

국립국어원 2019-01-52

발 간 등 록 번 호
11-1371028-000769-01

개인적 글쓰기 자료 수집 및 말뭉치 지식 강연회 개최

사업 책임자
나 윤 정

제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구용역 계약에 따라 ‘개인적 글쓰기 자료 수집 및 말뭉치 지식 강연회 개최’에 관한 연구 보고서를 작성하여 제출합니다.

■ 사업기간: 2019년 11월 ~ 2020년 6월

2020년 6월 30일

사업 책임자: 나 윤 정((주)메트릭스코퍼레이션)

사업 수행자 (주)메트릭스코퍼레이션, (주)레인보우커뮤니케이션

사업 책임자 나윤정

사업 참여자 박두진, 민지현, 김홍건, 김기현, 조일상,
조신애, 남궁건, 고유미, 박준현, 김규일,
이경화, 윤지은, 하지영, 채윤철, 유혜영,
임상호, 김수연, 김종서, 김은영, 이태연,
한이슬, 박지유, 권혁윤, 김다정, 김소현,
곽성호, 유영석, 이진철, 나진수, 이다연,
이종진, 이화복, 정선호, 황성하, 김소윤,
최세희

<사업 수행자> (주)메트릭스코퍼레이션 · 주식회사
 레인보우커뮤니케이션

사업 책임자	나운정((주)메트릭스코퍼레이션)
	박두진((주)메트릭스코퍼레이션)
	민지현((주)메트릭스코퍼레이션)
	김홍건((주)메트릭스코퍼레이션)
	김기현((주)메트릭스코퍼레이션)
	조일상((주)메트릭스코퍼레이션)
	조신애((주)메트릭스코퍼레이션)
	남궁견((주)메트릭스코퍼레이션)
	고유미((주)메트릭스코퍼레이션)
	박준현((주)메트릭스코퍼레이션)
	김규일((주)메트릭스코퍼레이션)
	이경화((주)메트릭스코퍼레이션)
	윤지은((주)메트릭스코퍼레이션)
	하지영((주)메트릭스코퍼레이션)
	채운철((주)메트릭스코퍼레이션)
	유혜영((주)메트릭스코퍼레이션)
	임상호((주)메트릭스코퍼레이션)
	김수연((주)메트릭스코퍼레이션)
	김종서((주)메트릭스코퍼레이션)
	김은영((주)메트릭스코퍼레이션)
	이태연((주)메트릭스코퍼레이션)
	한이슬((주)메트릭스코퍼레이션)
	박지유((주)메트릭스코퍼레이션)
	권혁윤((주)메트릭스코퍼레이션)
	김다정((주)메트릭스코퍼레이션)
	김소현((주)메트릭스코퍼레이션)
	곽성호((주)메트릭스코퍼레이션)
	유영석((주)레인보우커뮤니케이션)
	이진철((주)레인보우커뮤니케이션)
	나진수((주)레인보우커뮤니케이션)
	이다연((주)레인보우커뮤니케이션)
	이종진((주)레인보우커뮤니케이션)
	이화복((주)레인보우커뮤니케이션)
	정선호((주)레인보우커뮤니케이션)
	황성하((주)레인보우커뮤니케이션)
	김소윤((주)레인보우커뮤니케이션)
	최세희((주)레인보우커뮤니케이션)

개인적 글쓰기 자료 수집 및 말뭉치 지식 강연회 개최

본 사업은 인공 지능 산업 발전을 위한 대규모 고품질 우리말 자원에 대한 공공 인식 확산을 위해 추진되었다. ‘개인적 글쓰기 자료 수집’ 과업에서는 국민 참여의 장을 마련하여 글쓰기 자료를 수집하고 전문적인 교정을 거치지 않은 비출판물 말뭉치를 구축하였다. ‘말뭉치 지식 강연회 개최’ 과업에서는 일반인을 대상으로 언어 기반 인공 지능에 대한 강연회를 개최하였다. 이러한 대중 참여형 말뭉치 구축과 활용 사업으로 국어 정보화에 대한 국민의 관심을 높이고 우리말 자원을 활용한 연구 및 산업 환경 조성에 기여하고자 하였다.

개인적 글쓰기 자료 수집 과업은 ‘다양한 세대의 글쓰기 자료 수집’, ‘전문적인 교정 과정을 거치지 않은 개인적인 글쓰기 자료 확보’에 중점을 두어 추진하였다. 사업의 진행은 글쓰기 주제 구성, 수집 방법 등 자료 수집 전반에 걸친 세부 계획 수립, 추진 위원회 구성 및 운영, 홍보 진행, 글쓰기 자료 수집 실시 및 저작권 확보, 수집 저작물 목록 및 정보 정리 단계를 순차적으로 진행하였다.

글쓰기 자료를 단기간에 많이 모으기 위해 ‘한 편 줍쇼’라는 행사명을 정하고 행사용 누리집을 통해 자료를 주로 수집하였다. 전자 계약 방식으로 저작권 이용 허락 계약을 체결하고 참여자의 글을 받았다. 글의 종류는 참여자가 시(동시), 일기, 편지글, 소설(동화), 감상문, 기타로 분류하게 하였고, 블로그 게시물나 출판 등을 포함하여 어디에도 공개되지 않은 글로 참여를 제한하였다. ‘한 편 줍쇼’ 홍보는 온라인 커뮤니티와 도서관 및 주민 센터 게시판, 버스 정보 시스템 등을 활용하여 온라인과 오프라인에서 진행하였다. 누리집 이용이 어려운 참여자를 고려하여 오프라인 제출도 허용하였다.

초등학생부터 60세 이상 성인까지 총 6,044명이 참가하여 전체 10,919편, 9,250,097자의 글을 수집하였다. 참여자들은 최대 10편까지 글을 제출할 수 있으며, 적어도 총 1,000자 이상의 글을 제출하여야 했다. 만 7세~만 12세 800명, 만 13세~만 15세 497명, 만 16세~만 18세 1,058명, 만 19세~만 29세 2,720명, 만 30세~만 59세 648명, 만 60세 이상 321명이 참가하여 다양한 세대의 글쓰기 자료 수집이란 목적을 달성할 수 있었다. 수집한 글은 국립국어원 말뭉치 형식(JSON)에 부합하도록 정제하여 비출판물 원시 말뭉치를 구축하였다.

