분위기 특성상, 서바이벌 방송이 많다고 했습니다. 한편 오징어 게임이 재밌었고, 이정재 역할이 좋았다고 말했습니다. 그리고 오징어 게임과 형거 게임처럼 서바이벌 요소가 있는 작품이 전 세계적인 대중성을 띠는 점이 흥미롭다고 했습니다. 이와 함께 형거 게임의 열성 팬으로, 영화의 메시지가 마음에 든다고 말했습니다. SD2100783은 우리나라는 서바이벌 프로그램을 선호하는 경향이 있고, 슈퍼스타 케이 이후로 경쟁 프로그램이 많아진 것 같다고 말했습니다. 특히 우승자를 통해 대리만족하는 심리도 있다고 했습니다. 한편 한국 드라마를 선호하지만, 오징어 게임의 흥행은 예상하지 못했다고 말했습니다. 또한 형거 게임은 여주인공이 멋져서 재미있었다고 했습니다.",

"topic": [

{

"utterance_group": [

{

"sentence_set_id": "1"

"keyword": null,

"utterance_id": null,

"begin": null,

"end": null

},

{

"sentence_set_id": "2",

"keyword": "",

"utterance_id": "",

"begin": ,

"end":

},

(중략)

"topic_summary": {

"conversation_summary": "두 사람은 서바이벌 프로그램에 관한 대화를 했습니다. SD2100784는 스트리트 우먼 파이터에 불만인 부분도 있지만, 대중의 관심을 끌만큼 흥미로웠다고 말했습니다. 또한 우리 사회 분위기 특성상, 서바이벌 방송이 많다고 했습니다. SD2100783은 우리나라는 서바이벌 프로그램을 선호하는 경향이 있고, 슈퍼스타 케이 이후로 경쟁 프로그램이 많아진 것 같다고 말했습니다. 특히 우승자를 통해 대리만족하는 심리도 있다고 했습니다.",

{

"speaker_id": "SD2100783",

"summary": "SD2100783은 우리나라는 서바이벌 프로그램

을 선호하는 경향이 있다고 생각한다고 말했습니다. 특히 슈스케 이후로 경쟁 프로그램이 많아졌고, 1위를 한 참가자를 통해 대리만족하는 심리도 있다고 했습니다."

```
    },  
    "speaker_summary": [  
      {  
        {  
          "speaker_id": "SD2100784",  
          "summary": "SD2100784는 중장년층도 스트릿 우먼 파이터를 알고 있는 것을 보면, 프로그램의 인기가 상당한 것 같다고 말했습니다. 그리고 노출은 마음에 들지 않았지만, 팀워크나 리더 간의 관계를 보는 재미가 있었다고 했습니다. 또한 우리나라가 경쟁을 부추기는 분위기라서, 이와 같은 방송이 많아진 것 같다고 말했습니다."  
        }  
      ],  
      "decisions": "null"  
    ]  
  }
```

표 43 일상 대화 요약 분석 말뭉치의 요약 정보 예시

3.2 국회 회의록 요약 분석 말뭉치 구축 결과

3.2.1 회의록 요약 분석 말뭉치 통계

2023 대화 요약 분석 말뭉치 연구 분석 사업의 분석 대상이 된 회의록 문서에 대한 통계는 다음과 같다. 표 44에서 확인할 수 있는 것처럼 회의록 요약 분석 말뭉치는 평균 발화 수 501개, 평균 발화 길이 80개, 평균 어절 수 19개로 이루어져 있다.

항목	2021 회의록 요약 말뭉치	2022 회의록 요약 말뭉치	2023 회의록 요약 말뭉치(본 사업)
회의록당 회의 참석자 인원	16.24 (명)	15.7 (명)	15.88 (명)
회의록당 발화 개수 평균	475.26 (개)	485.18 (개)	501.88 (개)
발화 평균 길이 (글자 수)	68.89 (개)	70.62 (개)	80.36 (개)
발화 평균 길이 (어절 수)	16.41 (개)	16.62 (개)	19.05 (개)

표 44 회의록 요약 분석 말뭉치 관련 통계 비교

항목				내용		
회의록의 개수				300 개		
회의록 1개 기준	대표 요약문		평균 문장	13.84 문장		
			평균 어절	219.47 어절		
	이슈의 개수			2.81 개		
	이슈 1개 기준	이슈 요약문	평균 문장	6.21 문장		
			평균 어절	97.11 어절		
		세부 요약문의 개수			32.74 개	
		세부 요약문 1개 기준	평균 문장	2.04 문장		
			평균 어절	40.48 어절		

표 45 회의록 요약 분석 말뭉치 통계

항목	원시 말뭉치	대표 요약문	요약문 비율
회의록 당 평균 문장	485.18 발화	17,170 문장	3.53%
회의록 당 평균 어절 수	8,064.73 어절	167.76 어절	2.08%

표 46 회의록 원시 말뭉치에 대한 요약문의 비율

3.2.2 회의록 요약 분석 말뭉치 검수 결과

□ 오류 유형 분석

300건의 회의록 분석 데이터를 구축하면서 작업자 교차 검수를 수행하였으며 이를 바탕으로 자주 발생한 오류 유형을 예시와 함께 정리하였다.

1) 표현 및 띄어쓰기 오류

구어체로 이루어진 회의록의 경우 특히 원문의 구어체를 그대로 차용한 오류가 많이 발생하였다. 또한 일상 대화에 비해 상대적으로 많은 분량의 요약문을 작성해야 하는 회의록 작업의 특성상 오타도 많이 발생했다.

다음은 표현 및 띄어쓰기의 오류 유형과 그 예시이다.

유형	내용	대응
표현 오류	원문의 '우리 위원회', '우리나라'를 그대로 요약문에 작성	'우리'를 제외하고 작성하도록 검수 노트(note) 작성 및 SNS 전체 공지
부자연스러운 문장	비문 또는 부자연스러운 문장	자연스럽고 올바른 문장으로 수정하도록 검수 노트(note) 작성 및 SNS 전체 공지
구어체	~하느냐 / 그래가지고 / 그다음에 / 해갖고 등	문어체로 수정하도록 검수 노트(note) 작성 및 SNS 전체 공지
오타	~논의됐고 / ~때는 / ~했따 등	올바른 맞춤법으로 수정하도록 검수 노트(note) 작성 및 SNS 전체 공지
띄어쓰기	본회의 / 회의전 / 논의 했고 등	올바른 띄어쓰기로 수정하도록 검수 노트(note) 작성 및 SNS 전체 공지

표 47 회의록의 표현 및 띄어쓰기 오류 유형

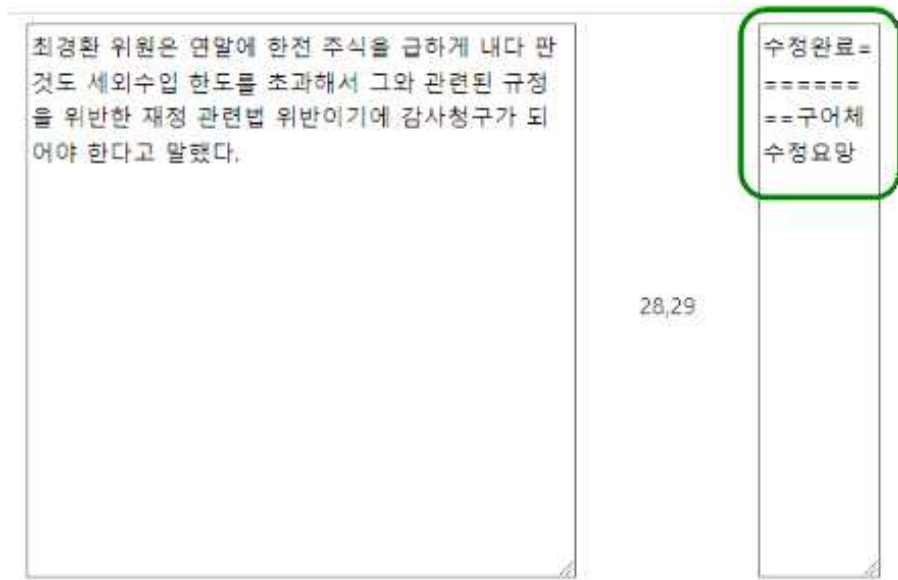


그림 54 회의록 요약문의 작성 오류 예시(구어체)

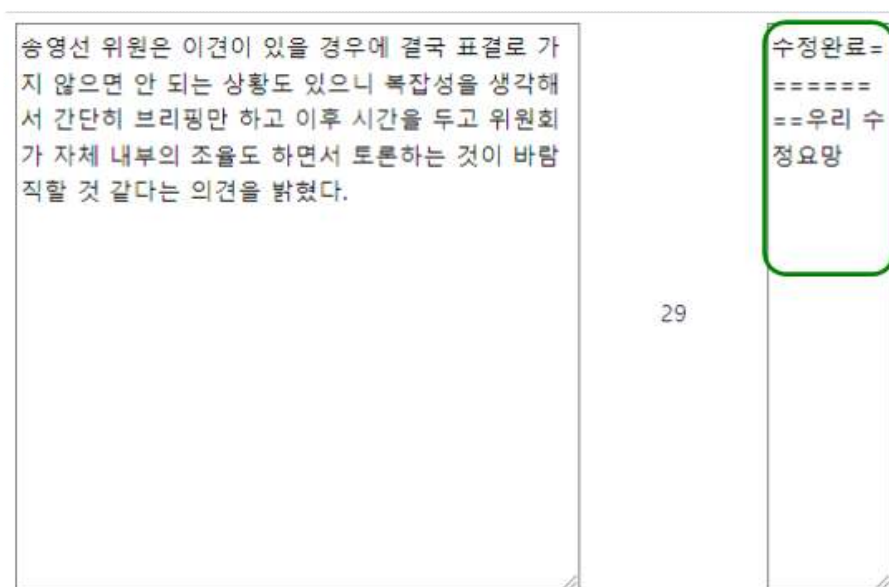


그림 55 회의록 요약문의 작성 오류 예시(표현 오류)

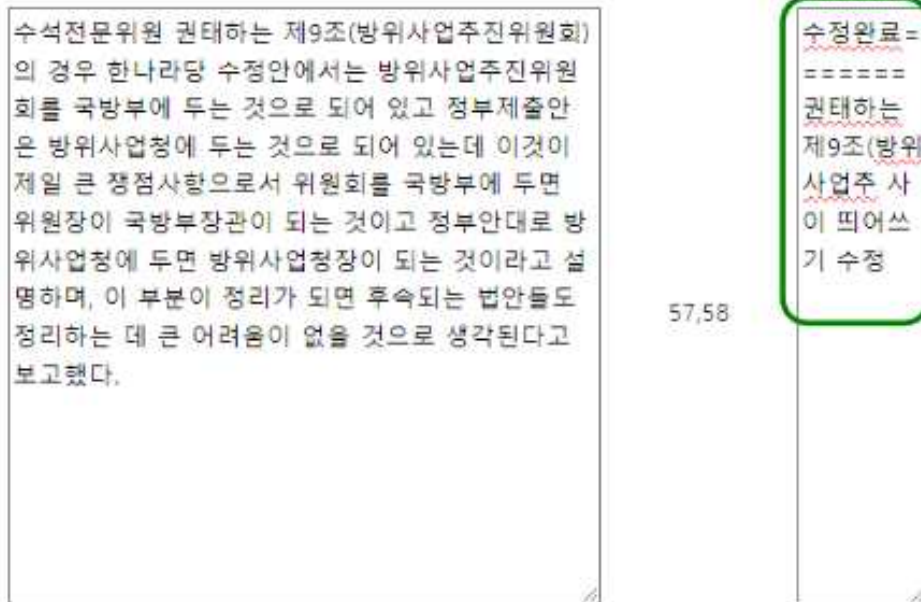


그림 56 회의록 요약문의 작성 오류 예시(띄어쓰기)

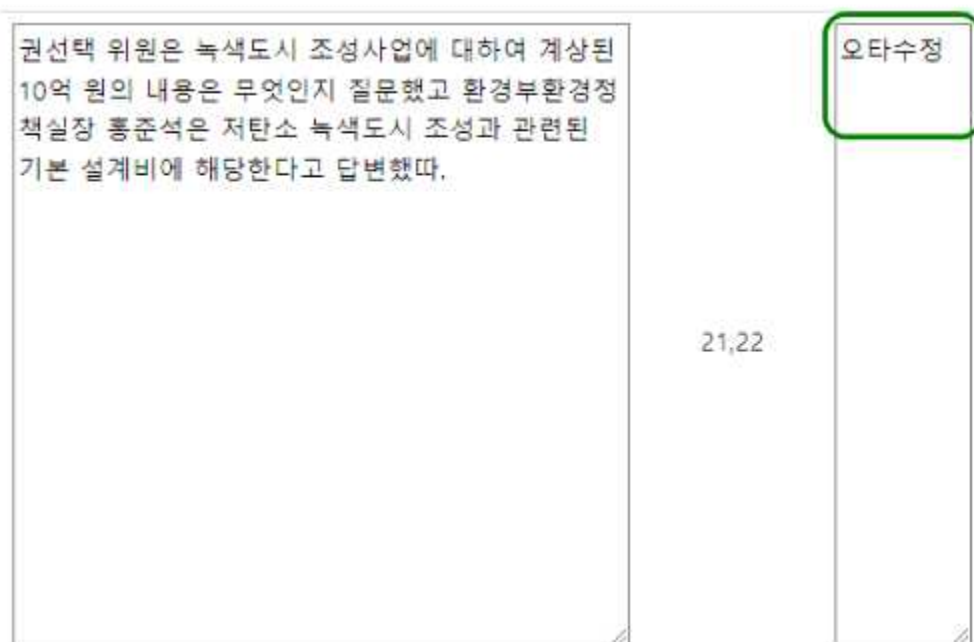


그림 57 회의록 요약문의 작성 오류 예시(오타)

2) 작성 누락 및 길이 오류

작성 누락 및 길이 오류는 요약문 자체나 요약문의 내용, 주어의 직책 등을 누락하는 경우나 요약문의 길이를 과도하게 길거나 짧게 작성하는 경우를 뜻한다.

표 48의 대응 항목과 같이 후속 조치, 결정 사항이 있는 경우에는 반드시 그에 해당하는 요약문을 추가로 작성해야 하며 모든 요약문은 독립성을 갖도록 작성해야 하기 때문에 구체적으로 작성하여야 한다. 또 요약문의 길이가 너무 긴 경우에는 발화를 나누어 작성하고 너무 짧은 경우에는 다른 발화와 함께 요약하여야 한다.

유형	내용	대응
요약문 누락 오류	후속 조치 또는 결정 사항을 작성해야 하는 데 작성하지 않은 경우	후속 조치/결정 사항이 있는 경우 반드시 후속 조치/결정 사항을 추가하여 작성하도록 검수 노트(note) 작성 및 SNS 전체 공지
내용 누락	어떤 질의를 했는지/주체가 누구인지/삭제하는 내용이 무엇인지 등의 내용 누락	요약문의 독립성을 위해 간단하게라도 질의, 답변 내용 등을 작성하도록 검수 노트(note) 작성 및 SNS 전체 공지
주어 직책 누락	주어 직책을 누락하고 작성한 경우	직책에 따라 주어 작성 기준이 바뀌므로 주의하여 지침에 맞게 작성하도록 검수 노트(note) 작성 및 SNS 전체 공지
요약문 길이	길이가 너무 짧거나 긴 경우	요약문 길이(특히 P/I)가 과도하게 짧거나 길지 않도록 검수 노트(note) 작성 및 SNS 전체 공지

표 48 회의록의 작성 누락 및 길이 오류 유형

다음은 교차 검수 시 작성한 작성 누락 및 길이 오류 유형의 요약문 및 검수 노트(note) 예시이다. 검수자의 피드백을 확인한 후 해당 내용을 수정하였다면 반드시 검수 노트(note)에 ‘수정 완료’를 작성하여 검수자에게 수정하였음을 알리도록 하였다.

<p>박대해 위원은 녹색도시 조성사업에 대한 자료는 언제 만든 것인지 질문했고 조원지 위원과 다시 자료를 만든 것으로 판단된다고 지적했고 호나경부환경정책실장 홍준석은 아니라고 답변하며 내부 보고 자료로 만든 자료라고 설명했고 환경부차관 이병욱은 11월에 마련된 자료라고 설명했다.</p>	<p>108,109,110</p>	<p>**전체적으로 ~했고 ~했고 로 접속사 중첩사용으로 인한 비문이 많습니다. 해당 내용은 전부 수정해주세요**</p>
--	--------------------	---

그림 58 회의록 요약문의 작성 오류 예시(비문)

<p>송영선 위원은 정부부처를 불러놓고 조목조목 따질 것이 아니라 소위끼리 의견을 먼저 조율하는 것이 더 바람직할 것 같다고 말했고, 임종인 위원은 정부가 내놓은 법안이기에 때문에 정부의 의견을 많이 들어야 될 것 같다는 의견을 밝혔다.</p>	<p>39,40</p>	<p>수정완료 = ===== ==요약문 이 짧으니 아래와 합 쳐서 요약 요망</p>
--	--------------	--

그림 59 회의록 요약문의 작성 오류 예시(길이)

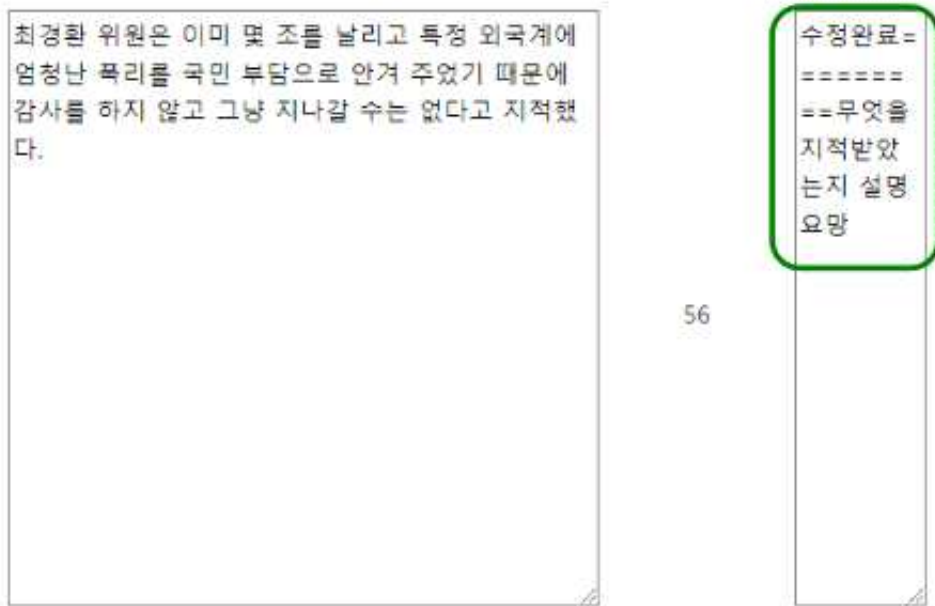


그림 60 회의록 요약문의 작성 오류 예시(작성 누락)

□ 보완 조치

회의록 분석 지침 외에도 자주 발생하는 오류를 바탕으로 교차 검수를 진행하였으며 사업 종료 후 보완 기간에도 이를 바탕으로 검수 및 수정을 진행할 예정이다. 지침 외에 추가로 참고할 수 있도록 오류 내용과 예시를 다음과 같이 정리하였다.