말뭉치 지식 강연회 과업에서는 ‘인공 지능 시대 언어 자원의 가치와 자연어 처리 현황’을 주제로 강연회를 개최하여 국어 자원 활용의 미래 전망을 제시하고자 하였다. 2020년 2월 오프라인 형식으로 개최할 계획이었으나 코로나바이러스감염증-19로 인해 2020년 6월 18일(목) 온라인으로 개최하였다. 행사의 성공적인 개최를 위해 영상 송출, 카드 뉴스 배포 등 다양한 온·오프라인 홍보를 진행했다.

온라인으로 진행한 말뭉치 지식 강연회에는 총 652명이 사전 참여를 신청하고 524명이 실제로 접속하였다. 강연자는 민규동 감독·주성철 평론가, 남세동 대표, 이경일 대표, 송길영 부사장으로 총 5명의 강연자가 4개의 강연을 진행했다. 강연당 평균 120~130명 내외 참여자가 참석했다. 다채로운 행사를 위해 강연회 휴식 시간에 전시 영상을 송출하고 문제 풀이 경품 행사를 실시했으며 강연회 종료 후 후기 작성 행사를 진행하여 참여자 반응을 확인했다. 구성이 알차고 어려운 내용을 쉽게 접할 수 있어서 좋았다는 반응이 다수였다.

강연회가 종료된 후 더 나은 말뭉치 지식 강연회를 위해 만족도 설문 조사를 진행했으며, 설문 조사 응답자(253명) 중 절반이 넘는 인원(191명)이 강연회에 대해 높은 만족도를 보였다. 설문 조사 응답자의 약 44%의 인원이 ‘민규동 감독, 주성철 평론가’와 ‘송길영 부사장’의 강연을 가장 기대했다고 답변하였고, 업무나 학업에 가장 도움이 된 강연으로는 48%가 ‘이경일 대표’의 강연을 꼽았다. 이는 말뭉치에 대한 쉬운 강의보다는 전문적인 강의에 대한 수요가 높음을 알 수 있다. 따라서 앞으로 말뭉치 관련 지식 강연회 개최 시 전문성을 강화할 필요가 있다.

주요어: 개인적 글쓰기 자료, 비출판물 말뭉치, 말뭉치 지식 강연회, 인공 지능

차례

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 개인적 글쓰기 자료 수집	7
2.1.1 글쓰기 자료 수집 개요	7
2.1.2 글쓰기 주제, 수집 방법 선정	8
2.1.2.1 자문 위원회 구성	8
2.1.2.2 자문 위원회 회의 개최	9
2.1.2.3 자문 위원 의견 취합 및 확정	10
2.1.3 글쓰기 자료 수집 공모 제목 선정	12
2.1.4 ‘한 편 줍쇼’ 누리집 관련 디자인	14
2.1.4.1 공모 제목 디자인	14
2.1.4.2 누리집 초기 화면	15
2.1.4.3 누리집 저작권 이용 허락 계약 시스템 구축	16
2.1.4.4 누리집 개인 인증 시스템 구축	17
2.1.4.5 누리집 메뉴 설정	18
2.1.5 ‘한 편 줍쇼’ 홍보 진행	19

차례

2.1.5.1 오프라인 홍보: 광고지, 소책자 제작 및 배포	19
2.1.5.2 온라인 홍보: 누리 소통망 행사, 온라인 커뮤니티	20
2.1.6 글쓰기 자료 수집 현황	23
2.1.6.1 ‘한 편 줍쇼’ 누리집 글 수집 결과	23
2.1.6.2 미성년자 법정 대리인 동의서 및 증명서 제출	23
2.1.7 표절 검사 진행	25
2.1.7.1 카피킬러 프로그램 이용	25
2.1.8 원시 말뭉치 정제	27
2.1.8.1 원문 자료 입력(초등학생 및 오프라인 제출자 글 대상)	27
2.1.8.2 원문 입력 규칙	28
2.1.8.3 자유 글감 ‘글감’ 분석	29
2.1.8.4 XML 문서 구조와 형식	30
2.1.8.5 말뭉치 파일명 부여 형식	31
2.2 말뭉치 지식 강연회 개최	32
2.2.1 말뭉치 지식 강연회 행사 개요	32
2.2.2 사전 홍보	36
2.2.2.1 오프라인 홍보	36
2.2.2.2 온라인 홍보(카드 뉴스)	38
2.2.2.3 온라인 홍보(유관 기관 대상)	41
2.2.2.4 온라인 홍보(커뮤니티 배너 광고)	45
2.2.2.5 언론 홍보	45

차례

2.2.3 행사 운영	46
2.2.3.1 온라인 강연회 사전 접수 신청	46
2.2.3.2 사전 준비	46
2.2.3.3 행사 운영	51
2.2.4 사후 관리: 후기 작성 행사, 강연 영상 제작, 설문 조사	56
2.2.4.1 후기 작성 행사	56
2.2.4.2 말뚝치 지식 강연회 만족도·개선 사항 설문 조사	60
2.2.4.3 강연 영상 제작	64

제3장 사업 수행 결과

3.1 개인적 글쓰기 자료 수집 결과	67
3.1.1 연령 및 직업별 수집 현황	67
3.2 말뚝치 지식 강연회 결과	69
3.2.1 말뚝치 지식 강연회 종합 평가	69
3.2.1.1 운영 결과(송출 부문)	69
3.2.1.2 운영 결과(강연 부문)	70
3.2.1.3 운영 결과(문제 풀이 행사 부문)	70
3.2.1.4 실시간 접속 결과	70

차례

부록

1. 개인적 글쓰기 자료 수집 저작권 이용 허락계약서75
2. 말뭉치 지식 강연회 접속자82
3. 말뭉치 지식 강연회 후기 작성 행사85
4. 말뭉치 지식 강연회 제작물88

표 차례

<표 1> 사업의 범위와 과업 내용	2
<표 2> 글쓰기 자료 수집 목표	7
<표 3> 글쓰기 자료 수집 결과	7
<표 4> 자문 위원 구성	8
<표 5> 자문 회의 논의 내용(2019. 12. 9. 개최)	9
<표 6> 글쓰기 자료 수집 글의 구체적인 내용	11
<표 7> 글쓰기 자료 수집 글의 공모 후보안 선정 과정	12
<표 8> 누리집 개인 인증 시스템	17
<표 9> 연령별 ‘한 편 줍쇼’ 참여자 구성 결과	23
<표 10> 표절 검사 결과	26
<표 11> 손글씨 제출자 연령별 구성	27
<표 12> 메타 정보 구조	30
<표 13> XML 정제 사항	31
<표 14> 강연회 진행 일정 개요 시간	33
<표 15> 강연회 진행 세부 일정	34
<표 16> 비예산 홍보_공공 기관 보유 매체 송출	36
<표 17> 온라인 홍보_관련 업체 연락 목록	41
<표 18> 온라인 홍보_대학 연락 목록	42
<표 19> 강연회 행사 당일 업무 구분	50
<표 20> 강연회 만족도 및 개선 사항 설문 조사	62

표 차례

〈표 21〉 강연회 만족도 및 개선 사항 설문 조사 대상 기본 정보	63
〈표 22〉 강연회 결과 동영상 목록	64
〈표 23〉 연령별 ‘한 편 줍쇼’ 참여자 구성	67
〈표 24〉 직업별 ‘한 편 줍쇼’ 참여자 구성	67
〈표 25〉 연령별 글 수집 결과	68
〈표 26〉 온라인 강연회 평균 접속자 결과	71
〈표 27〉 말뭉치 지식 강연회 참석자 명단	82
〈표 28〉 강연회 후기 작성 행사 참여자	85

그림 차례

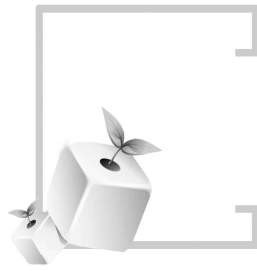
〈그림 1〉 사업의 배경	1
〈그림 2〉 개인적 글쓰기 자료 수집 사업 수행 절차	3
〈그림 3〉 자문 의견서	10
〈그림 4〉 ‘한 편 줍쇼’ 제목 도안	14
〈그림 5〉 ‘한 편 줍쇼’ 누리집 도안	15
〈그림 6-1〉 저작권 이용 허락 계약서(오프라인용)	16
〈그림 6-2〉 저작권 이용 허락 계약서(온라인용)	16
〈그림 7〉 ‘한 편 줍쇼’ 누리집 메뉴 설정	18
〈그림 8〉 ‘한 편 줍쇼’ 광고지 및 소책자	19
〈그림 9〉 누리 소통망 이용 홍보 행사	20
〈그림 10〉 ‘한 편 줍쇼’ 1, 2차 온라인 홍보	21
〈그림 11〉 (주)메트릭스 패널 누리집	22
〈그림 12〉 (주)메트릭스 패널 대상 발송 ‘한 편 줍쇼’ 홍보 전자 우편	22
〈그림 13〉 법정 대리인 동의서	24
〈그림 14〉 표절 검사 결과 확인서 예시	25
〈그림 15〉 원문 입력 예시: 수원동신초등학교 신재혁	28
〈그림 16〉 전문 요원 글감 입력 대상 글 예시	29
〈그림 17〉 XML 기본 구조	30
〈그림 18〉 비출판물 말뭉치 XML 예시	31
〈그림 19〉 말뭉치 지식 강연회 광고지	32
〈그림 20〉 말뭉치 지식 강연회 행사 장소	33

그림 차례

〈그림 21〉 버스 정보 시스템 홍보 모니터링 캡처(평택시청 소통홍보관실 제공)	37
〈그림 22〉 1차 카드 뉴스	38
〈그림 23〉 2차 카드 뉴스	39
〈그림 24〉 3차 카드 뉴스	40
〈그림 25〉 온라인 홍보_유관 기관 연락 결과	44
〈그림 26〉 스펙업 배너 광고, 페이스북 게시물, 카카오톡 메시지 발송	45
〈그림 27〉 1차, 2차 보도자료	45
〈그림 28〉 누리집 사전 접수 화면(왼쪽), 개별 입장 코드 발송(전자 우편, 문자)	46
〈그림 29〉 기념품 구성(왼쪽), 경품 구성(오른쪽)	47
〈그림 30〉 문화체육관광부 제1차관 오영우 축하	47
〈그림 31〉 국립국어원장 소강춘 개회사	47
〈그림 32〉 ‘성시경’ 팀의 인공지능 이야기 창작 보조 프로그램(웹 드라마 ‘블라인드’)	48
〈그림 33〉 ‘AI人’ 인공지능 기반 한국어 시 생성 프로그램	48
〈그림 34〉 행사장 동선 등	49
〈그림 35〉 현장 운영	50
〈그림 36〉 민규동 감독, 주성철 평론가	51
〈그림 37〉 남세동 대표	52
〈그림 38〉 이경일 대표	53
〈그림 39〉 송길영 부사장	54
〈그림 40〉 문제 풀이 행사 내용	55

그림 차례

〈그림 41〉 말뚝치 지식 강연회 후기 작성 행사	56
〈그림 42〉 말뚝치 지식 강연회 송출 화면	69
〈그림 43〉 저작권 이용 허락계약서	75
〈그림 44〉 말뚝치 지식 강연회 후기 1	85
〈그림 45〉 말뚝치 지식 강연회 후기 2	86
〈그림 46〉 말뚝치 지식 강연회 후기 3	87
〈그림 47〉 말뚝치 지식 강연회 제작물 1	88
〈그림 48〉 말뚝치 지식 강연회 제작물 2	89
〈그림 49〉 말뚝치 지식 강연회 제작물 3	90
〈그림 50〉 말뚝치 지식 강연회 제작물 4	91