- 1) 원문의 법안명, 이름, 직책 등의 대명사를 요약문에 작성 시 오류가 아니면 원문의 띄어쓰기, 표현 그대로 사용
- 2) 원문의 표현이나 띄어쓰기가 오류인 경우 올바른 표현으로 변경하여 작성
- 3) 높임말, 낮춤말, 구어체는 수정하거나 제외하고 작성
- 4) 기본 띄어쓰기, 맞춤법 준수

원문	틀린 예	옳은 예
교육부제1과과장v홍길동	교육부v제1과v과장v홍길동	교육부제1과과장v홍길동
2002년도예산안	2002년도v예산안	2002년도예산안
수석전문위원 : 제가 보고드린것처럼~	수석전문위원이보고드린것처럼~	수석전문위원이 보고한 것처럼
수석전문위원 : 홍길동 위원님께서말씀하신대로~	수석전문위원은 홍길동 위원님께서말씀하신대로~	수석전문위원은 홍길동 위원이 말한 대로
홍길동 위원 : 지역안전도 v라고 하는것이~	홍길동 위원은 ~법안과 관련하여 지역안전도v라고 하는것이~	홍길동 위원은 ~법안과 관련하여 지역안전도라는 것이~/ 지역안전도라고 하는 것이
홍길동 위원 : ~이 가능하다v라고 되어 있습니다.	홍길동 위원은 ~법안은 ~이 가능하다v라고 되어 있다고설명했다.	홍길동 위원은 ~법안은 ~이 가능하게 되어 있다고 설명했다./ '~가능하다'라고 되어 있다고~
홍길동 위원 : ~법안은 실행이 어렵습니다. ~ 때문에	홍길동 위원은 ~법안은 실행이 어려우며 그 이유는 ~ 때문에v라고말했다.	홍길동 위원은 ~법안은 실행이 어려우며 그 이유는 ~ 때문이라고 말했다. / ~법안은 ~ 때문에 실행이 어렵다

		고 말했다.
홍길동 위원 : ~법안은 ~~하기 때문에 그러니까 이렇게 해 가지고 통과를 하기가 어렵다 이 말입니다.	홍길동 위원은 ~법안에 대해 ~~하기 때문에 이렇게 해 가지고 통과를 하기가 어렵다고 주장했다.	홍길동 위원은 ~법안에 대해 ~하기 때문에 이런 방법으로 는 통과를 하기가 어렵다고 주장했다.
홍길동 위원 : 독일식 연동제 또는 일본식 병립제 냐 하는 것에 대해서~~	홍길동 위원은 ~법안과 관련하여 독일식 연동제 또는 일본식 병립제 냐 하는 것에 대해서~~	홍길동 위원은 ~법안과 관련하여 독일식 연동제인지 일본식 병립제인지에 대해서~~
홍길동 위원 : ~사항은 ~처럼 수정 하자 하는 애입니다.	홍길동 위원은 ~사항은 ~처럼 수정 하자 하는 애기라고 설명했다.	홍길동 위원은 ~사항을 ~처럼 수정 하자고 설명했다. ~처럼 수정 하자 는 애기라고 설명했다.
기타 띄어쓰기	재논의 v 하는 논의 v 하자고 논의 v 했고/됐고 논의 v 되고 검토 v 되었고 해야하는데 제 v 356회 제 v 3항 안전도 v 라고	재논의하는 논의하자고 논의했고/됐고 논의되고 검토되었고 해야 v 하는데 제356회 제3항 안전도라고

표 49 회의록 지침 외에 참고할 오류 유형 및 예시

다음은 사업을 수행하면서 진행한 교차 검수 및 수정 일정이며 회의록 분석 말뭉치 역시 사업 종료 후 재검토 및 수정을 통해 보완할 예정이다.

구분	기간	인원	비고
1차 교차 검수	2023. 8. 28.~2023. 9. 1.	테디썸 내부 인력 및 회의록 작업자 4명	완료
1차 수정	2023. 9. 4.~2023. 9. 8.	회의록 작업자 4명	

2차 교차 검수	2023. 11. 13.~2023. 11. 17.	테디섬 내부 인력 및 회의록 작업자 4명	
2차 수정	2023. 11. 20.~2023. 11. 29.	회의록 작업자 4명	
사업 종료 후 재검토 및 수정	2023. 12. 1.~	테디섬 내부 인력	예정

표 50 회의록 교차 검수 및 보완 일정

3.2.3 회의록 요약 분석 말뭉치 형식

□ 회의록 요약 분석 말뭉치의 메타 데이터

표 51은 회의록 요약 분석 말뭉치의 메타 데이터 구조를, 표 52는 실제로 분석에 사용된 메타 데이터의 실제 예시를 보여준다.

* 이때, 화자(speaker)의 original_id는 한자-한글 변환 적용 전 원문 그대로의 이름(즉, 한자 또는 한글)이며, id는 한자-한글 변환 적용 후의 이름(즉, 모두 한글)이다.

1수준	2 수준	3 수준	4 수준	설명
id				* 요약 분석 말뭉치 고유 아이디 - SBSC2300000220
metadata				* 파일의 메타 정보
	title			* 국회 회의록 요약 분석 말뭉치의 경우 "국립국어원 국회 회의록 원시 말뭉치 추출 SBSC2300000220"
	creator			* 생성자: 국립국어원
	distributor			* 배포자: 국립국어원
	year			* 작업세트 생성 연도 - 2021
	category			* 분류, 분류 단계는 '>'로 구분. 예) "신문 > 전국 종합지" - 국회 회의 요약 분석 말뭉치의 경우 "회의록 > 국회소위원회 > 豫算案調整小委員會"로 기재함
	annotation_level			* 국회 회의 요약 분석 말뭉치의 경우 "요약 분석"으로 기재함
	sampling			* 국회 회의 요약 분석 말뭉치의 경우 "본문 전체"로 기재함
document				* 문서 정보
	id			* 요약 분석 말뭉치의 경우 문서는 1개만 존재하는 것으로 간주함. 예) SBRW2100000001.1
	metadata			* 문서의 메타 정보
		title		* 문서 제목 - 제33회국회(임시회) 보건복지위원회회의록(법안심사소위원회) 제5호
		author		* 국회 회의 요약 분석 말뭉치의 경우 "국회사무처"
		publisher		* 국회 회의 요약 분석 말뭉치의 경우 "국회사무처"
		date		* 회의록 작성일 - 20150706
		topic		* 주제 - 국회 회의 요약 분석 말뭉치의 경우 "1. 2018년도 예산안(계속) > 가. 여성가족부 소관 / 2. 2018년도 기금운용계획안(계속) > 가. 양성평등기금, 나. 청소년육성기금"과 같이 기재함
		speaker		
			id	* 발화자 ID - 이명수(회의록: 한자 변환 적용 후)

			occupa tion	* 직업 - 회의록에 기재된 직위/직책: 소위원장
			original _id	* 발화자 ID - 李明洙(회의록: 한자 변환 적용 전)

표 51 회의록 요약 분석 말뭉치의 메타 데이터 구조

```

{
  "id": "SBSC2300000277",
  "metadata": {
    "title": "국립국어원 국회 회의록 원시 말뭉치 추출 SBSC2300000277",
    "creator": "국립국어원",
    "distributor": "국립국어원",
    "year": "2023",
    "category": "회의록 > 국회소위원회 > 小委員會",
    "annotation_level": [
      "요약 분석"
    ],
    "sampling": "본문 전체"
  },
  "document": [
    {
      "id": "SBRW2100000277.1",
      "metadata": {
        "title": "第222回國會(臨時會)財政經濟委員會會議錄(小委員會)第1號",
        "author": "國會事務處",
        "publisher": "國會事務處",
        "date": "20010620",
        "topic": "13. 國有財産法中改正法律案(장영신 의원 발의) / (중략) 33. 自由貿易  
地域의 지정 등에 관한 法律中改正法律案(이협의원, 산업자원위원회)",
        "speaker": [
          {
            "id": "강원순",
            "occupation": "재정경제부국유재산과장",
            "original_id": "姜元淳"
          },
          {
            "id": "김기태",
            "occupation": "재정경제부소득세제과장",
            "original_id": "김기태"
          }
        ]
      }
    ]
  }
}

```

표 52 회의록 요약 분석 말뭉치의 메타 데이터 예시

□ 발화 정보

각 문서의 발화 내용은 sentence에 기록되어 있으며 발화를 word 단위로 나눈 후 위치 정보와 문자열 정보를 모두 기록하였다. 또한 각 발화의 ID값은 원시 말뭉치의 발화 ID와 동일하게 적용되었다.

표 53에서 회의록 요약 분석 말뭉치의 발화 정보 구조를, 표 54에서 실제로 분석에 사용된 문서의 요약 분석 말뭉치 발화 정보 예시를 확인할 수 있다.

2 수준	3 수준	4 수준	설명
sentence			<ul style="list-style-type: none"> * 문장 - 회의록을 포함한 구어의 경우 원시 말뭉치의 utterance
	id		<ul style="list-style-type: none"> *문장 ID - 회의록을 포함한 구어의 경우 원시 말뭉치의 utterance > id에 해당함 - SBRW2100000001.1.1.1 (document 내 첫 번째 발화) - SBRW2100000001.1.1.2 (document 내 두 번째 발화)
	form		<ul style="list-style-type: none"> * 문장 정보 - 회의록을 포함한 구어의 경우 원시 말뭉치의 utterance > form에 해당함 - 예) "form": "바쁘신 중에도 이렇게 회의에 참석해 주신 여러분들께 감사를 드립니다."
	word		* 어절 정보
		id	* 어절 ID. 문장 내 순서로 1부터 시작
		form	<ul style="list-style-type: none"> * 어절 - 예) "form": "바쁘신"
		begin	<ul style="list-style-type: none"> * 어절의 문장 내 시작 위치 (UTF-8 문자 위치로 0부터 시작) - "begin": 0
		end	<ul style="list-style-type: none"> * 어절의 문장 내 끝 위치 (UTF-8 문자 위치로 0부터 시작) - "end": 3

표 53 회의록 요약 분석 말뭉치의 발화 정보 구조


```

"sentence": [
  {
    "id": "SBRW2100000277.1.1.1",
    "form": "그러면 제222회국회 재경위 제1차 소위원회를 개의합니다.",
    "speaker_id": "안택수",
    "word": [
      {
        "id": 1,
        "form": "그러면",
        "begin": 0,
        "end": 3
      },
      {
        "id": 2,
        "form": "제222회국회",
        "begin": 4,
        "end": 11
      },
      {
        "id": 3,
        "form": "재경위",
        "begin": 12,
        "end": 15
      },
      {
        "id": 4,
        "form": "제1차",
        "begin": 16,
        "end": 19
      },
      {
        "id": 5,
        "form": "소위원회를",
        "begin": 20,
        "end": 25
      },
      {
        "id": 6,
        "form": "개의합니다.",

```

		"begin": 26,
		"end": 32
	}	
]		
},		

표 54 회의록 요약 분석 말뭉치의 발화 정보 예시

□ 요약 정보

회의록 문서에 대한 요약문은 전체 요약문과 주제별 요약문으로 구분된다. 전체 요약문은 회의록 문서 전체에 대한 요약문으로 해당 회의에서 다루어진 안건(쟁점)들에 대한 요약 정보가 포함되어 있으며 각 안건에 대한 주요 논의 사항, 의결 정보 등이 추상 요약되어 있다.

주제별 요약문은 회의록 문서 내의 주제별 중심어와 주제별 요약문(Abstract)을 포함한 세부 요약문들의 집합으로 구성되어 있다. 표 56은 본 사업에서 실제로 분석한 말뭉치의 요약 정보 예시로, “issue_summary > issue > topic”에 해당하는 “13. 國有財産法中改正法律案(장영신 의원 발의) 14. 國有財産法中改正法律案(이성현 의원 대표발의)(이성현·김동욱·김덕룡·박혁규·임진출·윤여준·정인봉·이윤성·안상수·이주영·안영근·김명섭·김부겸·원희룡·김용갑·서상섭·이부영·박관용·박시균·이호웅·김경천·김민석·안택수·박세환·이상배·김종하·심재철·손태인·강삼재·정병국·오세훈·김성호 의원 발의)”이 이 회의의 안건이다. 또 “issue_summary > issue > keyword”에 이 회의의 중심어(Keyword) “국유재산법중개정법률안”이 기록되어 있으며 해당 회의에 대한 주제별 요약문이 “summary > abstract > form”에 제시되어 있다. 그리고 해당 요약문을 작성할 때 참고한 발화들의 목록이 “summary > abstract > reference”에 나열되어 있다.

또한 해당 주제 요약문에 대한 세부 요약문(문제 제기(Problems/Issues), 결정 사항(Decisions), 후속 조치(Actions))들 역시 “summary > sub”에 리스트 형태로 들어가 있으며, 각 세부 요약문의 유형(type) 및 참고한 발화의 목록 역시 같은 형태로 저장되어 있다.

2 수준	3 수준	4 수준	5 수준	6 수준	설명
SC					* 대화 요약 분석 정보
	main_summary				* 대화 전체에 추상 요약 분석 정보
	issue_summary				* 주제별 요약 분석 정보
		issue			* 요약 대상 문서의 주제별 정보 - 국회 회의 요약 분석 말뭉치의 경우, 분할된 주제 각각에 대한 정보를 의미함
			topic		* 안건에 대응하는 원시 말뭉치의 주제(topic) - "topic": "2. 2018년도 기금운용계획안(계속) > 나. 청소년육성기금"
			keyword		* 해당 주제를 대표하는 중심어(keyword) - 예) "법원조직법 일부개정법률안"
			sentence_id		* 중심어가 등장한 발화의 id - 예) SBRW210000023.1.1.2
			begin		* 중심어의 문장 내 시작 위치 (UTF-8 문자

					위치로 0부터 시작)
			end		* 중심어의 문장 내 끝 위치 (UTF-8 문자 위치로 0부터 시작)
		summary			* 주제별 요약 정보
			abstract		* 해당 주제의 추상 요약 정보
				form	* 주제별 요약문
				reference	* 해당 주제의 최종 요약문 작성 시 참고한 문장들. - document > sentence > id에 해당함(예. ["SBRW2100000001.1.1.1", "SBRW2100000001.1.1.15", "SBRW2100000001.1.1.40"])
			sub		* 해당 주제의 세부 추상 요약 정보.
				type	* 해당 주제의 세부 요약문의 유형. (Problems/Issues, Decisions, Actions)
				form	* 해당 주제의 세부 요약문
				reference	* 쟁점의 세부 요약 문장 작성 시 참고한 문장들. - document > sentence > id에 해당함(예. ["SBRW2100000001.1.1.1", "SBRW2100000001.1.1.10", "SBRW2100000001.1.1.15", "SBRW2100000001.1.1.30", "SBRW2100000001.1.1.40", "SBRW2100000001.1.1.60"])

표 55 회의록 요약 분석 말뭉치의 요약 정보 구조

"SC": {

"main_summary": "본 회의는 제222회 국회 제1차 재경소위원회로 총 15개의 법안이 심사되었다. 먼저 의사일정 제13항, 제14항 국유재산법중개정법률안의 경우 국유지를 무단 점용하고 있는 사람들에게 대해 그동안 변상금에 대해 고지하지 않다가 갑자기 5년 치를 한꺼번에 부과하는 것은 사람들이 납득하기 어렵다는 위원들의 지적이 있었고 이에 재경부가 종합적인 검토 후 대안을 제시한 이후로 심의를 연기하는 것으로 결정되었다. 의사일정 제15항 및 제16항 금융실명거래및비밀보장에관한법률중개정법률안은 추가 법안을 제출하겠다는 한나라당의 요청으로 심의보류 되었다. 다음으로 의사일정 제17항 국세기본법중개정법률안은 국회 및 국정감사에서 과세정보를 요구하는 경우 납세자의 동의를 얻어 제공할 수 있게 하고 다만 사용자 측에 맡긴다는 제한 조건을 제안하는 내용으로서 실효성이 없기에 현행대로 했으면 한다는 재경부와 국회 요청이 있을 시 과세정보를 제공할 수 있게 더 규제해야 한다는 위원들의 견해차가 있었다. 의사일정 제18항 국세기본법중개정법률안은 "세무공무원은 납세자의 탈세에 관한 객관적이고 명백한 혐의 또는 자료가 있거나 관련 법률의 규정에 의하지 아니하고는 어떠한 명목으로도 이미 확정된 세금에 대해서 세무조사를 할 수 없다."는 내용을 신설하는 것으로서, 재정경제부는 다른 법을 보완하지 않고는 상충되는 문제가 있기에 관련 법률을 보완해서 정기국회 때 고치는 것으로 해 달라고 말했다. 의사일정 17항은 이후 다시 논의하는 것으로 하였고 의사일정 18항은 정기국회 때까지 정부 측에 시간을 주는 것으로 하였다. 다음으로 의사일정 제19항 소득세법중개정법률안은 장애인이 공교육기관 외의 사회복지법인에서 교육받을 때도 소득공제를 해주자는 내용으로서, 장애인들이 외국에 나가서 공교육을 받을 때도 혜택을 줄 필요가 있다는 데 의견을 모았고 의사일정 제20항 소득세법중개정법률안은 양도 세율 인하가 필요하다는 내용으로서 재경부 측은 구조조정지원 세제까지 추가로 검토하겠다고 말했으며 의사일정 제21항 법인세법중개정법률안은 현행 한국 최고 법인세율인 28%를 25%로 낮추자는 제안 내용으로서 이 역시 재경부 측에서 세율을 낮추는 방향에 대해 전반적으로 검토하겠다고 말했다. 의사일정 제19항부터 제21항은 정기국회에서 다시 심의하는 것으로 결정되었다. 의사일정 제22항 및 제23항 조세특례제한법중개정법률안은 전체 회의에 부의하지 않는 것으로 결정되었고 의사일정 제24항 조세특례제한법중개정법률안은 형평성 문제로 인해 계류하는 것으로 결정했다. 마지막으로 의사일정 제33항 자유무역지역의지정등에관한법률중개정법률안은 관세자유지역의지정및운영에관한법률의 설립취지 역할분담을 분명히 차별화해 놓은 산자부와 재경부의 기능상 균형을 깨트리는 위험이 있다는 데 의견을 모아 개정하지 않는 쪽으로 재경위원회의 의견을 제출하는 것으로 하고 의사일정 제25항 조세특례제한법중개정법률안은 수석전문위원의 검토 보고 및 정부 의견만 듣고 토론은 다음으로 연기하며 산회를 선포하였다.",

"issue_summary": [

{

"issue": {

"topic": "13. 國有財産法中改正法律案(장영신 의원 발의) 14. 國有財産法中改正法律案(이성현 의원 대표발의)(이성현·김동욱·김덕룡·박혁규·임진출·윤여준·정인봉·이운성·안상수·이주영·안영근·김명섭·김부겸·원희룡·김용갑·서상섭·이부영·박관용·박시균·이호웅·김경천·김민석·안택수·박세환·이상배·김중하·심재철·손태인·강삼재·정병국·오세훈·김성호 의원 발의)",

"keyword": "국유재산법중개정법률안",

```

        "sentence_id": "SBRW2100000277.1.1.3",
        "begin": 43,
        "end": 54
    },
    "summary": {
        "abstract": {
            "form": "의사일정 제13항, 제14항 국유재산법중개정법률안이
(중략)
심사되었다. 국유지를 무단점용하고 있는 사람들에 대해 그동안 변상금에 대해 고지하지 않다가
갑자기 5년 치를 한꺼번에 부과하는 것은 사람들이 납득하기 어렵다는 위원들의 지적이 있었고
재정경제부는 총체를 파악하고 정리해서 감면 등에 대한 대책을 만들어 다시 보고하겠다고 말했
다. 의사일정 제13항, 제14항 국유재산법중개정법률안은 재경부가 종합적인 검토를 거쳐 대안을
제시한 이후로 심의를 연기하는 것으로 결정되었다.",
            "reference": [
                "SBRW2100000277.1.1.3",
                "SBRW2100000277.1.1.21",

                "SBRW2100000277.1.1.115",
                "SBRW2100000277.1.1.125"
            ]
        },
        "sub": [
            {
                "type": "Problems/Issues",
                "form": "정의화 위원은 장영신 위원의 발의한 내용이 이성
현 위원의 발의한 것과 같은 내용인데 이성현 위원 것은 완전히 면제가 아닌지 물었다.",
                "reference": [
                    "SBRW2100000277.1.1.4"
                ]
            }
        ]
    },

```

표 56 회의록 요약 분석 말뭉치의 요약 정보 예시

3.3 기타(검수 프로그램 개발)

사업 수행 중에 검수자와 사업 담당자가 실시하는 검수는 작업자의 작업과 동시에 데이터 구축 플랫폼을 통해 작업자와 소통하며 진행한다. 사업이 종료된 후 개선·보완 기간에 수행하는 검수는 국립국어원의 검토 의견과 자주 발생하는 오류 사항을 바탕으로 데이터를 뽑아 일괄 수정(기계), 수기 수정 목록을 작성하여 진행하게 되는데 이때 적지 않은 분석 데이터를 보다 정확하고 효율적으로 검수하기 위해 그림 61과 같은 검수 프로그램을 개발하였다.