제 1 장

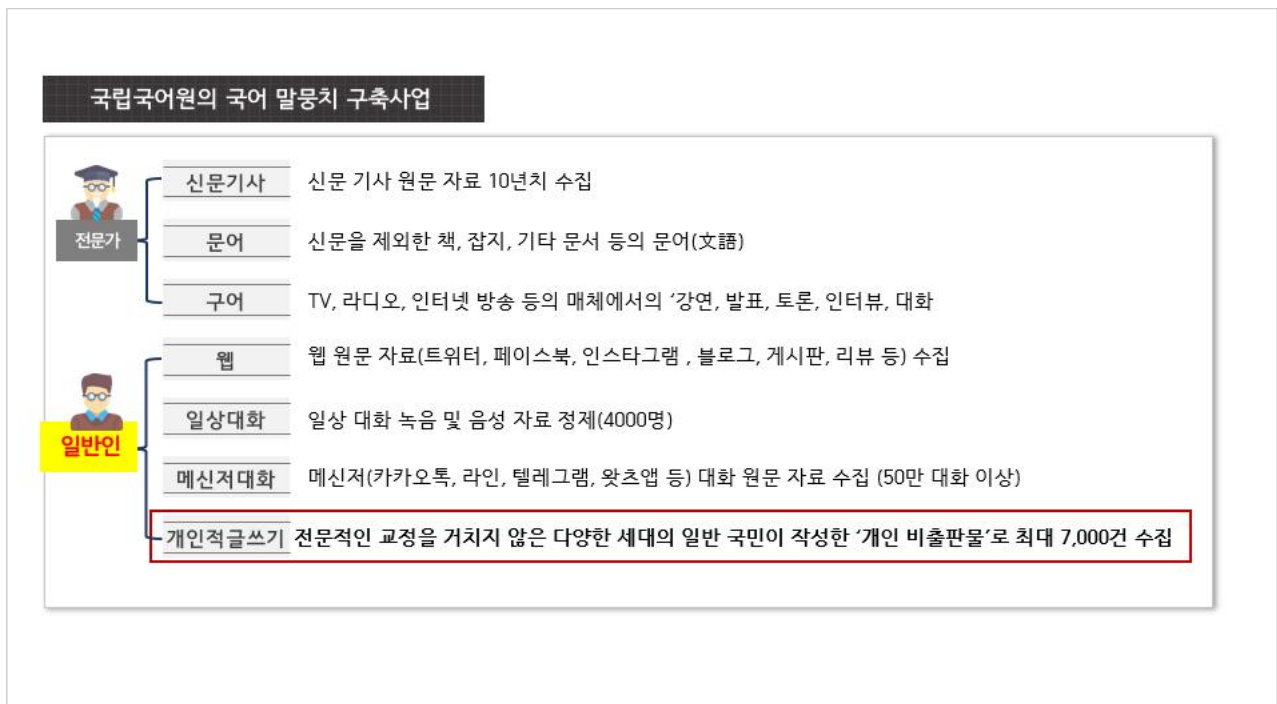
사업 개요



1.1 사업 목적

본 사업은 인공 지능 산업 발전을 위한 대규모 고품질 우리말 자원 수요 증대에 따라 말뭉치 자료를 다양화하고 말뭉치를 일반에 널리 알리기 위해 추진되었다. 첫 번째 사업 목표는 국민 참여의 장을 마련하여 글쓰기 자료를 수집하고, 비출판물 말뭉치를 구축하는 것이다. 두 번째 목표는 말뭉치 지식 강연회를 개최하여 국어 정보화에 대한 국민 관심을 유도하고 국어 자원 활용 연구 및 산업 환경을 조성하는 것이다.

개인적 글쓰기 자료 수집은 ‘다양한 세대의 글쓰기 자료 수집’, ‘전문적인 교정 과정을 거치지 않은 개인적인 글쓰기 자료 확보’에 중점을 두어 추진하였다. 말뭉치 지식 강연회에서는 ‘인공 지능 시대 언어 자원의 가치와 자연어 처리 현황’을 주제로 강연회를 개최하여 국어 자원 활용의 미래 전망을 제시하고자 하였다.



〈그림 1〉 사업의 배경

1.2 사업 수행 범위

본 사업의 수행 범위는 크게 ‘글쓰기 자료 수집’, ‘말뭉치 지식 강연회 개최’의 두 부분으로 나눌 수 있으며, ‘글쓰기 자료 수집’ 사업은 세부적으로 계획 수립, 추진 위원회 구성 및 운영, 홍보, 글쓰기 자료 수집, 글쓰기 자료 수집 및 이용 계약을 통한 저작권 확보, 수집 저작물 목록 및 정보 정리로 구성되었다. ‘말뭉치 지식 강연회 개최’ 사업은 온·오프라인 강연회 홍보, 참가자 유치, 강연자 섭외, 강연회 세부 실행 및 개최로 나뉘 볼 수 있다.