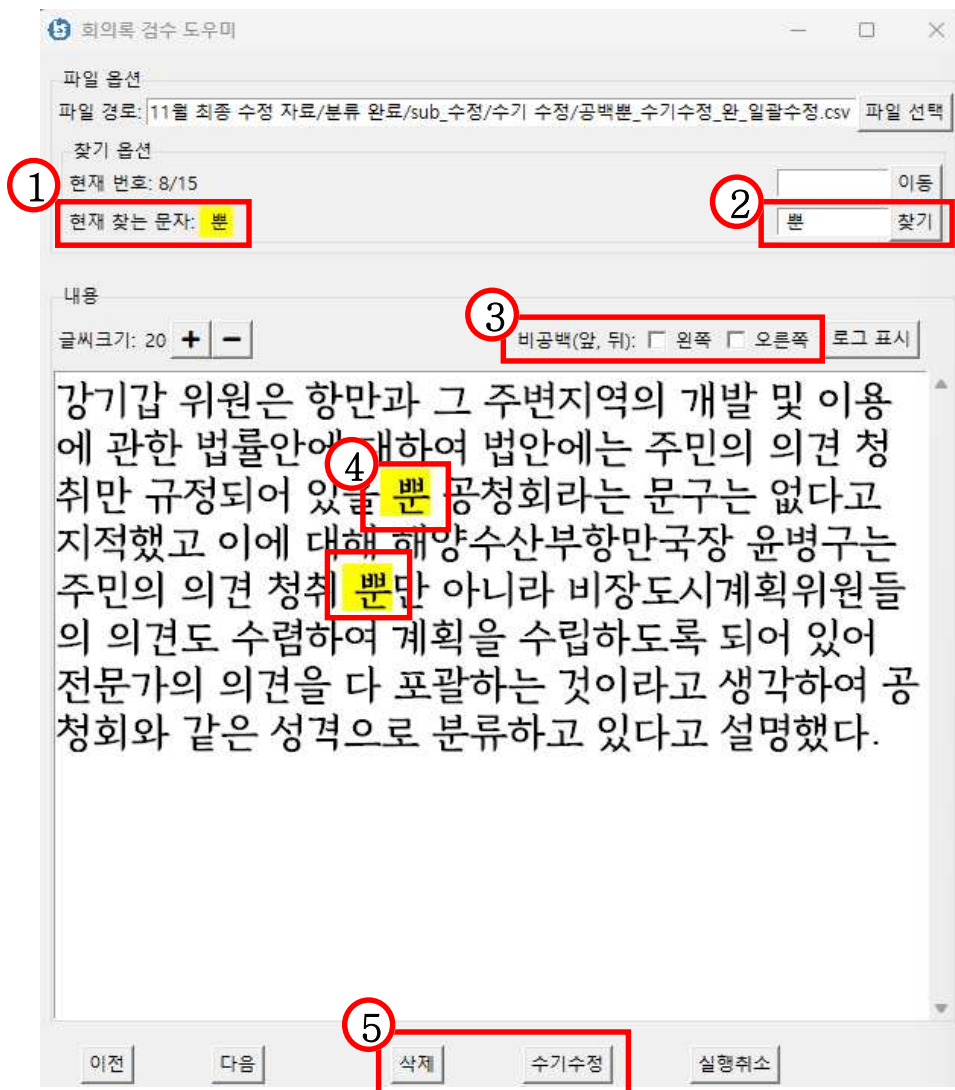


그림 61 대화 요약 검수 프로그램 화면

검수 프로그램의 기능을 간단히 살펴보면, 2번의 찾기 기능을 통해 공백을 포함한 오류 내용을 검색하고 1번에서 찾는 내용을 확인할 수 있다. 이렇게 검색한 분석 말뭉치 내의 오류 내용 위치를 4번과 같이 한눈에 확인할 수 있으며 오류 내용을 확인한 후 화면 하단의 버튼을 통해 일괄 및 수기 수정 목록을 작성한다.

또한 3번 기능을 통해 오류 내용의 앞, 뒤 비공백까지 검색할 수 있는데 비공백이 포함된 오류는 ‘~는것(~는^v것‘의 오류)’, ‘본회의는~(‘본^v회의는‘의 오류)’ 등이 있다.

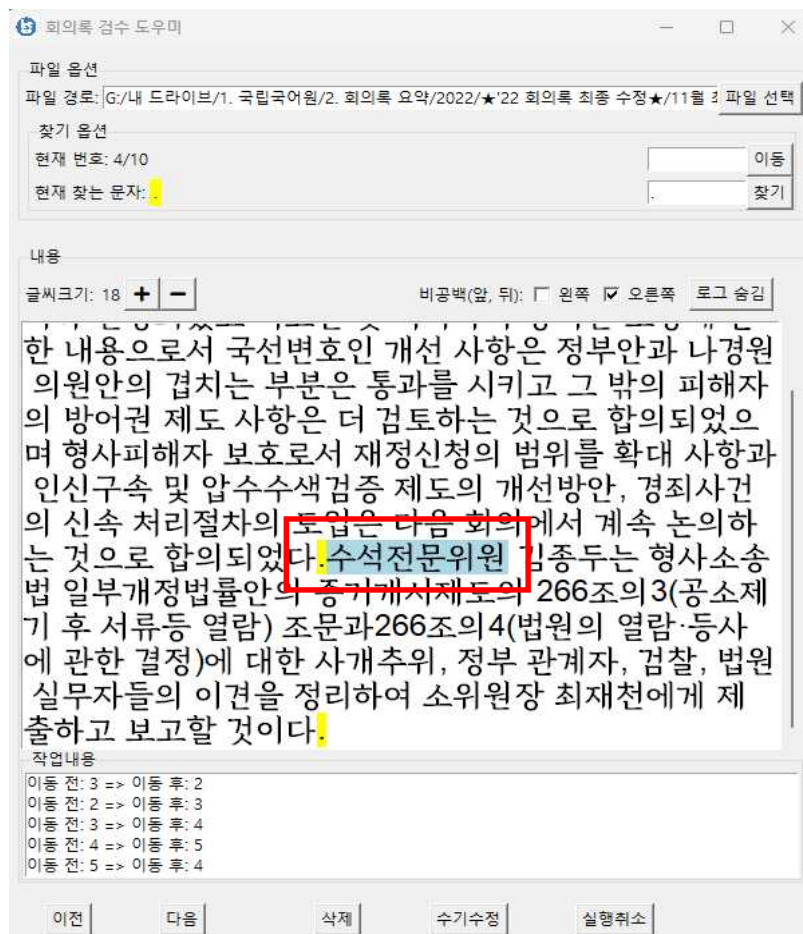
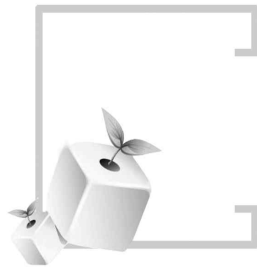


그림 62 대화 요약 검수 프로그램 화면(비공백 검색 예시)

비공백이 포함된 오류를 검색할 경우 그림 62와 같이 구분된 색으로 표시된다. 해당 화면은 2개 이상의 문장이 포함된 요약문에서 마침표 뒤에 공백이 없이 다음 문장이 이어진 경우로 마침표와 이어지는 문장 사이에 공백을 추가하여 수정하였다.



제 4 장

결론 및 추후 전망



4.1 프로젝트 요약

2023 대화 요약 말뭉치 연구 분석 사업의 개요를 그림 63, 64와 같이 정리하였다.

사업	일상 대화 요약	
목표	일상 대화 요약 분석 지침 수립 및 일상 대화 말뭉치 요약 분석(3,000건)	개선 사항 및 향후 계획
결과	<ul style="list-style-type: none"> - ‘일상 대화 요약 말뭉치 구축 지침’ 신규 수립 - 국립국어원이 제공하는 일상 대화 원시 말뭉치에 대한 요약 분석 말뭉치 3,000건 구축 (진행 중) - 일상 대화 쟁점 분할 및 분석 - 주제별 핵심 키워드 선택, 문장 및 화행 타입 주석, 중심 문장 간 관계 주석, 주제별/최종 요약문 작성 - 구축된 요약 분석 말뭉치에 대한 일관성 및 정확성 검증 방안 및 검증 체계 수립 및 제안 (교차검증 수행) 	<ul style="list-style-type: none"> - 작업자 및 관리자의 편의와 작업상의 오류 방지를 위한 플랫폼 기능 개선 - 11월 내에 완수하지 못한 데이터에 대해서는 12월 내로 작업 및 검수, 수정하여 제출 예정 - 데이터 구축을 통해 드러난 오류 유형 분석을 바탕으로 ‘일상 대화 요약 말뭉치 구축 지침’ 세밀화, 고도화 - 사업 종료 후 사전 계획한 데이터 보완 일정(12월~2월)과 국어원의 검토의견을 바탕으로 테디첨 내부 인력 및 기존 일상 대화 우수 작업자를 추가 투입하여 데이터 전수 검사 예정

그림 63 사업 개요(일상 대화 요약)

사업	국회 회의록 요약	
목표	논쟁적 대화 요약 분석 지침 고도화 및 국회 회의록 말뭉치 요약 분석(300건)	개선 사항 및 향후 계획
결과	<ul style="list-style-type: none"> - ‘논쟁적 대화 요약 말뭉치 구축 지침’ 고도화 - 국립국어원이 제공하는 국회 소위원회 회의록에 대한 요약 분석 말뭉치 300건 구축 - 회의록 쟁점 분할 및 분석 - 주제별 핵심 키워드 선택, 주제별 요약문 작성, 요약문 작성시 참조한 발화 명시, 최종 요약문 작성 - 구축된 요약 분석 말뭉치에 대한 일관성 및 정확성 검증 방안 및 검증 체계 수립 및 제안 (교차검증 수행) 	<ul style="list-style-type: none"> - 작업자 및 관리자의 편의와 작업상의 오류 방지를 위한 플랫폼 기능 개선 - 데이터 구축을 통해 드러난 오류 유형 분석을 바탕으로 ‘논쟁적 대화 요약 말뭉치 구축 지침’ 세밀화, 고도화 - 사업 종료 후 계획한 데이터 보완 일정과 국어원의 검토의견을 바탕으로 데이터 수정 - 데이터 보완은 지금까지 자주 발생한 오류 유형 분석 내용을 통해 기계적 일괄 수정과 비문, 맥락에 맞는 수기 수정으로 진행 예정

그림 64 사업 개요(회의록 요약)

4.2 일상 대화 요약 분석 말뭉치 구축 방향 제언

본 장에서는 일상 대화 원시 말뭉치 구축 및 후처리에 따른 특징을 확인하고 이를 토대로 양질의 데이터 구축을 위한 일상 대화 요약 분석 말뭉치 구축 방향을 제안하고자 한다.

□ 일상 대화 원시 말뭉치의 특징

국립국어원에서 제공한 ‘국립국어원 일상 대화 말뭉치 2020, 2021’에서 선정한 3,000건의 대상 문서는 평균 1,538어절, 발화 수 379개 내외 분량의 문서이다.

우선 특정한 세부 주제로 지정된 하나의 문서를 보다 효율적으로 분석하기 위해 소주제별로 발화를 분류하는 작업을 수행하였다. 이렇게 소주제 분할 작업을 통해 분류된 발화를 기준으로 요약 분석을 수행하면서 분석하기 어려운 여러 유형을 발견할 수 있었는데 그 내용은 다음과 같다.

1) 하나의 소주제에 발화자가 한 명인 경우

대부분의 일상 대화 데이터는 하나의 문서당 2~4개의 소주제로 이루어져 있으며 하나의 소주제에는 여러 발화자의 발화가 포함되어 있다. 그러나 하나의 소주제에 발화자가 한 명뿐인 경우도 있다. 이는 해당 발화자가 소주제에서 일부 벗어난 이야기를 하는 경우로 그 발화가 짧은 경우에는 다른 소주제에 포함시켜 분석하지만 발화가 긴 경우에는 별도의 소주제로 분할 후 분석하였다.

아래 예시는 제시된 각 소주제와 관계가 있다고 보기 어려운 소주제로 한 명의 발화자가 길게 발화하여 별도의 소주제로 분할한 경우이다.

€

이 데이터의 경우 스포츠, 레저가 주제인데
소양강과 춘천에 이사 와서 좋은 점,
병원에 간 에피소드에 대해
길게 하고 있어 별도로 주제분할하였음

그림 65 일상 대화 원시 말뭉치 중 하나의 소주제 안에 발화자가 한 명인 데이터 예시 1

이 데이터의 경우 즐겨 보는 방송과
미스터트롯 등이 주제인데
혼자 방탄소년단에 대한
이야기를 길게 하고 있어
별도로 주제분할하였음

그림 66 일상 대화 원시 말뭉치 중 하나의 소주제 안에 발화자가 한 명인 데이터 예시 2

2) 빈칸 개입이 많은 경우

일상 데이터 원시 말뭉치는 음성 파일을 전사한 데이터로 발화가 아닌 소리 (웃음, 기침 등)는 빈칸으로 표시되어 있다. 또 발화가 회의록 데이터와 같이 문장 단위로 끊어져 있지 않고 억양구로 끊어져 있기 때문에 발화자의 발화 습관이나 방식에 따라 짧은 발화 사이에 빈칸이 많이 개입되어 있는 원시 말뭉치가 다수 포함되어 있다.

SDHW2100000076 Title: 2인 일상 대화 (한국) Topic: 대중교통 > 서울대중교통 정책, 지방대중교통의 단점			SDHW2100000078 Title: 2인 일상 대화 (한국) Topic: 대중교통 > 서울대중교통 정책, 지방대중교통의 단점		
21	SD2100077	그리고 너는 좋은 게 부러워서라기 좋은 교통이 안 좋다는 게 좋은 게 싫어.	44	SD2100078	
22	SD2100077	근데 이게 적자라는 게도 계속 운영하고 있는 거 보면	45	SD2100078	통영들이 다 다르니까?
23	SD2100077		46	SD2100078	
24	SD2100077	운영비가 싼தாக도 못한 나는 좋아	47	SD2100078	그게는 좋은 거 같아 기차를 탈 때
25	SD2100078	그거에 대해서는 뭔가 아쉬움이 있고 할 수 있지	48	SD2100077	
26	SD2100078	너는	49	SD2100077	
27	SD2100078	이제 대중교통을 타고 매일 출퇴	50	SD2100077	그럼 기차가 좋아 비행기가 좋아?
28	SD2100078	퇴하 가 본 게 아니냐?	51	SD2100078	그 둘 다에 일단 많이 있는 거 같아.
29	SD2100077	제발 출퇴!	52	SD2100078	비행기는 빠르지만 후리가 막
30	SD2100077		53	SD2100077	
31	SD2100077	기차 타고 부산 갔을 때가 가장 편 가 같아.	54	SD2100078	
32	SD2100077		55	SD2100078	말론데
33	SD2100077	아	56	SD2100078	어떤 게 아니냐면 좀
34	SD2100077	기차 가는 게 좋아!			

그림 67 일상 대화 원시 말뭉치 중 빈칸 개입이 많은 데이터 예시

이 경우 한 명의 발화자의 발화를 묶어서 주석(문장, 화행 유형 주석)하기 어렵고, 분류되는 문장 주석 정보가 불필요하게 늘어난다. 빈칸 개입이 많은 데이터의 예시는 아래와 같다.

3) 발화가 너무 짧게 구분되어 있는 경우

역양구로 구분되는 일상 대화 원시 말뭉치의 특성상 발화자의 발화 습관 등에 따라 부득이하게 하나의 발화가 과하게 짧게 끊어져 있는 데이터가 다수 존재한다. 이 경우 앞뒤 발화를 연결하면 하나의 의미있는 발화로 주석할 수 있는데 구분된 대로 발화를 나누어 보면 그 내용을 파악하기 어려운 경우가 많다.

SDWW2100003816 Title: 4인 일상 대화 (오전, 동서리 지면) Topic: 음악 > tv음악프로그램, 좋아하는 가수 이야기			SDWW2100003816 Title: 4인 일상 대화 (오전, 동서리 지면) Topic: 음악 > tv음악프로그램, 좋아하는 가수 이야기		
19	SDQ103686	33번	45	SDQ103684	
20	SDQ103684		46	SDQ103686	
21	SDQ103685		47	SDQ103686	
22	SDQ103686		48	SDQ103687	노래를
23	SDQ103687	아 33호 얘기는 뭐가 하고 싶어요	49	SDQ103685	
24	SDQ103686		50	SDQ103687	노래를 듣고 있는데
25	SDQ103687	아	51	SDQ103686	
26	SDQ103687	33호 노래를	52	SDQ103687	내가 아 이 심사위원은 아니지만 심사위원들이 곡작성의 굉장히 강요해서 노래를 하나 씀니다
27	SDQ103686		53	SDQ103686	
28	SDQ103685		54	SDQ103686	
29	SDQ103687	아름다운 순간에	55	SDQ103686	
30	SDQ103686		56	SDQ103686	
31	SDQ103687	아	57	SDQ103687	아 현장에서 직접 들으면
32	SDQ103685		58	SDQ103687	저런 편곡선이 저렇게 나오면이나 싶을 정도로

그림 68 일상 대화 원시 말뭉치 중 발화가 너무 짧게 구분되어 있는 데이터 예시

□ 일상 대화 원시 말뭉치의 특징에 기반한 제안 사항

일상 대화 원시 말뭉치의 경우 문장 단위가 아닌 역양구 단위로 끊어져 있고 발화 사이에 빈칸 또는 의미 없는 짧은 발화가 상당수 존재한다. 또 회의록과 같이 떨어져 있는 각각의 발화를 다수 선택하거나 소주제 분할 단계에서 불필요한 발화를 제외하지 않기 때문에 문장 주석을 하기 어렵다는 특징이 있다. 그리고 일상적 주제를 바탕으로 한 자연스러운 대화이기 때문에 주제와 관련된 핵심 내용을 분별할 때 작업자 개인의 성향이 반영되어 분석 데이터의 일관성을 확보하기 어렵다.

이러한 일상 대화 원시 말뭉치의 특징을 바탕으로 다음의 몇 가지 사항을 제안하고자 한다. 일상 대화 원시 말뭉치는 음성 전사를 통해 만들어지며 음성 전사를 하는 경우 부득이하게 ‘웃음’이나 ‘기침’ 등 발화가 아닌 소리가 개입될 수밖에 없다. 발화가 아닌 소리가 개입되는 경우에는 빈칸이 발생하는데 후처리를 통해 빈칸을 제외할 수 있다면 많은 경우 빈칸으로 떨어져 있던 발화를 하나로 묶어 유의미한

발화 정보 데이터로 활용할 수 있을 것이다.

후처리를 통해 빈칸을 제외하는 방법을 적용하기 어렵다면 데이터 구축 작업의 소주제 분할 단계에서 회의록과 같은 방법을 적용할 수 있다. 소주제 분할 단계에서 빈칸이나 ‘아’, ‘근데’와 같은 의미 없는 발화를 사전에 제외한다면 원시 말뭉치가 억양구 단위로 끊어져 있어도 중요한 발화를 하나로 묶어서 주석할 수 있다.

일상 대화는 국회 회의록에 비해 상대적으로 발화자가 적고 화자별 요약문을 추가로 작성하기 때문에 문장 주석 단계에서 반드시 동일한 발화자의 붙어있는 발화를 하나로 묶도록 하고 있다. 그래서 상대 발화자의 의미 없는 발화(빈칸, 짧은 발화 등)가 개입되어 있는 경우에는 문장 주석의 수가 불필요하게 늘어난다. 동일한 발화자의 발화만 묶어서 주석하되 서로 떨어져 있는 발화도 선택하여 주석할 수 있도록 한다면 불필요하게 문장 주석의 수가 늘어나는 것을 방지할 수 있을 것이다.

마지막으로 작업자의 주관성을 교정하기 위해 작업 지침을 세밀화하는 것도 중요하지만 작업 지침을 충분히 이해하고 있는 검수자를 더 많이 확보하는 것이 중요하다. 많은 검수자가 계속해서 피드백을 주고 그 피드백을 바탕으로 작업 및 수정을 하다 보면 작업자의 주관성도 어느 정도 교정될 것이고 이를 통해 원시 및 분석 데이터의 일관성도 확보할 수 있을 것으로 기대한다.

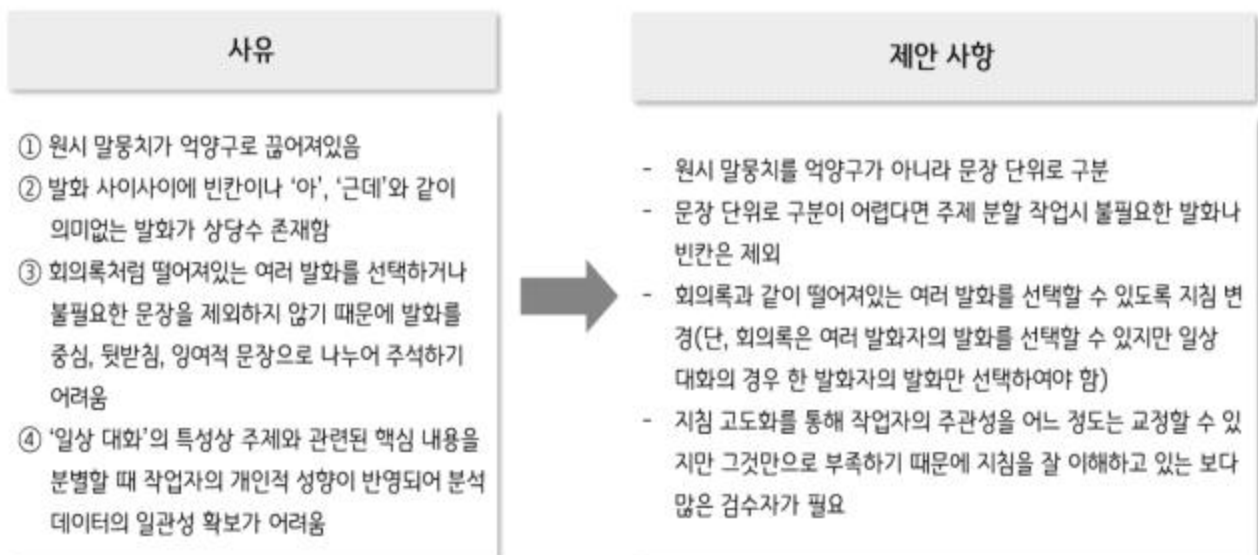


그림 69 일상 대화 원시 말뭉치의 특징에 기반한 제안 사항

4.3 향후 발전 가능성

본 사업팀은 사업을 수행하는 과정에서 여러 시행 착오를 경험했다. 특히 사업 초기에 필요한 자원과 일정을 정확하게 예측하지 못한 결과, 사업 일정이 일부 지연되고 추가 자원이 필요하게 되었다. 이로 인해 세밀하게 초기 계획을 수립하는 것이 데이터 구축 사업의 핵심적인 부분임을 명확하게 인식할 수 있었다.

또한, 자원 관리와 일정 계획 수립 단계에서 보다 철저한 예측과 위험 요소 관리가 필요하다는 점을 인지하였다. 이러한 교훈을 바탕으로 더욱 신중한 자원 할당과 일정 조정을 통해 예상치 못한 위험 요소를 고려하고 신속한 대처가 가능하도록 유연성을 높이는 방향으로 나아가고자 한다.

한편 사업 수행 과정을 통해 작업자가 보다 효율적으로 작업하고 작업 시의 오류를 미연에 방지할 수 있는 데이터 구축에 더욱 적합한 플랫폼을 개발할 수 있었다. 사업 계획 당시 제작한 데이터 구축 플랫폼을 바탕으로 데이터 구축 시에 필요한 기능을 추가 적용하였으며 개선 사항은 그림 70, 71과 같다.

로그 기록 확인 기능	작업자 계정별 로그 기록 확인 기능으로 작업자의 근무시간 관리 가능 *30분 단위로 확인 가능
오류 파일 교체 기능	원시 말뭉치의 오류 내용을 수정한 데이터를 개발자를 통하지 않고 관리자가 플랫폼 상에서 즉시 교체할 수 있는 기능 추가
방지, 제한 기능	①중심어 제한 : 중심어는 3어절 이하의 명사구로 작성해야 하므로 그 이상의 단어 조합으로 이루어진 중심어는 작성 후 저장될 수 없도록 공백이 3칸 이상의 키워드는 저장되지 않는 기능 추가 ②문장, 화행 타입 주석, 중심어 누락 방지 : 문장, 화행 타입 주석, 중심어를 누락한 경우 경고 팝업이 뜨고 저장되지 않는 기능 추가
개별 저장 기능	발화 선택 후 문장 및 화행 타입 주석시 개별 저장 기능 추가
작업 파일 변경/추가 기능	작업자별 작업 파일 변경 및 추가 기능 *작업자, 작업 상태, 마감일, 검수자 등 변경 가능

그림 70 일상 대화 요약 플랫폼 기능 개선 항목

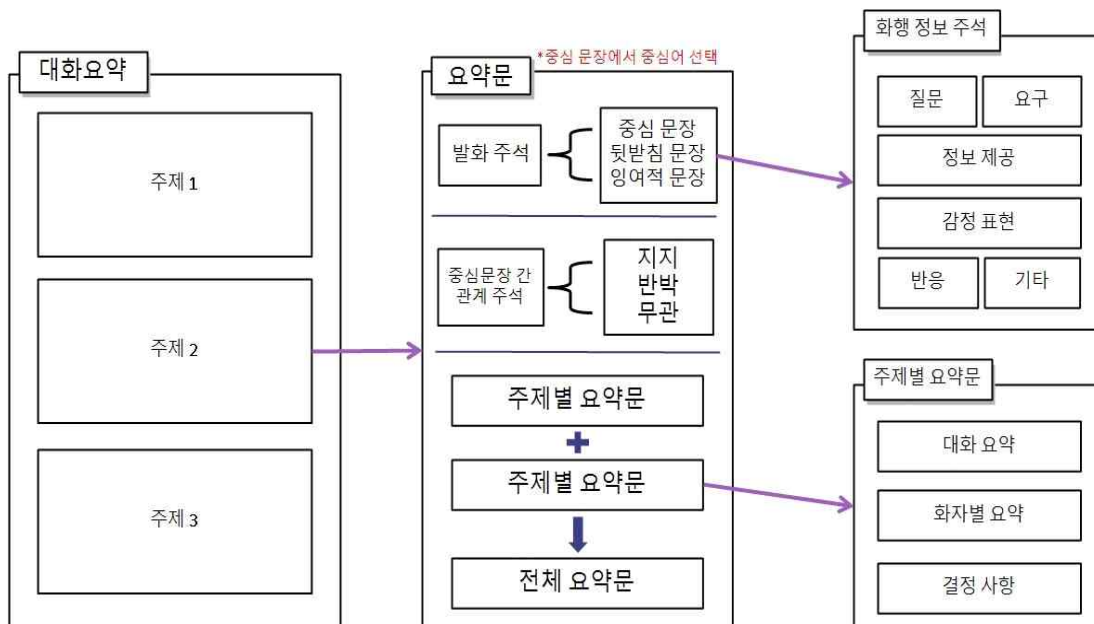
로그 기록 확인 기능	작업자 계정별 로그 기록 확인 기능으로 작업자의 근무시간 관리 *30분 단위로 확인 가능
검수 노트 기능 추가	최종 요약문 화면에 검수자가 검수 내용을 기재할 수 있도록 검수 노트 추가
중심어 기능 개선	법안명이 긴 경우 키워드가 입력되지 않는 문제 개선을 위해 플랫폼 내 키워드 선택의 입력 가능 글자 수를 100자 이상으로 확대
스크롤 기능 개선	요약문 저장, 삭제 작업 후 이전 요약작업 위치로 화면 회귀 기능 추가
방지, 제한 기능	요약문을 작성하지 않고 저장하는 경우 경고 팝업이 뜨고 저장이 되지 않는 기능 추가

그림 71 회의록 요약 플랫폼 기능 개선 항목

또한 데이터 구축 과정에서 신규 수립한 지침 외의 예상하지 못한 다양한 오류가 발생하였는데 이를 바탕으로 오류 유형을 분석하여 요약 작업 지침의 세밀화, 고도화가 가능할 것이다. 이렇게 대화 요약 말뭉치 연구 분석 사업을 수행하며 겪은 시행착오와 교훈을 토대로 더 철저하고 유연한 계획을 수립하고 가공 단계에서의 효율성을 높여 보다 현실적이고 수행 가능한 전략 수립이 가능해질 것으로 기대한다. 그리고 본 사업을 통해 구축된 데이터는 일상 대화 및 논쟁적 대화 요약 모델의 다양한 평가 지표를 제공하고, 대화 요약 기술의 발전과 향상을 위해 중요한 자원으로 활용될 것으로 기대한다.

일상 대화 요약 작업 지침

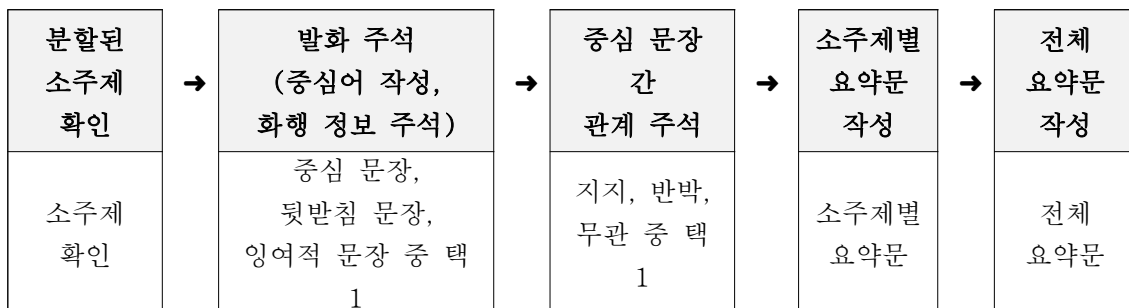
1. 일상 대화 요약 작업 개요



가. 작업 개요

1) 구성

가) 작업 순서



2) 소주제 분할 : 대화 소주제에 따른 발화 분류 수행

3) 발화 주석(문장 주석)

가) 발화 종류(문장 주석) : 중심 문장, 뒷받침 문장, 잉여적 문장 중 택 1
* 문장 주석을 한 다음 반드시 1개의 화행을 주석해야 한다.

- ① 중심 문장 : 소주제와 관련된 핵심적인 주장이나 의견, 질문이 포함된 발화(들)
 - 중심 문장에 적절한 중심어가 있는 경우 단어 또는 명사구로 작성하되 반드시 발화에서 그대로 가져와(복사-붙여넣기) 작성한다.
- ② 뒷받침 문장 : 중심 문장에 대한 뒷받침/부연설명 발화
 - 선택한 뒷받침 문장이 어떤 중심 문장에 대한 뒷받침 문장인지 반드시 선택해야 한다.
- ③ 잉여적 문장 : 중심 문장, 뒷받침 문장이 아닌 모든 발화(들)

나) 화행 정보 주석 : 질문, 요구(요청), 정보 제공, 반응, 감정 표현, 기타 중 택 1
- 모든 문장에 반드시 1개의 화행을 주석해야 한다.

- ① 질문 : 예/아니오 질문, 의문사 포함 질문 등
- ② 요구(요청) : 요청, 명령, 제안 등등
- ③ 정보 제공 : 주장, 정정, 추측, 소망, 의지 등등
- ④ 반응 : 질문에 대한 답변, 동의 표현, 비동의 표현, 수락, 거부, 회피 등
- ⑤ 감정 표현 : 인사, 칭찬, 비판, 감사, 사과, 비난, 감탄사 등
- ⑥ 기타 : 위의 화행 구분에 들어가지 않는 그 외 모든 표현

4) 중심 문장 간 관계 주석 : 지지, 반박, 무관 중 택 1

가) 문장 간 관계 주석 요소

- ① 지지 관계 : 중심 문장 간 관계가 긍정적일 때 주석
(지지, 선호, 인정, 공감 등)
- ② 반박 관계 : 중심 문장 간 관계가 부정적이며 상충할 때 주석
(반박, 거절 등)
- ③ 무관 관계 : 중심 문장 간 관계가 성립하지 않는 경우일 때 주석

5) 소주제별 요약문 작성 : 중심 문장, 뒷받침 문장 등의 내용을 중심으로 소주제에 대한 요약문 작성

구분	내용
대화 요약	
화자 1 요약	
화자 2 요약	
결정 사항	(대화를 통해 결정 사항이 나온 경우에만 작성)

6) 전체 요약문 작성 : 소주제별 요약문의 ‘대화 요약’ 내용을 바탕으로 전체 요약문 작성

구분	내용
대화 요약	

나. 원시 말뭉치의 예시

화자	발화
SD210242 6	너는 지금 결혼 한 거에 대해서 되게 행복하다고 생각하니?
SD210242 5	후회 하지도 않고 행복하지도 않고 중간이라고 생각해 그게 예전에는 그게 몰랐는데 지금은 좀 나이가 먹어갈수록 이게 가족이 된 거 같다는 그거 있잖아 언니 그냥 옛날에 사랑했던 그런 게 아니라 지금은 가족 같은 분위기 그러니깐 들어오면 말도 별로 없고 그런 게 많이 많더라고 언니는 요즘에 그런 거 많이 안느껴?
SD210242 6	나 같은 경우에는 결혼 임신을 해서 결혼했기 때문에 그 처음에는 그 되게 힘들었어 그 결혼 생활 10년이 되게 힘들었거든 근데 나이가 먹으면서 안정권을 찾는 거 같아 신랑이랑 나랑 그리고 어느 순간에는 서로 이해하려고 되게 많이 노력했던 거 같아 결혼 10년은 엄청 많이 싸웠어 그러니까 내가 생각해도 나는 결혼하래 그때 10년은 난 결혼 안 한 단 소리를 했을 거 같아 그 그리고 지금 결혼 하고 나한테 그

	<p>진짜로 생각하고 싶지 않았던 정도로 그런데 지금은 어 내가 봤을 땐 서로 이해하려고 하려고 안 안정권이라고 생각해 그리고 부모 우리가 조금 편해야 자식들도 편하잖아 그렇지 그래서 나는 나름 지금 우리 서로 건강한 상태에서 이렇게 가면 나는 지금도 그냥 행복하다고 느껴</p>
SD2102425	<p>나도 언니랑 비슷한 거 같아 나도 처음에는 한 8년 동안은 되게 힘들었거든 그리고 나서 셋째를 낳고 그다음부터는 좀 안정기에 들었다 그래야 되나 좀 남편이 자식들도 챙기고 많이 저거 하면서 그다음부터 이제 좀 행복하게 좀 돌아다니고 많이 저거 했고 지금은 애들하고도 잘 지내는 거 같기도 하고 그냥 지금 만족스럽기는 해 많이 좋아하지</p>

다. 요약의 예시

구분	내용
대화 요약	이 대화에서 화자 SD2102425는 결혼 생활의 만족도를 중립으로 답변하였고, SD2102426은 결혼 초기의 어려움 극복 이후 가족 관계의 변화와 현재의 행복함에 대해 이야기했습니다. 또한 공통적으로 어려움을 겪었던 초반에 비해 현재는 안정되고 행복한 상태라는 이야기를 하였습니다.
SD2102425 요약	SD2102425는 결혼 생활의 만족도에 대해 중립적으로 답변하는 한편, 어려웠던 결혼 생활 초기에 비해 셋째 출산 이후 비교적 안정을 찾았음을 이야기했습니다.
SD2102426 요약	SD2102426은 결혼 초기의 어려움을 극복한 후 현재 행복함을 느끼며, 그 과정에서 가족과의 관계가 어떻게 변화하였는지에 대해 이야기했습니다.
결정 사항	(결정 사항 없음)

라. 작업 원칙

1) 기본 원칙

- ① 요약문은 분할된 소주제를 기준으로 작성
- ② 요약문은 ‘중심 문장’ 과 ‘뒷받침 문장’ 을 중심으로 작성하며, 필요시 ‘잉여적 문장’ 도 참고해 작성
- ③ 요약문만 읽고 대화의 소주제와 내용을 확인할 수 있도록 작성

- 어떤 주제가 다루어졌는지, 어떤 발화자의 어떤 의견이 있었는지, 해당 주제에 대해 어떻게 결정이 났는지 등을 확인할 수 있어야 한다.