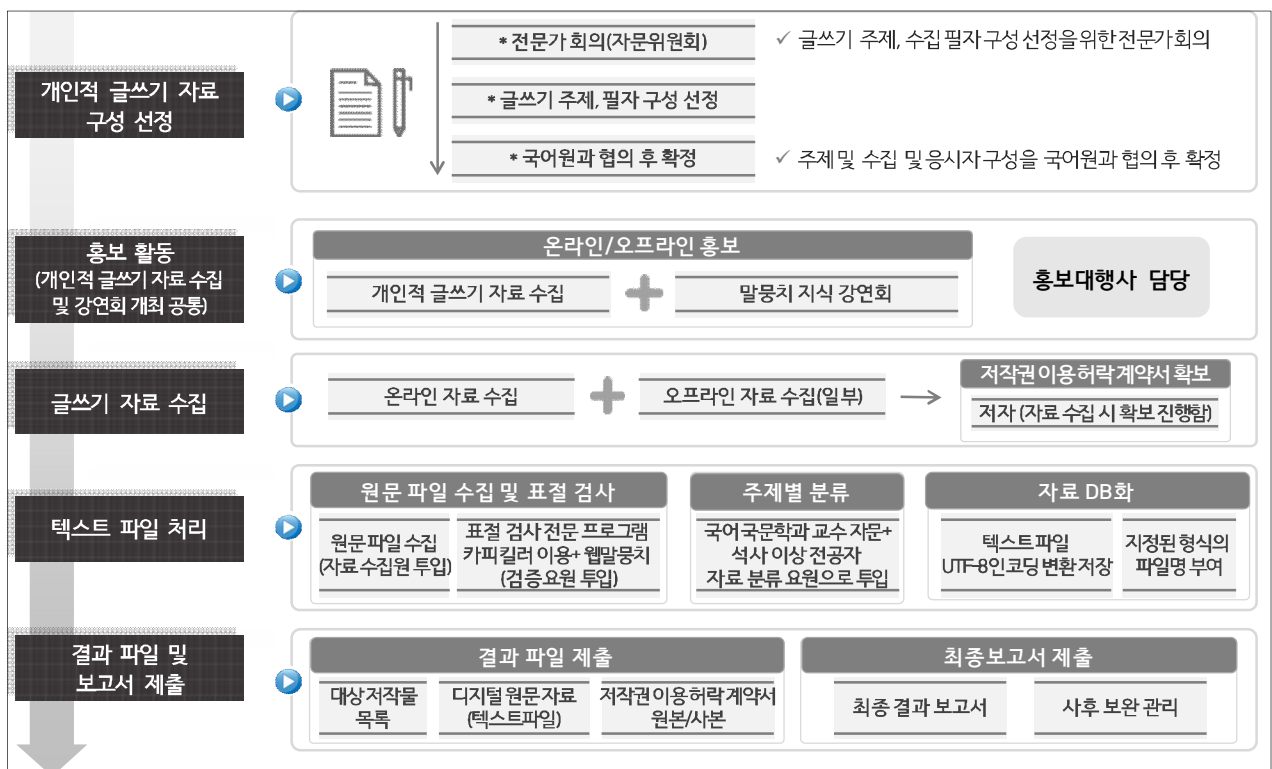
〈표 1〉 사업의 범위와 과업 내용

사업의 범위	과업 내용
글쓰기 자료 수집	<ul style="list-style-type: none"> ■ 글쓰기 주제 구성, 수집 방법 등 자료 수집 전반에 걸친 세부 계획 수립 ■ 글쓰기 자료 수집 추진 위원회 구성 및 운영(주제, 분류, 수집 방법 등) ■ 글쓰기 자료 수집 홍보: 온라인, 오프라인 홍보 진행 ■ 글쓰기 자료 수집 실시 ■ 글쓰기 자료 수집 및 이용 계약을 통한 저작권 확보 ■ 수집 저작물 목록 및 정보 정리
말뭉치 지식 강연회 개최	<ul style="list-style-type: none"> ■ 강연회 세부 실행 계획 수립 ■ 강연회 추진 위원회 구성 및 운영 ■ 강연회 홍보: 온라인 및 오프라인 홍보 ■ 참가자 유치 및 등록 방안 수립 ■ 강연회 개최

1.3 사업 수행 절차

개인적 글쓰기 자료 수집의 구성을 위한 전문가 회의를 진행하고 이를 통한 구성안을 선정한 후, 온라인 글쓰기 자료 수집 시스템을 구축하고, 온라인 및 오프라인 홍보 진행 후 온라인 자료 수집을 위주로 자료를 수집하고, 온라인에서 수집하기 어려운 연령의 글은 오프라인으로 자료를 수집하였다. 글쓰기 자료 수집과 동시에 저작권 이용 허락 계약서를 확보하면서 진행하였다.

다음으로 수집된 원문 자료에 대한 표절 검사를 시행해서 표절률을 확인하고 표절률이 있는 글을 확인해서 표절로 판명됐을 시에는 자료로 포함하지 않았다. 표절 검사를 통과한 글은 글감을 저자가 ‘자유 주제’ 또는 ‘자유 글감’을 선택하였거나 작성하지 않은 경우에 자료 분류 요원이 글감을 입력하고 글의 메타 정보 및 글이 입력된 완성된 최종 파일은 UTF-8 인코딩 변환 저장하였다. 결과 파일은 대상 저작물 목록, 디지털 원문 자료(텍스트 파일), 저작권 이용 허락 계약서를 제출하였다.

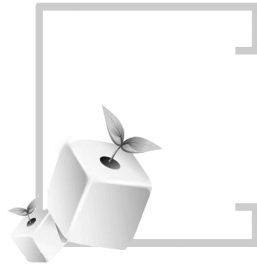


〈그림 2〉 개인적 글쓰기 자료 수집 사업 수행 절차

‘말뭉치 지식 강연회’는 2월에 오프라인 지식 강연회 형식으로 개최할 계획이었으나, 코로나바이러스감염증-19로 인해 한 차례 연기해 6월에 온라인 방식으로 개최하였다. 말뭉치 지식 강연회를 성공적으로 개최하기 위해 온·오프라인 홍보를 진행하였다. 주로 유관 기관 종사자 및 관련 기업 관계자를 대상으로 홍보했으며, 관련 학과 학생 및 유관 분야 취업 희망자 위주로 강연회 참가를 유도했다. 국민 접점 홍보를 위해 공공 기관 보유 매체 및 사업 수행자 보유 네트워크를 활용하여 강연회 홍보 영상을 오프라인 매체에서 3주간 송출하였다.

말뭉치와 말뭉치 지식 강연회를 널리 알리기 위해 누리 소통망을 활용했다. 말뭉치 관련 핵심 내용을 압축하여 ‘말뭉치’와 같은 생소한 정보를 효과적으로 노출하고 행사를 알릴 수 있도록 홍보물을 제작했다. 말뭉치 지식 강연회 개최 전 보도 자료 초안 작성을 지원하여 언론 홍보를 보조했다.

본 행사는 최종적으로 온라인상에서 개최했으며, 강연 자료 유출을 막기 위해 외부 송출 플랫폼, 쿡쿡(CongKong)을 활용했다. 사전 신청자에게 개별 입장 코드를 발송하여 강연회에 입장할 수 있도록 했다.



제 2 장

사업 수행



2.1 개인적 글쓰기 자료 수집

2.1.1 글쓰기 자료 수집 개요

본 과업은 개인적 글쓰기 자료 7,000건 수집을 목표로 발주되어 7,000명의 글을 수집하기로 협상하고 과업을 추진하였다. 그러나 코로나바이러스감염증-19의 유행으로 말뚝치 지식 강연회를 6월로 연기하며 위약금 등 비용 손실이 발생하였다. 이에 개인적 글쓰기 자료 수집 목표량을 하향 조정하여 계약 금액 내에서 손실 비용을 마련하였다. 개인적 글쓰기 자료 수집 과업의 목표는 6,000명으로 변경하였다.

〈표 2〉 글쓰기 자료 수집 목표

초기 목표 설정(사업계획변경 전)			변경된 목표 설정(사업계획변경 후)		
구분		모집 수(명)	구분		모집 수(명)
학생	초등	1,200	학생	초등	1,000
	중등	1,200		중등	1,000
	고등	1,200		고등	1,000
	대학(원)	1,300		대학(원)	1,100
일반(성인)	중장년	1,300	일반(성인)	중장년	1,100
	노년	800		노년	800
계		7,000	계		6,000

글쓰기 자료 실제 수집량은 6,044명, 10,919편이며 글자 수는 9,250,097(공백 포함 전체)자로 목표를 초과 달성하였다.

〈표 3〉 글쓰기 자료 수집 결과

최종 수집 결과				
구분		모집 수(명)	수집 글(편)	글자 수(자)
학생	초등 (만 7세 ~ 만 12세)	800	3,119	1,063,836
	중등 (만 13세 ~ 만 15세)	497	854	765,639
	고등 (만 16세 ~ 만 18세)	1,058	1,688	1,754,297
일반(성인)	대학(원) 외 (만19세 ~ 만 29세)	2,720	3,926	4,325,375
	중장년 (만30세 ~ 만 59세)	648	862	899,281
	노년 (만 60세 이상)	321	470	441,669
계		6,044	10,919	9,250,097

2.1.2 글쓰기 주제, 수집 방법 선정

2.1.2.1 자문 위원회 구성

국어 관련 전문가들과 홍보 전문가로 자문 위원회를 구성하였으며, 자문 회의나 개별 인터뷰를 통해 글쓰기 자료 수집 사업 추진 방법에 대한 의견을 듣고 반영하였다.

〈표 4〉 자문 위원 구성

자문 위원	소속 및 전공	활동 내역
박정희	<ul style="list-style-type: none"> 서울대학교 교수학습개발센터 연구교수 국어국문학 전공 	자문 회의 참석 자문 의견서 제출
김미형	<ul style="list-style-type: none"> 상명대학교 국어문화원장 국어국문과 전공 	자문 회의 참석 자문 의견서 제출
박상규	<ul style="list-style-type: none"> 전국국어교사 모임 사무총장 목동고등학교 교사 국어과 	자문 회의 참석 자문 의견서 제출
우경숙	<ul style="list-style-type: none"> 서울구로초등학교 교사 	인터뷰 자문 자문 의견서 제출
이정옥	<ul style="list-style-type: none"> 양원주부학교 홍보부장 (사)한국문해교육협회 학습자지원분과 이사 	인터뷰 자문 자문 의견서 제출
이일우	<ul style="list-style-type: none"> 한국외식업중앙회 광고홍보학 박사 	공모 제목 자문 자문 의견서 제출

2.1.2.2 자문 위원회 회의 개최

자문 회의 안건은 글의 종류 구분(연령별, 내용별, 분량)과 참여자 구성 비율 적정성, 글감과 장르의 정의 등으로 세부적인 내용은 다음과 같다.

〈표 5〉 자문 회의 논의 내용(2019. 12. 9. 개최)

구분	내용
① 글의 종류 구분	<p>연령</p> <p>유아동: 그림일기, 편지 초등: 일기, 편지, 동시, 기행, 동요(가사), 독서 일기, 학교 과제물 포함 여부 중고등, 대학(원): 논술, 학교 과제물 포함 여부</p>
	<p>내용</p> <p>- 전문 분야 관련 글의 포함 여부: 대학생 학교 수업에서 작성한 레포트 포함 여부 ex) 국어국문학과 수업 등 글쓰기 수업 참여시 과제물 제출 가능 여부 등 - 자기소개서 등 특수 목적의 글 포함 여부 - 맛집 및 음식 관련 등 비완결적인 글 포함 여부 - 사진, 그림, 동영상 포함 여부(그림일기, 사진 용량 등) - 특정 주제 제시 또는 자유 주제 글쓰기 가능 여부 등</p>
	<p>분량</p> <p>- 글의 분량(글의 장르별 분량 제시: 일정 기준 분량 이상) - 1인당 1편 제출 또는 2~3편 제출 가능 여부: 시, 가사의 경우 3편 제안 논의 - 파일을 올리는 경우: 장르별, 연령별 기준 제시 필요 예) 산문-초등학생 1~3장, 중고등학생 2장 이상, 성인 3장 이상</p>
② 응시자 구성 비율 적정성 (지역, 성별 일정 비율 필요 여부)	<p>유아동 350(5%), 초등 1,190(17%), 중등 1,190(17%), 고등 1,190(17%), 대학(원) 1,190(17%), 중장년 1,190(17%), 노년 700(7%)</p>
③ 글감과 글의 종류 구분 방식	<p>참여자도 글감을 제출하면서 선택하는 방식</p>

2.1.2.3 자문 위원 의견 취합 및 확정

자문 회의와 자문 위원과의 인터뷰 이후 각 위원들의 자문 의견서를 취합하였으며 의견을 정리하여 사업에 반영하였다.