2) 기본 내용

- ① 대화의 소주제(주제 분할)
- ② 대화를 구성하는 문장의 의도(문장 주석, 화행 정보 주석)
- ③ 각각의 소주제에 대한 대화 내용(소주제별 요약)
- ④ 전체 대화의 내용(전체 요약)

2. 작업 순서

가. 기본 작업 순서

- ① 소주제 분할 : 대화 소주제에 따른 분류 수행
- ② 발화 주석(문장 주석, 화행 정보 주석)
 - 발화 주석 : 중심 문장, 뒷받침 문장, 잉여적 문장 중 택 1
 - 중심어 작성 : 중심어는 소주제를 잘 나타내는 핵심 단어 또는 명사구로 1개 이상 작성해야 하며 반드시 중심 문장 내에서 그대로 가져와(복사-붙여넣기) 작성(‘중심어’는 최대 3어절의 복합 명사(구)까지 허용)
 - 화행 정보 주석 : 문장 주석 후 화행 선택
(질문, 요구(요청), 정보 제공, 반응, 감정 표현, 기타 중 택1)
 - 중심 문장-뒷받침 문장 연결 : 뒷받침 문장의 경우 어떤 중심 문장에 대한 뒷받침 문장인지 선택
- ③ 중심 문장 간 관계 주석 : 지지, 반박, 무관 중 택 1
- ④ 소주제별 요약문 작성 : 중심 문장, 뒷받침 문장 간의 관계를 활용하여 소주제에 대한 요약문 작성
- ⑤ 전체 요약문 작성 : 소주제별 요약문을 바탕으로 전체 요약문 작성

*작업 시 구글 크롬의 맞춤법 검사기능을 통해 표시되는 빨간 밑줄을 확인하며 오타와 문법적 오류에 주의하여 작성(자세한 방법은 플랫폼 작업 순서 참고)

나. 맞춤법 검사 기능 활용

1) 구글 크롬의 맞춤법 검사 기능 활성화(필수)

요약문 작성 시 구글 크롬의 맞춤법 검사 기능을 활용하여 오타 및 띄어쓰기 오류가 발생하지 않도록 주의한다.



2) 일상 대화 요약 작업 방법

① 로그인 - 작업 항목 선택



② 작업 현황 확인(작업자 화면)

teddysum | 2023학년도 1학기 대화 구조 플랫폼

작업현황 작업 작업자별 QA 한지현님

주간 작업현황

	의뢰건	작업 건	완료 대기	수정 필요	완료 비율
한지현	9	0	0	0	0 (0%)
박민지	9	0	0	0	0
김지현	9	0	0	0	0
복금희	9	0	0	0	0
이다영	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

The screenshot shows the 'Job Application' page on the teddydynam website. The page has a sidebar on the left with a navigation menu. The main content area displays a table of job applications. A red box highlights the 'Application Status' column, and a red arrow points to the 'Apply' button next to the application with ID 1234567890148. A red box also highlights the 'Apply' button next to the application with ID 1234567890148.

Application ID	Application Name	Application Type	Application Date	Application Status
123456789000000	대형 교재	리작성	2023-08-08	작업하기
123456789000001	대형 교재	리작성	2023-08-08	작업하기
123456789000002	대형 교재	리작성	2023-08-08	작업하기
123456789000003	대형 교재	리작성	2023-08-08	작업하기
123456789000004	대형 교재	리작성	2023-08-08	작업하기
123456789000005	대형 교재	리작성	2023-08-08	작업하기
123456789000006	대형 교재	리작성	2023-08-08	작업하기
123456789000007	대형 교재	리작성	2023-08-08	작업하기
123456789000008	대형 교재	리작성	2023-08-08	작업하기
123456789000009	대형 교재	리작성	2023-08-08	작업하기
123456789000010	대형 교재	리작성	2023-08-08	작업하기
123456789000011	대형 교재	리작성	2023-08-08	작업하기
123456789000012	대형 교재	리작성	2023-08-08	작업하기
123456789000013	대형 교재	리작성	2023-08-08	작업하기
123456789000014	대형 교재	리작성	2023-08-08	작업하기
123456789000015	대형 교재	리작성	2023-08-08	작업하기

tucdysum 2023년 하반기 데이터 구축 일정표				
주요 일정				
과일명	작업종류	작업현황	작업기간	작업비율
32HW2-10000408	주제 분석	작업 예시	-	확인하기
32HW2-10000811	주제 분석	작업 예시	-	확인하기



⑥ 화행 선택



⑦ 중심어 작성



⑧ 뒷받침 문장 - 중심 문장 연결

- 뒷받침 문장이 어떤 중심문장에 대한 뒷받침 문장인지 작성
 - 중심문장은 ID:숫자 라고 기재되어 있는 발화 ID의 숫자 입력
 - 뒷받침 문장은 중심문장의 바로 앞뒤에 붙어있지 않을 수도 있음
 예시) ID:35에 위치한 뒷받침 문장과 연결되는 중심문장이 ID:33일 수 있음

중심문장 아이디 선택 후 반드시 '확인' 버튼을 눌러 유효성 검증해야 함

⑨ 중심 문장 간 관계 주석

- 중심문장 간 어떤 관계인지 주석 (지지, 무관, 반박 중 택1)
 - 문장의 등장 순서대로 앞뒤 문장에 대해서만 주석

① 지지 관계 : 지지, 선호, 인정, 공감 등
 ② 반박 관계 : 반박, 거절 등
 ③ 무관 관계 : 관계가 성립하지 않는 경우

* 특히 질문과 답변의 문장 관계의 경우 기본적으로는 '무관'으로 주석하되, 명백하게 반박, 거절의 의사를 표현한 경우에는 '반박'으로 주석

⑩ 주제별 요약문 작성

- 대화 요약 : 대화에서 드러나는 주제와 그 주제에 대한 화자별 의견, 결정 사항을 명료하게 작성
 - 화자별 요약 : 주제에 대한 화자별 의견을 근거 등을 포함하여 구체적으로 작성
 - 결정 사항 : 대화를 통해 나온 결정 사항이 있는 경우 작성

* 분홍색 바탕 : 중심 문장
 노란색 바탕 : 뒷받침 문장
 회색 바탕 : 영역적 문장

[illegible]

주간 작업현황					
	마작업	작업종	접수 대기	수정 필요	최종 완료
연시간	9	0	0	0	0 00%
백업작	9	0	0	0	0
설치작	9	0	0	0	0
유동작	9	0	0	0	0
합계	0 0%	0 0%	0 0%	0 0%	0 0%

2023 제1회 태국어 대회 우수 문제집

[학업지원](#)
[학업](#)
[학업지원](#)
[Q&A](#)
[학업지원](#)

No.	날짜	일문자	질문 유형	제출 방법	발제인	질문 내용	답변
12	2023-01-07	태국어	기타	2023 제1회 태국어 대회 우수 문제집	태국어	태국어 시험 문제집을 4월 14일까지 제출하려고 합니다. 제출한 문제집을 4월 14일까지 제출한 학생들에게는 100%의 장학금을 지급합니다. 제출한 학생들에게는 100%의 장학금을 지급합니다.	태국어 시험 문제집을 4월 14일까지 제출한 학생들에게는 100%의 장학금을 지급합니다.
13	2023-01-12	태국어	기타	2023 제1회 태국어 대회 우수 문제집	태국어	태국어 시험 문제집을 4월 14일까지 제출하려고 합니다. 제출한 문제집을 4월 14일까지 제출한 학생들에게는 100%의 장학금을 지급합니다. 제출한 학생들에게는 100%의 장학금을 지급합니다.	태국어 시험 문제집을 4월 14일까지 제출한 학생들에게는 100%의 장학금을 지급합니다.
14	2023-01-13	태국어	기타	2023 제1회 태국어 대회 우수 문제집	태국어	태국어 시험 문제집을 4월 14일까지 제출하려고 합니다. 제출한 문제집을 4월 14일까지 제출한 학생들에게는 100%의 장학금을 지급합니다. 제출한 학생들에게는 100%의 장학금을 지급합니다.	태국어 시험 문제집을 4월 14일까지 제출한 학생들에게는 100%의 장학금을 지급합니다.

★질문 올리기 전에 다른 사람이 올린 질문과 답변 확인★

질문 올리기

질문 올리기

질문 올리기

질문 올리기

질문 올리기

질문 올리기

질문 올리기

질문 올리기

질문 올리기

질문 올리기

질문 올리기

질문 올리기

3. 요약문 작성 원칙

가. 기본 작성 원칙

1) 기본 원칙

- ① 소주제별로 하나의 요약문을 작성하되 대화 요약, 화자별 발화요약, 결정 사항을 나누어 작성

－ 결론의 경우 대화를 통해 결정 사항이 나온 경우에만 작성한다.

예) 대화를 통해 약속 시간이나 장소를 정한 경우 등

구분	내용
대화 요약	
화자 1 요약	
화자 2 요약	
결정 사항	(대화를 통해 결정 사항이 나온 경우에만 작성)

- ② 모든 요약문은 명료하게 작성

- ③ 요약문은 대화 내에 명시적으로 제시된 대화의 내용을 위주로 작성

－ 요약문을 작성할 때 작업자의 주관적인 의견을 반영하지 않고, 객관적인 대화의 내용을 충실히 담아서 작성한다.

- ④ 전체 요약문은 소주제별 요약문을 잘 아우를 수 있도록 작성

나. 요약문 작성 예시

1) 2인 대화 - 논쟁적 대화

가) 소주제 : 주식과 관련한 청년들의 자산 축적 방법

(1) 발화 주석

***중심어는 파란색으로 표시**

화자ID	발화 주석	내용	화행 정보	비고
SD210234 7	중심 문장1	요즘 2 30대 젊은 층들이 그 가장 고민하고 있는 게 취업하고 그리고 주택 이 주택 문제가 심각하다고 하던 데 뭐 얼마 전까지만 해도 빛을 내서라도 뭐 주식을 구입해야 되는 주식 광풍 뭐 이런 게 부는데	정보 제공	

		그게 참 이 그 요즘 큰 문제 사회문제로 대두되는 거 같아요.		
SD210234 8	중심 문장2	주식 에 투자하는 거 자체는 문제가 아니라고 봐요	반응	
	뒷받침 문장	그리고 내가 20대 때만 해도 진짜 경제에 관 한 개념이 너무 없었는데 요새 20대들을 보면 정말 감탄스러울 정도로 자기의 재무목표를 건설하게 세 우고 그리고 투자를 어디에 해야 되는 지에 대한 고민을 굉장히 많이 해서 좋은 방향으로만 흘러 간다면 꽤 찮은 관심인 거 같아요.	정보 제공	중심 문장 2에 대한 뒷받침 문 장
SD210234 7	잉여적 문장	그렇지. 근데 지금 주식이나 집이나 요즘 젊은 층들 이 그쪽에 왜 몰입을 하는가 이거에 대해서도 한 번 생각을 해 볼 필요가 있는 게	정보 제공	
	중심 문장3	그거 말고는 어떤 재산 증식 자산을 축적할 수 있는 수단이 잘 없다고 생각을 하기 때문에 차라든지 아니면 집이라든지 이런 쪽으로 몰리는 게 아닌가 생각되거든?	정보 제공	

(2) 중심 문장 간 관계 주석

화자ID	발화 주석	내용	비고
SD210234 7	중심 문장1	요즘 2 30대 젊은 층들이 그 가장 고민하고 있는 게 취업하고 그리고 주택 이 주택 문제가 심각하다고 하던데 뭐 얼마 전까지만 해도 빚을 내서라도 뭐 주식을 구입해야 되는 주식 광풍 뭐 이런 게 부는데 그게 참 이 그 요즘	

		큰 문제 사회문제로 대두되는 거 같아요.	
SD210234 8	중심 문장2	주식에 투자하는 거 자체는 문제가 아니라고 봐요	반박
SD210234 7	중심 문장3	그거 말고는 어떤 재산 증식 자산을 축적할 수 있는 수단이 잘 없다고 생각을 하기 때문에 차라든지 아니면 집이라든지 이런 쪽으로 몰리는 게 아닌가 생각되거든?	무관

(3) 소주제별 요약문

구분	내용	
대화 요약	이 대화에서 SD2102347과 SD2102348은 젊은 층의 주택 문제와 주식 투자에 대한 의견을 공유했습니다. SD2102347은 젊은 층의 주식 투자 광풍 현상의 배경이 자산 축적 수단의 부재 때문이라는 의견을 내며 이를 사회 문제로 보았습니다. 반면 SD2102348은 재무 목표 수립과 투자에 대한 관심 등 주식 투자의 순기능에 대해 이야기하며 이는 문제가 아니라고 이야기했습니다.	
SD2102347 요약	요	SD2102347은 주택 문제와 주식 투자에 대한 젊은 층들의 관심이 사회적인 문제로 대두되고 있고 그 이유는 자산 축적 등을 위한 다른 수단이 제한적이기 때문이라고 이야기했습니다.
SD2102348 요약	요	SD2102348은 주식 투자는 문제가 아니며, 요즘 젊은 층들이 재무 목표를 세우고 투자에 관심을 갖는 것은 긍정적인 동향이라는 의견을 냈습니다.
결정 사항	(결정 사항 없음)	

나) 소주제 : 부동산 문제와 사회적 차원의 분산 필요성

(1) 발화 주석

화자ID	발화 주석	내용	화행 정보
SD210234 7	잉여 적 문장	일단 집값이나 이런 부동산은 그 기본적으로 구입을 하면 언젠가는 올라갈 것이다 근데 그 증가 속도가 최근에 엄청나게 어 그 증가폭이 확대가 되었고 급격하게 증가가 된 그런 부분들이 있고 그리고 그	정보 제공

		부동산 밖에 투자할 게 없다	
	중심 문장1	이런 인식이 강하다 보니까 결국은 돈이 대출을 받으면 거의 8 90퍼센트는 이 돈이 도는 게 아니라 부동산 에 투자를 하다 보니까 전체적으로는 사회적으로는 좋은 영향이 있는 거 같진 않아	정보 제공
SD210234 8	중심 문장2	그러면 결국은 주식투자를 해서 돈을 벌어서 우리나라는 부동산을 사는 게 답이다. 이렇게 사람들이 생각하고 있는 거네요?	정보 제공

(2) 중심 문장 간 관계 주석

화자ID	발화 주석	내용	비고
SD210234 7	중심 문장1	이런 인식이 강하다 보니까 결국은 돈이 대출을 받으면 거의 8 90퍼센트는 이 돈이 도는 게 아니라 부동산 에 투자를 하다 보니까 전체적으로는 사회적으로는 좋은 영향이 있는 거 같진 않아	무관
SD210234 8	중심 문장2	그러면 결국은 주식투자를 해서 돈을 벌어서 우리나라는 부동산을 사는 게 답이다. 이렇게 사람들이 생각하고 있는 거네요?	

(3) 소주제별 요약문

구분	내용
대화 요약	이 대화에서 화자 SD2102347과 SD2102348은 부동산과 주식에 대한 젊은 층들의 관심이 커지는 것에 대해 이야기했습니다. SD2102347은 급격한 부동산 가격 상승으로 인해 주택 투자에 대한 인식이 강해져 이가 사회에 좋지 않은 영향을 끼친다고 이야기하였습니다. SD2102347는 일부 사람들이 주식 투자를 통해 돈을 벌어 부동산을 구입하는 것이 해결책이라고 생각하는 것 같다는 의견을 공유합니다.
SD2102347 약	요 SD2102347은 부동산의 가격 상승 속도가 급격해지면서 주택 투자에 대한 인식이 강해지고, 이는 사회적으로는 좋은 영향을 미치지 않는다고 생각했습니다.
SD2102348 약	요 SD2102348은 주식 투자로 돈을 벌어서 부동산을 구입하는 것이 답이라고 생각하는 사람들이 있는 것 같다고 생각했습니다.
결정 사항	(결정 사항 없음)

다) 소주제 가, 나에 대한 전체 요약문

구분	내용
대화 요약	화자 SD2102347과 SD2102348은 젊은 층들이 자산 축적 방법, 특히 주식이나 주택에 큰 관심을 갖고 있다는 점에 대한 사회적인 문제나 이유 등을 이야기했습니다. 주식과 부동산 투자에 대한 젊은 층들의 관심이 커지는 것이 사회적으로는 좋은 영향을 미치지 않는 것 같다는 의견과 재산 증식을 위해 목표를 세우고 관심을 갖는 것은 좋은 현상이지만 일부 잘못된 인식을 갖고 있는 사람들이 있다는 의견을 공유했습니다.