자 문 의 건 서	
1. 개인적 글쓰기 자료 수집의 전체 방향 1) 국어 빅데이터 사업의 일환, 향후 이 자료는 인공 지능 개발을 위한 자료, 기타 국민 글쓰기 현황을 살펴보는 자료 등으로 활용 2) 훌륭한 글쓰기 공모가 아니므로 누구나 참여 의욕을 가질 수 있도록 최소한의 분량 (원고지 5매 이상)으로 하고 최대 분량도 제한을 두지 않는 것이 좋다. 3) 자료 활용 가치를 위해 장르와 주제 마킹이 필요하겠으나, 국민이 자유롭게 자기 글을 올릴 수 있게 하기 위해 장르와 주제 매칭을 반드시 하게 하지는 말자. 최대한 자유롭다고 느끼게 하자. 4) 무엇보다 국민에게 널리 홍보하는 일이 최대 관건이므로 언론 보도 유발 필요하다. 5) 비출판이라는 요건을 정확히, 잘 알릴 수 있어야 하고, 저작권 관련 통의서를 별리 갖추어야 한다.	
2. 참여 대상자 1) 초등학교 1학년부터 한글 교육과 받아 쓰기, 그런 일기 교육이 시작되므로 유치원 아동은 제외하자. 2) 다문화 가족, 이주자들도 대한민국 국적이면 참여 가능, 특별히 이를 구분할 필요 없다. 3) 참여자 신원 구분 항목을 체계적으로 해야 함. (예를 들어 대학원생과 일반인 구분은 어떻게 할 것인가)	
3. 글의 종류 1) 학생 과제물은 국어 시간 글쓰기 과정은 배웠으나 특정 과목의 과정은 개인적 글쓰기가 아님. 2) 굳이 소설을 제의할 필요는 없음. 3) 글의 종류를 제시하는 목적은 국민 이해를 돕기 위한 것이지 그들로 엄격히 구분하기 위한 것이 아니므로 국민의 개인 글쓰기 상황으로 제시하면 된다. (예를 드는 형식으로, 수필, 기행문, 감상문, 편지, 일기, 연설문, 시 등) 4) 자기소개서는 그 글이 어딘가에 제출되었던 것은 제외하는 것으로 해야 한다.	
4. 참여 유도 방안 1) 초등, 중등 교사 협의회 활용 (지도 사례비 필요) 2) 문예 소모임이나 동아리 참여 유도 (경품 추첨 등)	
2019년 12월 10일 소 속 : 삼명대학교 직 위 : 교수 성 명 : 김미형 인	

자 문 의 건 서	
1. 다양한 자료 확보를 위해 장르를 제한하는 것보다는 응시 대상을 제한하는 방식이 적절해 보임. - 장르를 제한하지 않되, 장르의 예시를 보여줄 필요는 있음. 2. 연령에서 청소년은 학교 안 학생과 학교 밖 학생의 구분이 필요함. - 유치원 / 초등 1,2학년 / 초등 3,4학년 / 초등 6학년 / 중학교 / 고등학교 / 대학교 - 초등학생은 성장 단계에 따라 글쓰기 수준의 차이가 크므로, 초등 교사의 자문이 필요함. - 학교 밖 청소년의 경우, 하자 센터나 소년원 등 청소년이 모여 있는 기관이나 단체의 협조 필요함. 3. 제한된 시간에 효율적으로 자료를 수집하기 위해 초중등의 경우 과제물을 포함할 필요가 있음. - 학교 현장에서 다양한 종류의 글쓰기 수행 활동이 이루어지고 있음. : 감상문, 서평, 보고서, 시, 노래가사, 캐리커처, 토론 개요서, 행사 후 소감 등 - 교사들이 가지고 있는 자료를 수집하는 것이 효율적일 수 있으나 행정적, 실무적 검토 지원 필요함. - 교사들이 이 과정에 대한 이해도가 가장 높을 수 있음. - 전국국어교사모임이나, 초등국어교사모임 등과 협조하는 방안도 모색할 필요가 있음. - 국어와 외에도 다양한 교과에서 글쓰기 수행 활동이 이루어지고 있음. 다양한 교과와 교과 모임과의 접촉도 필요해 보임. 4. 개인 글 공모에도 개인적 접근과 함께 모임을 통한 접근이 필요해 보임. - 학교 안과 청소년: 다양한 동아리를 통한 집단 공모 - 대학생: 학과, 글쓰기 동아리, 다양한 동아리를 통한 집단 공모 - 중장년: 글쓰기 동아리, 여행이나 독서, 스포츠 등 취미생활 동호회, 문화센터, 노인정, 마을회관 등 5. 연령대별로 최소 분량 제시가 필요함. - 유아: 20자 / 초등1,2: 50자 / 초등3,4: 200자 / 초등5,6: 300자 / 중등: 400자 / 고등 600자 ~ 6. 개인적 글쓰기의 개념 및 사업의 목적에 대한 구체적 안내가 필요함.	
2019년 12월 10일 소 속 : 전국국어교사모임 직 위 : 사무총장 성 명 : 박상규	

자 문 의 건 서	
<input type="checkbox"/> '개인적 글쓰기'의 범주 설정 문제 - '미간행, 미공개 글쓰기 자료'(SNS 글, WEB 글 제외)에 초점 맞추기 - 기 작성된 글의 성격 가운데 자유로운 목적과 형식으로 작성된 글을 포함시킴	
<input type="checkbox"/> 연령별 글쓰기 특징을 고르게 반영할 수 있는 방안 - 초중고/대학생의 경우 자발적 참여와 단체참여 유도가 수월할 것으로 예상됨 - 30~40대의 글쓰기 자료 수집에 어려움이 예상됨(특히 남성) - 개인별 사이트 접속 참여 방법과 아울러 동아리 등 각종 단체가 참여할 수 있는 방안 필요 * 중장년, 노년층의 자료 확보를 위해 지역 및 각종 동아리 등 단체 참여 방안 모색	
<input type="checkbox"/> 글의 주제 및 형식(종류) 등 - 사이트 구축: <기 작성한 글 탐색> 코너와 <참여형 작성> 코너를 나누어 2트랙으로 진행하는 방안 <참여형 작성> - 글의 형식 및 장르를 카테고리를 정하기보다 많은 사람이 흥미를 가지고 참여하여 부담없이 작성할 수 있도록 하는 방안이 적절해보임 - 글감 혹은 상황을 부여하고 '자유롭게' 작성하는 글쓰기 지향 - 분량은 고교 등을 고려하여 A4 1매(200자 인고지 10매) 이상으로 작성 * 평가 대상의 글이 아니라는 점을 강조하여 전문적인 지도를 받은 글이 탑재되지 않도록 유의시키는 일이 중요(맞춤법부터 완결성까지 '스스로' 작성한 글 수집) <기 작성한 글 탐색> - '안내사항'을 구체적으로 작성할 필요가 있음 - 글의 형식이나 종류를 강조하지 않을 때 더 많은 자료를 수집할 수 있을 것임 - 기 작성한 글의 예로 이런 것들이 있을 수 있다는 식의 다양한 사례 제시	
2019년 12월 10일 소 속 : 서울대학교 교수학습개발센터 글쓰기교실 직 위 : 연구교수 성 명 : 박정희 인	