2) 2인 대화 - 일상적 대화

가) 소주제 : 자주 듣는 노래와 좋아하는 장르

(1) 발화 주석

화자ID	발화 주석	내용	화행 정보	비고
SD2102418	중심 문장1	임영웅이 노래 잘하지	정보 제공	
SD2102418	뒷받침 문장	그리고 엄마 옛날에 왜 강릉 갈 땐 가 차에서 너네가 틀어 준 노래 할아버지 생각나서 막 울었잖아	정보 제공	중심 문장1에 대한 뒷받침문장
SD2102417	중심 문장2	아 그거는 영탁이가 부른 거지 영탁이가 부른 막걸리 한잔 이었나	정보 제공	
SD2102418	잉여적 문장	맞아 막걸리 한잔 그 노래 너무 슬펐어 할아버지 생각나고 막 울었잖아 엄마	정보 제공	
SD2102417	잉여적 문장	맞아 그 후로 엄마가 그 노래를 좀 많이 들었지	반응	

(2) 중심 문장 간 관계 주석

화자ID	발화 주석	내용	비고
SD210241 8	중심 문장1	임영웅이 노래 잘하지	
SD210241 7	중심 문장2	아 그거는 영탁이가 부른 거지 영탁이가 부른 막걸리 한잔 이었나	반박
SD210241 8	중심 문장3	나는 발라드도 좋아하고 힙합도 좋아하고 팝송도 좋아하고 난 다 좋아해	무관
SD210241 7	중심 문장4	그리고 엄마가 좋 엄마는 그 7080 노래 옛날 그 7080 통기타 치면서 하는 노래 들 그런 노래 좋아하지	무관

(3) 소주제별 요약문

구분	내용
대화 요약	이 대화에서 화자 SD2102418은 트로트를 들을 때 자신의 아버지가 생각나 울었다고 말했습니다. 7080과 같은 옛날 노래를 좋아한다고 이야기했습니다. 화자 SD2102418과 다르게 화자 SD2102417은 발라드, 힙합, 팝송을 좋아한다고 말했습니다.
SD2102418 요약	화자 SD2102418은 트로트를 들었을 때 돌아가신 아버지가 생각나 눈물을 흘린 경험을 이야기하며, 트로트 장르를 자주 들었다고 말했습니다. 또 통기타를 치며 부를 수 있는 7080과 같은 노래를 좋아한다고 이야기했습니다.
SD2102417 요약	화자 SD2102417은 발라드, 힙합, 팝송과 같은 모든 장르의 음악을 좋아한다고 답했습니다.
결정 사항	(결정 사항 없음)

나) 소주제 : 가족들의 상황별 애창곡

(1) 발화 주석

화자ID	발화 주석	내용	화행 정보	비고
SD2102418	중심 문장1	무주 갔을 때 그 한가위 행사 할 때 우리 다 같이 외삼촌 집에 가서 놀던 거 생각나지?	질문	
SD2102417	잉여적 문장	어 근데 그때 할머니 댁에서 조금 술을 먹다가 거기로 간 건가?	질문	
SD2102418	잉여적 문장	어	반응	
SD2102418	중심 문장2	그래서 엄마 그때 무슨 노래 불렀어?	질문	
SD2102417	중심 문장3	그때 엄마 당돌한 여자 불 렀잖아	반응	
SD2102418	중심 문장4	엄마가 제일 좋아하는 노래 지	정보 제공	
SD2102417	뒷받침 문장	그렇지 근데 엄마는 못 부 르면서 맨날 그 노래 부르 지	정보 제공	중심 문장3에 대한 뒷받침문장
SD2102418	잉여적 문장	그러지	반응	
SD2102417	잉여적 문장	왜 그때도 술 취해서 가사도 제대로 모르면서 이상하게 막 불렀었어 엄마	정보 제공	
SD2102418	잉여적 문장	맞아	반응	

(2) 중심 문장 간 관계 주석

화자ID	발화 주석	내용	비고
SD2102418	중심 문장1	무주 갔을 때 그 한가위 행사 할 때 우리 다 같이 외삼촌 집에 가서 놀던 거 생각나지?	
SD210241	중심	그래서 엄마 그때 무슨 노래 불렀어?	무관

8	문장2		
SD2102417	중심 문장3	그때 엄마 당돌한 여자 불렀잖아	지지
SD2102418	중심 문장4	엄마가 제일 좋아하는 노래 지	지지

(3) 소주제별 요약문

구분	내용
대화 요약	이 대화에서는 화자 SD2102418과 SD2102417이 가족 한가위 행사 때 부른 노래를 공유하고 가족 구성원들이 좋아하는 노래를 말했습니다. 화자 SD2102418은 7080 노래를 좋아하지만 음악 취향이 반대인 아빠는 트로트를 좋아한다고 했습니다. 화자 SD2102417은 신나는 음악을 좋아해 노래방에서 랩을 즐겨 부르지만 음악 취향이 반대인 언니는 발라드를 좋아한다고 말했습니다.
SD2102418 요약	화자 SD2102418은 가족 한가위 행사 때 당돌한 여자를 불렀으며, 다른 가족들이 부른 노래에 대해 질문했습니다. 화자 SD2102417의 노래 실력을 칭찬하며 노래방에서 불러준 노래가 좋아 운전하면서도 자주 들었다고 이야기했습니다. 아빠와는 노래 취향이 다르지만 트로트를 잘부른다고 말했습니다.
SD2102417 요약	화자 SD2102417은 가족 한가위 행사 때 언니와 함께 노래를 불렀고, 아빠는 부르지 않았다는 정보와 name7 오빠가 부른 노래 정보를 공유했습니다. 노래방에서는 랩을 부르고, 신나는 노래를 좋아하여 친구들에게 노래를 추천한다고 말했습니다. 화자 SD2102417은 신나는 노래인 랩을 좋아하지만, 언니는 조용한 발라드를 선호하여 음악 취향이 다르다고 했습니다. 또 엄마는 트로트와 7080 노래를 좋아한다고 말하며 name7이 좋아하는 가수의 정보를 공유했습니다.
결정 사항	(결정 사항 없음)

다) 소주제 가, 나에 대한 전체 요약문

구분	내용
대화 요약	이 대화에서 화자들은 좋아하는 음악과 장르에 대해 공유했습니다. 화자 SD2102418은 7080과 같은 옛날 노래를 좋아하고, 화자 SD2102417은 발라드, 힙합, 팝송을 좋아한다고 말했습니다. 또한 가족 한가위 행사 때 부른 노래와 가족 구성원들이 좋아하는 노래를 공유했습니다. 아빠는 트로트를 좋아하고, 언니는 발라드를 좋아한다고 말했습니다.

4. 작업 방법 세부 원칙

가. 기본 원칙

- ① 일상 대화 본문에서 같은 소주제에 대해 이야기하는 핵심 문장(=중심 문장)이 2개 이상 등장할 경우 하나의 주제로 보고 같은 소주제로 분류
- ② 소주제에 대한 판단은 대화 내 중심이 되는 화제로 판단되는 명사인 ‘단어’로 구분
- ③ 동일한 소주제에 대한 대화가 꼭 이어지지 않고 다른 대화 이후에 다시 나오더라도 하나의 소주제로 간주
 - * 여기서 하나의 소주제란, 중심 문장이 내용적으로 동일한 경우.
- ④ 소주제별 발화 개수를 너무 짧거나 길게 분류하지 않도록 주의
 - 소주제별 발화 개수를 평균적으로 100개 이상으로 잡되, 60~80개의 발화더라도 그 내용이 충실한 경우 하나의 소주제로 분류할 수 있다.
 - 소주제를 분류하기 애매하거나 내용이 빈약한 경우 발화 개수가 약 300개더라도 하나의 소주제로 분류할 수 있다. 단, 소주제별 발화 개수가 400개를 넘어가지 않도록 해야 하며 판단하기 어려운 경우 담당자에게 문의한다.

나. 대화의 종류에 따른 주제 분할

1) 논쟁적 대화

- ① 휴가, 대중교통, 스포츠 등 주제에 대한 토론, 논박 등
- ② 논쟁적 담화에서의 주제는 대화 참여자들의 주장에서 공통적으로 언급하고 있는 것으로 간주

2) 일상적 대화

- ① 일상적 대화의 주제는 발화자들이 이야기하는 하나의 이야기(세부 주제)
- ② 일상적 대화에서 발화자들은 서로 대답하며 다양한 소주제에 대해 이야기할 수 있음
 - 소주제는 대답, 긍정, 반박 등의 형식으로 나타나며 대답, 반응으로 생성되는 소주제는 세부 주제에 포함되는 것으로 본다.

다. 발화(문장) 주석 원칙

1) 중심 문장 주석

- ① 중심 문장은 하나의 소주제에서 1개 이상 등장
- ② 논쟁적 대화에서는 화자의 핵심 주장이나 의견을 중심 문장으로 간주
- ③ 일상적 대화에서는 단순 의견, 선호, 취향, 제안, 질문 등 여러 종류의 중심 문장이 등장
- ④ 각 대화에서 발화자의 의견을 가장 잘 나타내는 문장을 중심 문장으로 간주
- ⑤ 소주제를 던지는 발화나 소주제와 관련된 핵심적인 질문의 경우에도 중심 문장으로 간주

가) 중심어 주석 원칙

- 중심어는 소주제별로 1개 이상 작성
- 발화자의 발언에서 소주제를 가장 잘 설명할 수 있는 단어(핵심 어휘) 혹은 명사구(최대 3어절)
- 요약자의 임의적 판단에 의한 수정은 허용하지 않으며, 반드시 실제 발화(중심 문장)에서 그대로 추출하여 입력할 수 있는 단어 혹은 명사구만 허용

예) 취업용 자기소개서, 장기적 팀 프로젝트

- 적절한 중심어를 찾기 어려운 경우 담당자에게 문의

소주제	중심어 예시	잘못된 예
주식과 관련된 청년들의 자산 축적 방법	주식에 투자하는 거 자체는 문제가 아니라고 봐요	주식에 투자 (‘주식 투자’였다면 중심어로 잡을 수 있음)
	그거 말고는 어떤 재산 증식 자산을 축적할 수 있는 수단이 잘 없다고 생각을 하기 때문에 차라든지 아니면 집이라든지 이런 쪽으로 몰리는 게 아닌가 생각되거든요?	자산을 축적할 수 있는 수단(최대 3어절의 명사구여야 함)
부동산 문제와 사회적 차원의 분산 필요성	이런 인식이 강하다 보니까 결국은 돈이 대출을 받으면 거의 8 90퍼센트는 이 돈이 도는 게 아니라 부동산에 투자를 하다 보니까 전체적으로는 사회적으로는 좋은 영향이 있는 거 같진 않아	부동산에 투자 (‘부동산 투자’였다면 중심어로 잡을 수 있음)
	그러면 결국은 주식투자를 해서 돈을 벌어서	부동산 (같은 주제의 중심 문장)

	우리나라는 부동산을 사는 게 답이다. 이렇게 사람들이 생각하고 있는 거 네요?	내에 같은 중심어가 있는 경우 중심어를 작성하지 않음)
트로트 경연 TV프로그램에 나온 참가자와 좋아하는 곡	임영웅이 노래 잘하지 그리고 엄마 옛날에 왜 강릉 갈 땐 가 차에서 너네가 틀어 준 노래 할아버지 생각나서 막 엄마 울었잖 아	너네가 틀어 준 노래 (주제를 포괄할 수 있는 최소한의 단어 또는 명사 구만 선택하여야 함)
	아 그거는 영탁이가 부른 거지 영탁이가 부른 막걸리 한잔 이었나	막걸리 한잔 (주제를 포괄할 수 있는 최소한의 단어 또는 명사 구만 선택하여야 함)
	아 어디지 그 이찬원이 개가 부르는 노래도 좋던데 난 찬원이 부르는 노래 미스트롯 에서 나온 거	개가 부르는 노래 (주제를 포괄할 수 있는 최소한의 단어 또는 명사 구만 선택하여야 함)
	그거는 기억까지 안 나 나는 근데 동원이 정동원 정동원이 부른 노래가 한 개 있거 든 눈물비 눈물비 알아?	정동원이 부른 노래 (주제를 포괄할 수 있는 최소한의 단어 또는 명사 구만 선택하여야 함)
엄마가 좋아하는 7080 노래	그래? 그리고 엄마 운전할 때 듣는 노래 엄마는 뭐 우연히 그리고 7080 노래 들이 좋아 통기타 치고 옛날 노래 다 까먹었네 엄마도	엄마 운전할 때 듣는 노 래 (주제를 포괄할 수 있는 최소한의 단어 또는 명사 구만 선택하여야 함/최대 3어절의 명사구여야 함)
	그리고 엄마가 좋 엄마는 그 7080 노래 옛날 그 7080 통기타 치면서 하는 노래들 그런 노래 좋아하지	7080 통기타 치면서 하는 노래 (최대 3어절의 명사구여 야 함)

2) 뒷받침 문장 주석

- ① 중심 문장을 뒷받침/부연설명하는 형태로 중심 문장 앞 또는 뒤에 위치
- ② 뒷받침 문장 주석 시, 반드시 어떤 중심 문장에 대한 뒷받침 문장인지 선택
- ③ 뒷받침 문장과 중심 문장은 반드시 화자가 일치하도록 주석
(예시 : 발화자A의 중심 문장-발화자A의 뒷받침 문장은 성립, 발화자A
의 중심 문장-발화자B의 뒷받침 문장은 성립하지 않음)
- ④ 뒷받침 문장은 없을 수도 있음

3) 잉여적 문장 주석

- ① 잉여적 문장은 중심 문장과 뒷받침 문장을 제외한 모든 문장
- ② 추가 정보 전달, 의견 제시, 공감, 동의, 반응 등의 형태

4) 화행 정보 주석

- ① 문장 주석 시 반드시 1개의 화행 정보(질문, 요구(요청), 정보 제공, 반응, 감정 표현, 기타)를 주석

가) 질문 : 예/아니오 질문, 의문사 포함 질문 등

* 질문의 경우 해당 문장만 읽고 질문의 내용을 이해할 수 있는 문장에만 ‘질문’ 으로 주석하고, 아래 예시의 ‘너는?’ 과 같이 질문의 내용을 알 수 없는 경우에는 아래/위 문장에 포함시켜 주석한다.

발화 내용(예시)	해당 여부
너는 요즘 뭐 듣고 있어?	O
그러면 너는 슬플 때는 어떤 음악을 주로 들어?	O
너는?	X

나) 요구(요청) : 요청, 명령, 제안 등

다) 정보 제공 : 주장, 정정, 추측, 소망, 의지, 설명 등등

라) 반응 : 질문이나 의견에 대한 직접적인 답변, 동의 표현, 비동의 표현, 수락, 거부, 회피 등

* 질문/의견에 대한 발화가 직접적이고 짧은 답변과 부연 설명(근거)으로 나뉘는 경우 직접적 답변은 ‘반응’ 으로, 부연 설명(근거)은 ‘정보 제공’ 으로 주석한다.

화자ID	발화 주석	발화 내용(예시)	화행 정보
SD210117 4	중심 문장	혹시 배워 보고 싶은 악기는 있어?	질문
SD210117 5	뒷받침 문장	내가 중학교 때 수행평가로 바이올린을 잠깐 했었는데 바이올린을 칠 때 나는 소리가 너무 내 스타일이고	정보 제공
	중심 문장	그래서 바이올린을 배우고 싶어.	반응

화자ID	발화 주석	발화 내용(예시)	화행 정보
SD210230 5	잉여적 문장	딸이나 아들이 뭐 어디 주로 어디에서 근무하고 있어? 어떤 부류의 일을 하고 있어?	질문
SD210230 4	잉여적 문장	엄마 주변에는 삼성 전자도 있고 또 저기	정보 제공
		뭐 두산 두산 중공업?	
		뭐 이런 데도 있고 뭐 이런데 두산 중공업 같은 경우에는	
		거기 아들이 너무 힘들어 해 가지고 어 아아 (후략)	

마) 감정 표현 : 인사, 칭찬, 비판, 감사, 사과, 비난, 감탄사 등

바) 기타 : 위의 화행 구분에 들어가지 않는 그 외 모든 표현

* 문장이 화행에 따라 분할되지 않은 경우, 발화 주석 단위를 고려하여 하나의 화행으로 통일한다.(발화의 주된 내용이 무엇인지 판단하여 결정)

화자ID	발화 주석	내용	화행 정보	비고
SD210241 8	중심 문장1	무주 갔을 때 그 한가위 행사 할 때 우리 다 같이 외삼촌 집에 가서 놀던 거 생각나지?	질문	
SD210241 7	잉여적 문장	어 근데 그때 할머니 댁에서 조금 술을 먹다가 거기로 간 건가?	질문	

참고
SD2102417의 발화는 반응(‘어’)과 질문(‘근데 그때 할머니 댁에서 조금 술을 먹다가 거기로 간 건가?’)이 함께 나타나고 있음. 이때 ‘어 근데 그때’와 같이 화행 정보와 문장 구분의 단위가 일치하지 않는 경우 하나의 발화 주석 단위(잉여적 문장)를 고려하여 하나의 화행으로 통일할 수 있음. 위 사례는 질문이 주된 화행이므로, ‘질문’으로 통일함.

라. 중심 문장 간 관계 주석

1) 문장 관계 주석 원칙

- ① 중심 문장 간의 관계를 주석
- ② 문장의 등장 순서대로 앞뒤 문장에 대해서만 주석
예) 3~4인 대화에서 중심 문장이 A-B-C의 순서일 때는 A-B의 관계, B-C의 관계만 주석한다.
* A-B의 관계가 지지인데, B-C의 관계가 반박이라면 자연스럽게 A-C의 관계가 반박임을 자연스럽게 추론할 수 있기 때문.
- ③ 중심 문장이 A-A-B 순서이면 문장 순서에 맞게 A-A, A-B의 관계만 주석
- ④ 문장 간 관계는 지지, 반박, 무관 중 택1

2) 문장 관계 주석 요소

- ① 지지 관계 : 중심 문장 간 관계가 긍정적일 때 주석
(지지, 선호, 인정, 공감 등)
* 앞선 화자의 의견을 요약하거나, 환언하여 진술할 경우는 지지 관계
* 화자의 의견 표명으로 볼 수 없는 단순 재진술, 이해 확인 질문 등의 형태일 때는 지지 관계로 주석하지 않도록 주의
* 앞선 화자의 의견에 다른 견해나, 근거, 예시를 단순 추가했다면 지지 관계로 주석하지 않도록 주의
- ② 반박 관계 : 중심 문장 간 관계가 부정적이며 상충할 때 주석
(반박, 거절 등)
- ③ 무관 관계 : 중심 문장 간 관계가 성립하지 않는 경우일 때 주석
* 특히 질문과 답변의 문장 관계의 경우 기본적으로는 ‘무관’으로 주석하되 명백하게 반박, 거절의 의사를 표현한 경우에는 ‘반박’으로 주석

마. 요약문 작성

1) 요약문 작성 기본 원칙

- ① 요약문은 화자, 소주제 등을 반드시 포함하여 작성
- ② 요약문은 발화 주석 후 소주제별 요약문을 작성하고, 소주제별 요약문이 작성된 후 최종적으로 전체 요약문을 작성
- ③ 원문의 구성과 표현을 활용하여 작성하되 원문을 그대로 차용하는 것은 지양(대화 내 등장하는 단어들의 동의어, 유의어 등의 단어를 활용)
- ④ 하나의 요약문 내에서 동일한 단어나 표현을 사용하지 않도록 주의

- ⑤ 화자 정보(주어)의 경우 플랫폼 내 표시되는 화자 ID를 그대로 복사하여 사용

예) SD2102426, SD2100002, SD2100003

- ⑥ 요약문은 합쇼체로 작성하되 시제는 아래와 같이 작성

- 기본 : 과거형(~했습니다.)
- 계획 : 미래형(~것입니다.)
- 결론 : 과거형(~했습니다.)

- ⑦ 결정 사항의 경우 대화를 통해 결정된 사항이 있는 경우에만 작성

예) 대화를 통해 약속 시간이나 장소를 정했거나, 진로 고민 후 결정하는 경우 등

- ⑧ 요약문 작성 시 자의적 판단이 개입되지 않도록 주의하고, 대화 내 명시적으로 제시된 내용을 요약.

참고) ‘의견이 서로 대립하고 있다’ , ‘필요성이 논의된다’ 와 같은 표현은 대화문에 대한 작업자의 판단이 들어간 표현이므로 부적절

2) 소주제별 요약문

- * 요약문은 아래 표의 파란색 부분과 같이 화자, 소주제 등을 반드시 포함하여 작성하여야 하며, 표현은 자유롭게 변형할 수 있다.

구분	내용
대화 요약	이 대화에서 (화자ID)과 (화자ID)은/는 _____에 대해 _____했습니다.
SD2100003 요약	(화자ID)은/는 _____라고 _____했습니다.
SD2100002 요약	(화자ID)은/는 _____라고 _____했습니다.
결정 사항	이 대화에서는 _____에 대해 _____는 _____을 _____했습니다.

- ① 중심 문장과 중심어를 중심으로 요약하되, 필요에 따라 뒷받침 문장을 활용하여 작성

- ② 중심 문장을 활용해 하나의 소주제에 대한 발화자들의 의견을 정리하는 방식으로 작성

- ③ 항목별 요약 내용

- 대화 요약 : 대화에서 드러나는 소주제와 그 소주제에 대한 화자별 의견,

결정 사항을 명료하게 작성

- 화자별 요약 : 소주제에 대한 화자별 의견을 근거 등을 포함하여 구체적으로 작성
- 결정 사항 : 대화를 통해 나온 결정 사항이 있는 경우 작성

3) 전체 요약문

구분	내용
대화 요약	

- ① 대화문의 소주제가 1개 이상인 경우 소주제별 요약문의 ‘대화 요약’ 내용을 통합한 후 문장 간 이어지는 흐름이 자연스럽게 느껴지도록 수정하여 작성
- ② 대화문의 소주제가 1개인 경우 소주제별 요약문의 ‘대화 요약’ 내용이 전체 요약문을 대체

5. 기타 사항

가. 작업 시 주의사항

- 요약문 작성 시 **ChatGPT 사용을 금지**한다.

나. 검수 및 수정 시 주의사항

1) 작업자 주의사항

- ① 검수 완료된 파일의 노트(note) 내용을 꼼꼼하게 확인한 후 수정
- ② 피드백 받은 내용을 다음 작업에 적용
- ③ 노트(note) 내용은 참고하되 삭제하지 않고 제출

2) 검수자 주의사항

- ① 검수 노트(note)는 작업자가 피드백 내용을 잘 이해하여 다음 작업에 반영할 수 있도록 작성
- ② 수정 완료된 파일이 본인이 작성한 노트(note) 내용에 맞게 수정되었는지 확인하고 노트(note)를 삭제한 후 최종 완료 처리

국회 회의록 요약 가이드

1. 작업 원칙

1.1 기본 원칙

- 요약문은 분할된 주제를 기준으로 작성한다.
- 요약문은 ‘논쟁(논의)’를 중심으로 작성한다.
- 요약문만 읽고 대화의 주제와 내용을 확인할 수 있어야 한다.
 - 어떤 주제가 다루어졌는지, 어떤 발화자의 어떤 의견이 있었는지, 해당 주제에 대해 어떻게 결정이 났는지 등을 확인할 수 있어야 한다.

1.2 기본 내용

- 중요한 정보
 - ① 대화의 주제는 무엇인가?(주제 분할)
 - ② 대화를 구성하는 문장들은 어떠한 의도로 쓰였는가?(문장 주석)
 - ② 각각의 주제에 대해 어떤 이야기가 오갔는가?(주제별 요약)
 - ③ 전체 대화의 내용은 무엇인가?(최종 요약)

2. 작업 개요

2.1 구성

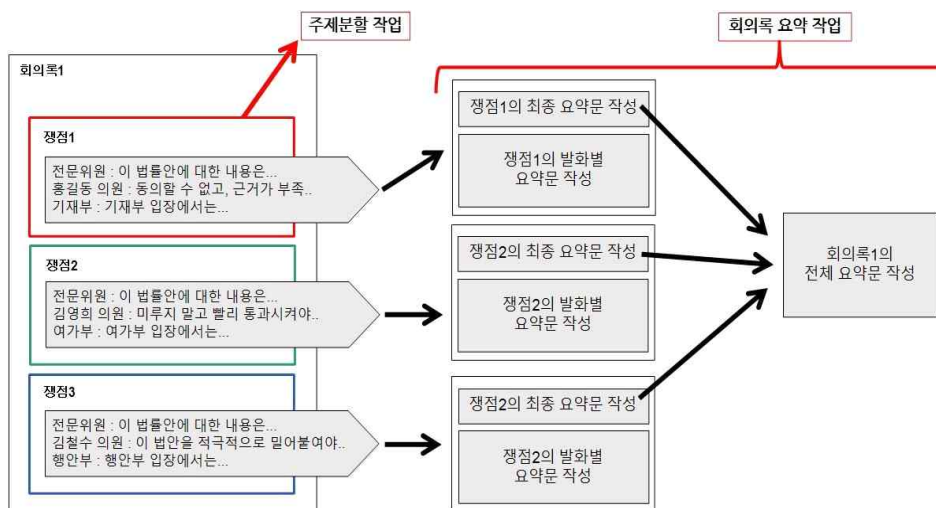


그림 1 작업 개요

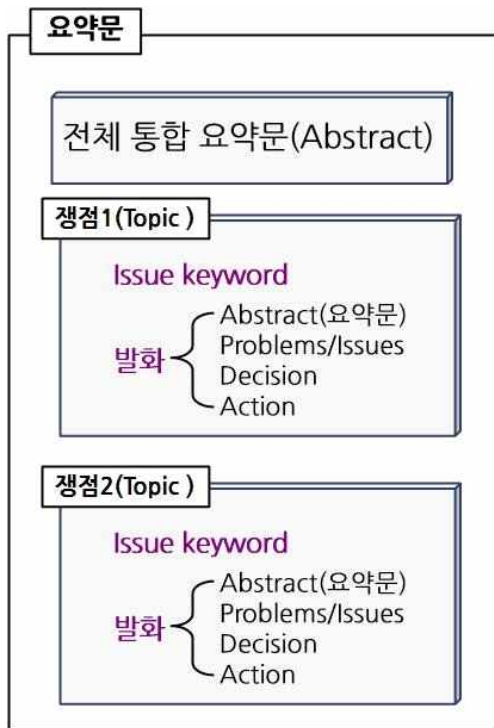


그림 2 완성된 요약문의 세부 구조

• 구분

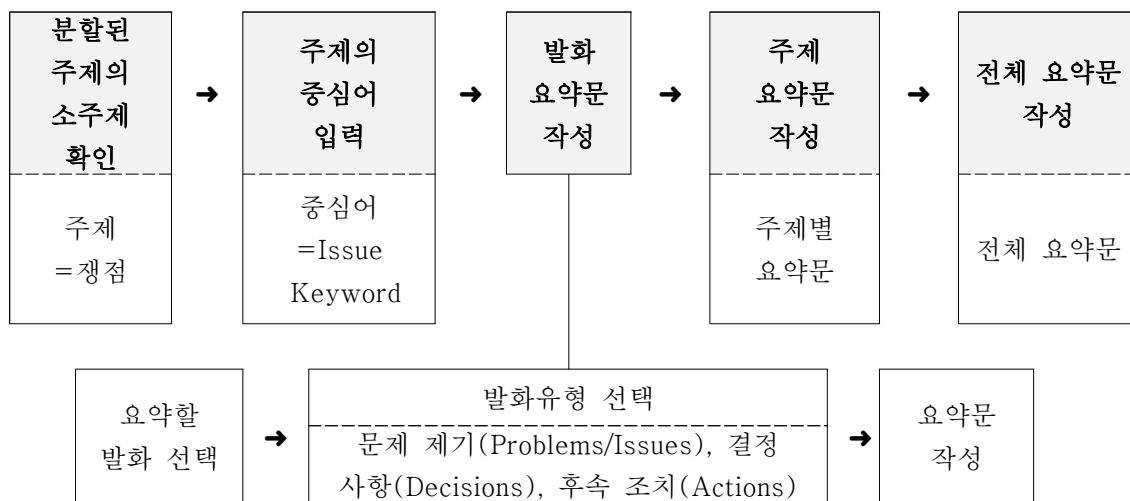
- ① 전체 통합 요약문 : 하나의 회의록 전체에 대한 요약문
- ② 주제(Topic) : 해당 쟁점과 연관이 있는 심사된 안건
- ③ 중심어(Issue keyword) : 중심어. 해당 쟁점을 포괄적으로 나타낼 수 있는 단어 혹은 명사구
- ④ 발화 : 해당 쟁점에 대해 논의한 내용

• 발화 종류

- ① 주제별 요약문(Abstract) : 논쟁적 대화 전체를 포괄하는 요약문
*주제별로 반드시 1개의 주제별 요약문(Abstract)를 작성
- ② 문제 제기(Problems/Issues) : 논쟁적 대화에서 논의의 대상이 되는 쟁점/문제/이슈
*발화의 대부분을 차지함
- ③ 결정 사항(Decisions) : 논쟁적 대화에서 도출된 결정 사항
- ④ 후속 조치(Actions) : 논쟁적 대화를 통해 결정된 후속 조치
*회의록에 따라 결정 사항이나 후속 조치가 없는 경우에는 결정 사항(Decisions), 후속 조치(Actions)는 작성하지 않음

2.2 작업 순서

- ① 분할된 주제의 소주제를 확인한다.
- ② 해당 주제를 포괄적으로 나타낼 수 있는 중심어(단어 혹은 명사구)를 ‘복사-붙여넣기’하여 입력한다.(중심어(Issue keyword))
- ③ 해당 주제 내에서 발화에 대한 요약문을 작성한다.
(요약할 발화 선택 → 발화유형 선택 → 요약문 작성)
(문제 제기(Problems/Issues), 결정 사항(Decisions), 후속 조치(Actions))
- ④ 해당 주제에 대한 요약문을 작성한다.(주제별 요약문)
- ⑤ 주제별 요약문을 참고하여 회의록 전체를 아우르는 요약문을 작성한다.



***작업 시 구글 크롬의 맞춤법 검사기능을 통해 표시되는 빨간 밑줄을 확인하며 오타와 문법적 오류에 주의하여 작성(자세한 방법은 플랫폼 작업 순서 참고)**

2.3 플랫폼 작업 순서

- 구글 크롬의 맞춤법 검사 기능 활성화(필수)

① 구글 크롬의
오른쪽 화면 상단 버튼 클릭

② '설정' 클릭

③ 왼쪽 화면 하단의
'언어'탭 클릭

④ 맞춤법 검사 버튼 활성화

회의록 요약

작성 현황

작성

작성 지침

Q&A

teddy13님

로그아웃

키워드 입력하기

발화ID

키워드

begin

end

세부 요약문

Type

요약문

발화ID

note

이동

삭제

이제일 위원은 학생중식사업비와 관련하여 학생중식 지원문제는 결식아동이나 가정형편이 어려운 학생들을 어떻게든 지원하는 것이 중요하니까 그런 관점에서 이것을 받아들였으면 좋겠다고 생각합니다.

- ⑤ 작성한 요약문 아래 빨간 밑줄이 그어진 부분을 확인하고 맞춤법, 띄어쓰기를 고려하여 수정

* 이름, 부처명, 직함, 대명사, 법안명 등은 띄어쓰기나 맞춤법상 오류가 있더라도 원문을 따라 수정없이 그대로 작성하며 특히 자주 발생하는 오류에 주의한다.

– 자주 발생하는 오류 : 타이핑 오류, 이름 오타 등

X	O
~할때, ~할 때 ~때문에	~할 때, ~때문에
~는것, 는걸로	~는 것, ~는 걸로
~할수 있는	~할 수 있는

• 회의록 요약 작업 방법

회의록 요약 작업현황 및 작업기한 확인

회의록 목록

5월

6월

7월

8월

9월

10월

11월

과일명	작업종류	작업현황	작업기한	작업버튼
216예결소위(예산안조정)04차	주제 분할	미작업	2022-06-20	작업하기
216예결소위(예산안조정)07차	주제 분할	미작업	2022-06-20	작업하기
215교육소위(법안)01차	주제 분할	작업 중	2022-07-20	작업하기
215교육소위(법안)01차	회의록 요약	미작업	2022-08-20	작업하기
215교육소위(예산)01차	회의록 요약	작업 중	2022-08-20	작업하기

✓ 작업탭에서 작업할 회의록의 작업하기 버튼 클릭

① 키워드 입력

② 키워드를 추출한 발화ID 입력

③ 확인버튼 클릭

④ 요약할 발화 선택 (shift 버튼 사용 가능)

⑤ 화살표 버튼 클릭

Title: 2001회계연도교육부 소관예산안
Topics: 1. 2001회계연도교육부 소관예산안

선택	발화ID	식별	발화사	발화
<input type="checkbox"/>	7	교육부기획관리실장	이기우	교육부 기획관리실장입니다.
<input type="checkbox"/>	8	교육부기획관리실장	이기우	여제 여러 위원님들께서 시정발 증액 및 세정에 대한 실의를 주셨습니까?
<input type="checkbox"/>	9	교육부기획관리실장	이기우	여기에 대해서 기획관리실, 교육정책기획관, 국제교육협력관장, 학교정책실장으로 해당 실국장님이 보고를 드리도록 하겠습니다.
<input type="checkbox"/>	10	교육부기획관리실장	이기우	그러나 보고드리기 전에 말씀드릴 것은 위원님들께서 걱정하는 내용에 대한 설명을 주셨는데 저희 교육예산은 턱없이 부족합니다. 그리고 지금까지 작업에 대해서는 교육위원회 분과위원회에서 해결위에 요구한 것이 없었기 때문에 저희들이 보고드리는 것은 정책하도록 하겠습니다.
<input type="checkbox"/>	11	교육부기획관리실장	이기우	그 점을 명쾌하게 주시면 좋겠습니다.
<input type="checkbox"/>	12	교육부기획관리실장	이기우	그러면 기획관리실부터 말씀드릴 것입니다. 기획관리실은 기획예산담당관이 보고를 드리도록 하겠습니다.
<input type="checkbox"/>	13	위원	이재정	장관님요, 예산을 확보하기 위한 정부쪽의 노력과 연구에 대해서는 제가 두말할 필요가 없습니다. (다)이는 여제 위원회에서 일단 시정요구를 한 것 자체에 대해서도 이 위원회의 일단 보고가 되어서 기록으로 남는 것이 상당히 중요

키워드 입력하기

발화ID: [발화ID 입력] [확인]

키워드: [키워드 입력]

begin

end

세부요약문

Type	요약문	발화ID	note	이동	삭제
------	-----	------	------	----	----

※ 작업 중간중간 저장 필수

[이전] [작업저장] [다음]

- ✓ 해당 쟁점의 중심어(Issue keyword)를 발화 내에서 복사-붙여넣기 하여 입력
- ✓ 중심어가 포함되어 있는 발화ID(발화번호)를 입력 - 확인
- ✓ 쟁점에 대한 논의/문제제기/요청사항 등 요약할 발화 선택 (복수 선택 가능)
- ✓ 화살표 버튼 클릭

④ 발화 타입 선택
-PI
-action
-decision
-abstract

⑦ 내용 요약

⑧ 요약 작업 반복 후 저장
⑨ 다음 버튼 클릭
(다음 토픽 요약 작업)

※abstract는 반드시 작성
abstract를 작성하지 않으면 넘어가지
않음. 반드시 쟁점별 abstract를
작성할 것!!

- ✓ 발화 타입 선택(Abtract, Problems/Issues, Decision, Action)
- ✓ 선택한 발화에 대한 요약문 작성(반드시 하나의 문장이어야 함)
- ✓ 요약문 작성 반복
- ✓ 주제(쟁점)별 Abstract는 반드시 1개를 작성(복수 문장으로 작성 가능)
- ✓ 다음 주제(쟁점) 요약문 작성(작업저장 후 다음 버튼 클릭)

⑩ 토픽별 요약문 확인

⑪ 최종 요약문 작성

⑫ 작업저장
⑬ 제출

※작업 중간중간 저장 필수

- ✓ 주제(쟁점)별 요약문(Abtract) 확인
- ✓ 주제(쟁점)별 요약문(Abtract) 을 바탕으로 최종 요약문 작성
(복수 문장으로 작성 가능)
- ✓ 작업저장 - 제출

3. 요약문 작성 원칙

- 각 요약문은 최대한 아래의 정보를 담아야 한다.
 - ① 누가
 - ② 무엇을
 - ③ 어떻게
 - ④ 왜
- 모든 요약문은 반드시 1개의 문장으로 작성한다.
- 예외적으로 전체 요약문은 복수의 문장으로 작성할 수 있다.
- 모든 요약문은 요약 시 참고한 모든 발화를 선택하여 작성한다.
- 모든 요약문은 문장별로 각각 최대한 독립성을 갖도록 작성한다.
 - 다른 문장과 상관없이 그 문장만 읽었을 때도 내용을 이해할 수 있어야 한다.
 - ‘이 법안’ ‘해당 제안’과 같은 모호한 표현은 지양하며, 내용이 길어지더라도 구체적인 표현으로 작성한다.

✓ 교육부지방교육지원국장 정병걸은 김상희 위원의 지적에 대해 보건복지부가 사회복지법인의 재산을 분할하는 것에 대해 부정적인 입장을 취하고 있음을 밝히며, 법인 재산 분할 관련 문제는 현재까지도 해결되지 않았다고 답변했다.

→ A가 제시한 의견(질의)에 대한 답변, 동의, 반박 등의 내용이라면 “A”를 명시한다.

✗ 교육부지방교육지원국장 정병걸은 보건복지부가 사회복지법인의 재산을 분할하는 것에 대해 부정적인 입장을 취하고 있음을 밝히며, 법인 재산 분할 관련 문제는 현재까지도 해결되지 않았다고 답변했다.

→ 누군가의 질의에 대한 답변임이 분명하지만, 누가 누구인가 요약문에 표현되지 않음 질의자가 누구인지

✓ 전문위원은 안민석 위원이 발의한 평생교육법 일부개정법률안에 대하여 현행법과 유사한 형태들이 나열되고 있으나 현행법과 달리 세부적인 내용이 없고 현행법 체계와 맞지 않는 측면이 있어 수정 및 보완을 제안했다.

→ 전문위원이 ‘무엇’에 대해 수정 및 보완을 제안하고 있는지 명시한다.

✗ 전문위원은 현행법과 유사한 형태들이 나열되고 있으나 현행법과 달리 세부적인 내용이 없고 현행법 체계와 맞지 않는 측면이 있어 수정 및 보완을 제안했다.

→ 바로 위 요약문에 안민석 위원안이 있어서 연이어 읽으면 이해할 수 있으나, 이 문장만 읽었을 때는 ‘무엇’에 대한 내용인지 확인할 수 없음

- 요약문의 순서는 발화유형과 상관없이 발화ID 순서에 따라 자연스럽게 작성한다.
- 주어 작성 기준 : 누가(주체)

① 정부측 의견의 경우

- 정부 명칭은 되도록 해당 부처명을 사용한다.
(국토교통부, 행정안전부 등)
- 실무자의 의견과 정부 의견을 구분하여 명시한다.

✓ 정부부처의 입장일 경우

→ '통일보'는 이러한 입장을 밝혔다'와 같이 부처명 사용

✓ 정부부처의 실무자 입장일 경우

→ '통일보 차관 황보기'는 홍길동 위원의 질문에 대해 이러한 입장을 밝혔다'와 같이 부처명+실무자 이름으로 작성

- 보통 절차에 따라 정부의 의견을 발표하는 경우에는 정부 부처의 입장, 그 외에 회의 내에서 질의응답을 진행하는 경우에는 실무자 개인의 의견으로 보고 주어를 결정한다.
예) '정부측 말씀해주시기 바랍니다'와 같은 사회자의 발화를 통해 실무자 개인이 아닌 정부 부처의 의견임을 알 수도 있다.
- 행정부 전체 결정 사항이거나 부처 구분이 크게 의미가 없을 경우 '정부' 명칭을 사용할 수 있다.

② 호칭

- 직책과 이름을 모두 표기하며 반드시 아래의 순서를 따라 표기한다.

전문위원 A / 소위원장 B / 부처명 + 직급 C / D 위원

✓ 전문위원 홍길동은 정부가 발의한 교원의 지위 향상 및...

✓ 교육부 차관 이영은 시간강사 처우개선 방안 추진에 있어...

✓ 소위원장 조승래는 다음 소위원회까지 개정안을...

✓ 김민기 위원은 이 안건을 오늘 처리하지 않는 것이 당의 입장이라며...

✗ 위원 이미경은 주택정책에 대한 주요 쟁점 중에서...

✗ 박백범 교육부차관은 평생교육법 일부개정법률안 전문위원 검토의견을...

- 공청회 회의록에서 진술인의 표기는 '진술인 000'으로 표기하며 일관성을 위해 공청회를 제외한 모든 회의록은 위의 표기법을 따른다.

5. 작업 방법 세부 원칙

5.1 중심어(Issue keyword)

- 해당 쟁점의 중심 화제를 대표할 수 있는 단어 혹은 명사구
- 쟁점당 1개만 입력한다.
- 반드시 실제 발화에서 요약자의 임의적 판단에 의한 수정을 하지 않고 그대로 (복사-붙여넣기) 가져와서 입력할 수 있는 단어 혹은 명사구만 허용한다.
- 대부분 사회자(소위원장) 또는 전문위원의 발언에서 찾을 수 있다.
- 적절한 키워드를 찾기 어려운 경우 담당자에게 문의한다.

□ ○소위원장 김성태
그래서 오늘 안건 4번, 자원의 절약과 재활용촉진에 관한 법률 일부개정법률안을 순서를 일부 변경해서 의사일정 제4항을 먼저 상정하여 심사하도록 하겠습니다.

□ ○소위원장 조승래
좋습니다. 그러면 이것도 의결을 잠시 보류하고요. 의사일정 제3항을 심사하도록 하겠습니다. 안민석 의원이 대표발의한 평생교육법 일부개정법률안을 심사하겠습니다. 전문위원, 보고해 주시기 바랍니다.

□ ○전문위원 박용수
4항 보고드리겠습니다. 사이버 따돌림 예방과 피해자 지원 관련 법률의 보완·강화에 관한 청원입니다.

5.2 문제 제기(Problems/Issues)

- 어떤 내용이 P/I인가?
 - ① 안건(논의 대상)
 - ② 안건과 관련된 쟁점/문제(지적)/제안/요청/주장 등
 - ③ 해당 회의를 거쳐야만 하는 결정, 변화, 문제제기 등(회의 목적 감안)
- 포함할 요소
 - ① 누가 문제를 제기하였는가?
 - ② 어떤 문제가 이야기되었는가?
 - ③ 근거가 드러난 경우, 해당 근거를 문장에 포함한다.
- 유의사항
 - ① 각각의 문제/이슈는 반드시 한 문장으로 작성해야 한다.
 - ② 과거형으로 작성한다.
예) 홍길동 위원은 ~ 라고 주장하였다.

- ③ 한 명의 화자가 다수의 이슈를 언급했다면 이슈별로 분리하여 별도의 P/I로 새로운 줄에 입력한다.
- ④ 여러 사람이 유사한 의견을 낸 경우에는 통합하여 작성한다.
예) 김철수, 김영희, 홍길동 위원은 정부가 발의한 ○○○개정안에 동의하였다.
- ⑤ 너무 많은 내용을 무리하게 한 문장에 작성하지 않고, 읽고 이해하는데 어려움이 없도록 내용에 따라 나누어 작성한다.

• 유형별 - 발표(주로 전문위원)

- ① 독립된 쟁점인 경우
 - 쟁점별로 작성한다.
 - 문장이 너무 길어질 경우 세부 주제에 맞게 분할하여 작성한다.
(법안심사의 경우 ‘조(항)’을 기준으로 작성)
 - 발표는 결정 사항(Decisions), 후속 조치(Actions)가 될 수 없다.
- ② 독립된 쟁점이 아닌 경우
 - 분량/이슈 내용을 고려하여 최소한으로 작성한다.

• 유형별 - 논쟁

- ① 안전에 대한 의견(쟁점/문제(지적)/제안/요청/주장 등)
- ② 주요 안전과 완전히 별개인 의사진행발언의 경우 제외할 수 있다.

□ 소위원장 조승래는 다음 소위원회까지 개정안을 발의한 의원, 전문위원 그리고 정부의 의견을 조율하여 대안 완성 시 의결할 것을 제안했다.

□ 교육부는 개인정보 보호법의 처벌 수위와 형평성 유지를 위해 형량과 벌금 상향 조정에 동의했다.

□ 김민기 의원은 이 안건을 오늘 처리하지 않는 것이 당의 입장이라고 밝히며, 2011년 개정된 법안에 대해 2019년 9월에 개정안이 새로 발의되었다면 한 달 만에 빨리 처리할 것이 아니라, 충분한 논의를 거칠 필요가 있다고 주장했다.

□ 교육부차관 이영은 시간강사 처우개선 방안 추진에 있어, 시간 강사 처우 부분을 평가 지표로 넣는 등 대학들의 변화를 유도하는 한편, 평균 임금 상승을 위한 재정사업도 만들고 협의체를 구성해서 운영할 것이라고 정부의 의견을 밝혔다.

□ 전문위원은 정부가 발의한 교원의 지위 향상 및 교육활동 보호를 위한 특별법 일부 개정법률안과 관련하여 현행법 8조 4항을 삭제하면 공무원인 위원에 관해서는 위원 신분을 보장하기 어려우므로 개정안에 ‘공무원이 아닌 위원’ 중 ‘공무원이 아닌’을 빼고 ‘위원은’으로 수정할 것을 제안했다.

□ 이미경 의원은 주택정책에 대한 주요 쟁점 중에서 임대주택법 일부개정법률안, 임대차등록제 관련 내용만 유독 이유 없이 제외되었다며 소위원장의 안전 수용 태도에 문제가 있다고 지적했다.

5.3 결정 사항(Decisions)

- 어떤 내용이 결정 사항(Decisions)인가?

– 해당 안건에 대한 논의 후 결정된 사항

***회의 전에 이미 결정된 사항은 결정 사항(Decisions)으로 볼 수 없다.**

- 포함할 요소

① 누가 결정하였는가?

– 다수 화자(회의 전체)에 의한 결정일 경우 생략 가능

② 어떤 결정이 이루어졌는가?

③ 근거가 드러난 경우 해당 근거를 문장에 포함한다.

④ 제안이나 요구사항이 전제된 안건이라면 그 주체와 내용도 포함한다.

- 유의사항

① 결정 사항은 **반드시 한 문장으로 작성**해야 하며, 결정 사항이 복수인 경우 새 요약문을 추가하여 건별로 나누어 작성한다.

② **과거형**으로 작성한다.

결정 주체 결정 근거 결정 내용 요구사항/전제사항

□ 지방자치단체 재난 안전 사업 투자 현황 수집·분석 근거 마련의 건은 보류했다가 다음 심사에서 심사하기로 하였다.

→ 결정 사항이 다수의 결정(의결)일 경우 결정 주체 생략 가능

□ 정우택 위원이 발의한 주택도시보증법안과 윤후덕 위원이 발의한 국민주택보증법안의 논의는 마쳤으나 야당 소위 위원들이 참석하지 않아 의결은 보류되었다.

□ 새마을금고법 일부개정법률안은 소위에서 계속 심사하기로 결정되었다.

□ 기동민, 박정, 이우현, 이명수, 정병국, 김두관, 강길부 각 의원이 발의한 재난 및 안전 관리 기본법 일부 개정 법률안 7건은 소위에서 계속 심사하기로 하였다.

□ 주승용, 이종명, 장정숙, 윤재옥, 김정우 각 의원과 정부가 발의한 재난 및 안전 관리 기본법 일부 개정 법률안 6건은 본회의에 부의하지 않고 현재까지의 심사 결과를 반영하여 위원회 대안을 제안하기로 의결하였다.

5.4 후속 조치(Actions)

- 어떤 내용이 후속 조치(Actions)인가?

- 회의 참여자 혹은 관계자가 수행해야 할 ‘과업(후속 조치)’
- 향후 조치 사항이 결정된 사항
예) 추가 제출, 재논의, 재검토 등

*회의 전에 결정되었거나 이미 수행되고 있는 사항, 회의와 관계없이 수행할 사항은 후속 조치(Actions)에 포함하지 않는다.

- 포함할 요소

- 누구에 의해서 과업이 지시되었는가?
- 어떤 과업인가?
- 누가 수행해야 하는 과업인가?
- 근거가 드러난 경우, 해당 근거를 문장에 포함한다.

- 유의사항

- 후속 조치/수행 과업 등은 **반드시 한 문장으로 작성**해야 하며, 후속 조치/수행 과업 등이 복수인 경우 새 요약문을 추가하여 건별로 나누어 작성한다.
- 시제는 **미래형**으로 작성한다.
- 수행 주체를 반드시 포함하여야 하며, 원문 내용에 따라 다르게 표현될 수 있으므로 원문 표기를 그대로 따른다.

과업(후속조치)
지시/요청 주체

과업(후속조치)
지시/요청/수행 근거

과업(후속조치) 내용

수행 주체

☐ 경찰청 소관 법률안인 총포·도검·화약류 등의 안전 관리에 관한 법률 일부 개정 법률안에 대한 심사는 경찰청이 준비되지 않아 차후에 진행될 것이다.

☐ 국토교통부는 공적 보증에 관한 중복지원 방지를 위해 가능한 사항을 시행령에 구체화할 것이다.

☐ 수석전문위원 박수철은 지역 누락 문제를 보완할 것이다.

☐ 재건축초과이익 환수에 관한 법률 폐지법률안은 의결정족수가 되지 않아 다음 회의에서 논의가 이어질 것이다.

5.5 주제별 요약문(Abstract)

- 주제별 요약문(Abstract)는 무엇인가?
 - ① 논쟁적 대화를 포괄하는 요약문이다.
 - ② 모든 주제는 반드시 1개의 주제별 요약문(Abstract)를 가진다.
 - ③ 모든 주제별 요약문(Abstract)은 내용을 읽으면, 해당 주제별 요약문(Abstract)이 포함하는 범위에 대한 대략적인 내용을 이해할 수 있도록 작성되어야 한다.
- 포함할 요소
 - ① 주요 논의
 - ② 상호 간의 주장
 - ③ 회의를 통해 결정된 사항(Decisions)
 - 결정 사항이 없는 경우 작성하지 않는다.
 - ④ 회의 이후에 진행해야 할 후속 조치(Actions)
 - 후속 조치가 없는 경우 작성하지 않는다.
- 유의사항
 - ① P/I 등을 먼저 작성한 후 최종적으로 작성한다.
 - ② 주제별로 반드시 1개의 주제별 요약문(Abstract)만 작성한다.
 - ③ P/I 등과 달리, **복수의 문장으로 작성**할 수 있다.
- 주제별 요약문(Abstract)
 - ① 해당 안건의 주요 논의 내용과 결과를 함께 작성한다.
 - ② 필요시 안건과 관련한 논의의 종류와 과정, 대립 관계 등을 중심으로 작성한다.
 - ③ 해당 주체에서 무슨 안건이 있었고 어떤 결과가 내려졌는지를 알 수 있도록 작성한다.
 - ④ 구체적인 주요 논쟁 주제와 참여자들의 입장, 회의 분위기 등을 파악할 수 있도록 작성한다.

□ 정우택 의원이 발의한 주택도시보증법안과 윤후덕 의원이 발의한 국민주택기금법안에 대해 전문위원은 두 법안을 일원화할 것을 제안하고 추가 수정의견을 제시했으며, 국토교통부는 이를 모두 수용했다. 위원들은 중복 지원 등 일부 우려되는 사안을 제외하면 대체로 법안 내용에 찬성했으나, 야당 위원들이 출석하지 않아 의결은 보류되었다.

□ 보건복지부는 감염 관련 조치에 의한 손실보상을 규정하고 있는 수정의견에 동의하며, 구체적인 사항을 대통령령으로 정해야 할 필요가 있음을 주장하였다. 손실보상심의위원회의 위원장은 보건복지부차관과 민간위원이 공동으로 말하는 것으로 결정하였으며, 보건복지부와 위원들은 유급휴가 부여에 대한 수정의견과 제70조의3을 수용하였다. 보건복지부는 제64조제8호의 '의사' 및 '의료업자'라는 표현 수정에 대한 의견을 제출할 것이다.

□ 교육부가 국회의 승인 없이 연말 부정수급 방지 보조 인력 단기 일자리 사업을 추진한 사안과 관련하여 시정요구사항 중 '조직적으로'라는 표현은 교육부의 요청을 받아들여 삭제하고 그 외의 문구는 유지하며 시정요구유형 또한 주의로 유지하기로 하였다.

- 전체 요약문

- 회의록 전체 내용을 대상으로 하며, 주제별 요약문을 참고하여 작성한다.

예) 본 회의는 어떠한 회의로 어떤 안건이 상정되었으며 그에 대해 이러한 결론이 내려졌다.

□ 본 회의는 국회(정기회) 제3차 법안심사소위원회로 지난 회의 때 심사를 마치지 못한 새마을금고법 일부개정법률안과 19건의 개인정보 보호법 일부개정법률안을 심사했으며, 모두 합의점을 찾지 못해 계속 심사하기로 결정되었다.

□ 본 회의는 제326회 국회(임시회) 국토교통위원회 제1차 국토법안심사소 위원회로 정우택 의원이 발의한 주택도시기금법안과 윤후덕 의원이 발의한 국민주택기금법안, 신동우 의원이 발의한 재건축초과이익 환수에 관한 법률 폐지법률안 등을 심사했다. 회의를 통해 주택도시기금법안과 국민주택기금법안의 일원화가 결정되었으며 재건축초과이익 환수에 관한 법률 폐지법률안은 의결정족수 미만으로 의결이 보류되었다.

* 주제별 & 전체 요약문 참고 예시

□ **쟁점별 abstract**

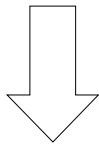
주택임대차보호법 일부개정법률안에 찬 반 의견이 나뉘었고 공청회를 하자는 주장이 나왔으며 '주택임대차보호법 법무부 종합의견' 문건 중 1항, 2항, 3항, 6항 관련해서 위원회안으로 가결되었고 시행일 관련해서 1항, 3항, 6항은 2014년 1월 1일부터 나머지 부분은 공포일로부터 시행하는 것으로 수정 의결되었다.

상가건물 임대차보호법의 계약갱신청구권의 제한과 범위에 대한 논의가 이루어졌고 상가건물 임대차보호법 일부개정법률안은 '법무부 종합의견' 문건 1항, 3항, 4항, 6항, 7항을 위원회안으로 제시하여 가결하기로 하였으며 채택이 되지 않은 부분에 대해서 계속 심사하기로 하였다.

법원조직법 일부개정법률을 회의에서 논의해서 수정한대로 의결하기로 하였다.

채권의 공정한 추심에 관한 법률 일부개정법률안은 논의가 이루어지지 않았다.

상설특별검사의 임명 등에 관한 법률안은 다음 소위에서 논의하기로 하였다.



□ **최종 abstract**

본 회의는 제316회 국회 제4차 법안심사제1소위원회회의로 의사일정 제21항 법원조직법 일부개정법률안이 수정 의결되었고 주택임대차보호법 일부개정법률안과 상가건물 임대차보호법 일부개정법률안은 소위에서 논의한 안건을 위원회 안으로 제시해서 가결하기로 결정되었다. 채권의 공정한 추심에 관한 법률 일부개정법률안은 논의가 이루어지지 않았으며 상설특별검사의 임명 등에 관한 법률안은 다음 소위에서 논의하기로 결정되었다.

6. 기타 사항

6.1 법안

- 회의록에서는 법안 심사가 목적이므로 가장 명확하고 핵심적인 정보인 법령의 ‘조(항)’을 중심으로 문장을 분리한다.
- 법률안 처리과정
 - ① 해당 상임위의 법안심사소위원회 논의(가장 활발한 논의)
 - ② 소위원회 의결
 - ③ 상임위 전체회의 의결(소위원회 거치면 대부분 통과)
 - ④ 법제사법위원회 의결
 - ⑤ 본 회의 상정

6.2 예/결산 심사

- 소위원회에서 담당 기관이 예산을 어떻게 사용했는지를 확인 후, 부적절하다고 판단한 사항에 대해 ‘시정 요구’를 하여 후속 조치를 하도록 하기 때문에 ‘시정 요구’를 중심으로 문장을 분리한다.
- 의원들의 시정 요구에 대해 정부측이 어떻게 받아들이기로 결정했는지(수용/일부 수용/수용 불가)도 포함하여 작성한다.
- 시정 요구 요소
 - ① 시정 요구 사항(내용)
 - ② 시정 요구 유형(변상/징계/시정/주의/제도개선)

6.3 검수 후 주의사항

- 작업자 주의사항
 - ① 검수 완료된 파일의 노트(note) 내용을 꼼꼼하게 확인한 후 수정한다.
 - ② 피드백 받은 내용을 다음 작업에 적용한다.
 - ③ 노트(note) 내용은 참고하되 삭제하지 않고 제출 버튼을 누른다.
- 검수자 주의사항
 - ① 검수 노트(note)는 작업자가 피드백 내용을 잘 이해하여 다음 작업에 반영할 수 있도록 작성한다.
 - ② 수정 완료된 파일이 본인이 작성한 노트(note) 내용에 맞게 수정되었는지 확인하고 노트(note)를 삭제한 후 최종 완료 버튼을 누른다.

<기획·연구>

국립국어원 강미영 언어정보과장

국립국어원 유희정 학예연구사

국립국어원 서혜진 연구원

<연구 참여자>

연구 책임자 함영균(테디 쌤)

공동 연구원 김한샘(연세대학교)

연구 보조원 정용빈, 이이슬, 오유진(테디 쌤)

박재완, 이지인(테디 쌤)

김유진, 강조은(연세대학교)

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9727

인쇄일: 2023년 11월 30일

발행일: 2023년 11월 30일

인 쇄: 올프린팅

※ 이 보고서는 국립국어원의 용역비로 수행한 ‘2023 대화 요약 말뭉치 연구 분석’ 사업의 결과물을 발간한 것입니다.