'개인적 글쓰기 자료 수집 및 말동지 지식 강연회 개최' 자문회의	
자 문 의 건 서 말동지 모음에 대해 초등학생 참여와 글쓰기에 대한 의견입니다. *생활글 특성-초등어린이들의 경우 -초등학생은 생활글쓰기(경문일쓰기)에 매우 왕성한 표현욕구를 갖고 있습니다. 완성된 문장은 아니더라도 사물, 친구, 육구, 갈등 등 한 말이 참 많습니다. -생활글쓰기에 대한 자료수집을 예시자료별로 제시가 필요함. (예: 일요일, 친구, 핸드폰, 편의점, 학원 등) -친구 이름 등 실명은 꼭 정보보안 처리해주시기 바람. -이 글쓰기는 수업을 목적으로 하지 않음을 뚜렷하게 밝힘. -초등생들의 언어는 말과 글, 글이 서로 닮음. -저 중 고학년에 공동적으로 글쓰기 분량은 짧은 편이다.(전작발, 매체 자막 등 영향) -분량 몇 자 이상이 익숙하지 않으니 예시자료로 제시가 필요함. -어린이들의 분량이 단분이고 구어에 가까운 글쓰기가 많음. -글의 종류는 생활글, 시조가 어린이들이 가장 개성이 뚜렷함. -정서, 육구와 갈등이 가장 솔직하게 드러나는 것이 시조기임. *글 수집 -개인글에 대한 보호자 동의의 여부가 절차적으로 명확 -자료로서 활용이 있는 글이라 글 올릴 동기가 생김. -홍보는 sns로 교사동아리나 교사모임에 연결. -글쓰기에 대한 보상이 커서 유인 효과가 큼.	
2019년 12월 26일 소 속 : 서울구로초등학교 직 위 : 교사 성 명 : 우경숙 인	

<그림 3> 자문 의견서

글쓰기 자료 수집의 구체적인 실행 방법은 다음과 같이 정하였다. 초등학교부터 정식으로 글씨를 배우는 것을 고려하여 초등학생부터 글쓰기 자료를 수집하기로 하였다. 또한 단기간(한 달 이내) 많은 글을 수집하여야 하므로 누리집을 구축해 온라인상에서 수집하기로 하였다. 수집 대상 글의 종류는 시(동시), 일기, 편지글, 소설(동화), 감상문, 기타로 분류하였으며 글의 종류 및 주제에 제한을 두지 않았다.

비출판 글을 기준으로 하였으므로 출판은 물론 블로그 등에도 공개되지 않은 글로 참여를 제한했다. 1인당 글자 수는 공백 포함 1,000자 이상으로 최소 한 편당 50자 이상 작성 가능하며 최대 10편까지 제출할 수 있도록 했다. 제목은 참여자가 직접 입력하거나 ‘제목 없음’을 선택하게 하였으며, 글감은 시스템에서 보여주는 글감을 선택하거나 ‘자유 글감’을 선택하도록 하였다. 작성 연도는 제출하는 글을 처음 쓴 해를 적게 하였다. 또한 초등학생에 한해 손 글씨로 쓴 원문(이미지, 피디에프 파일)을 첨부하도록 하였다. 초등학생 참여자도 글쓰기 창에 글을 입력하게 하였으나 이는 글자 수(1,000자 이상) 확인을 위함이며, 실제 원문 입력은 전문 요원이 별도로 진행하였다.

〈표 6〉 글쓰기 자료 수집 글의 구체적인 내용

구분	내용
대상	초등학생 이상 전 국민, 1인당 1회의 참여로 제한
참가 방법	누리집(http://www.iwrite.kr)에서 글 작성: 누리집 접속이 어려운 경우에 별도 접수
수집 대상 글의 종류	·시(동시) ·일기 ·편지글 ·소설(동화) ·감상문 ·기타 ※ 글의 종류 및 주제 제한 없음
비출판 글	·어디에도 공개(출판, 본인 및 타인의 블로그 게시 등)되지 않은 글
글의 분량	·1인당 글자 수 1,000자 이상 제출(50자 이상씩 최대 10편 제출 가능)
제목	제목 혹은 ‘제목 없음’ 입력
글감	‘글감 뽑기’에서 선택하거나 ‘자유 글감’ 선택
작성 연도	제출하는 글을 처음 쓴 해
원문 제출 대상	초등학생에 한해 손 글씨로 쓴 원문(이미지, 피디에프 파일)을 제출하고 글자 수 확인을 위해 글쓰기 창에도 글을 입력함

2.1.3 글쓰기 자료 수집 공모 제목 선정

글쓰기 자료 수집 사업의 공모 제목을 선정하기 위한 여러 후보안이 제안되었으며, 회의 시 자유 의견 취합과 논의 등의 과정을 통해 최종적으로 ‘한 편 줍쇼’로 결정했다. 제목 선정은 사업의 성격을 반영하면서 기억하기 쉽도록 하는 데에 주안점을 두었다.

〈표 7〉 글쓰기 자료 수집 글의 공모 후보안 선정 과정

시기	회의안	내용
19.12.3	회의 시 논의 후보	① 티글모아 빅데이터 모음전 ② 한글한글 모아 큰 정보를 이룬다 ③ 너도나도 모은 글이 큰(빅)데이터를 이룬다 ④ 전 국민 글 사랑 모음전 ⑤ 너도나도 모은 글이 빅 데이터가 된다. 글 모음전에 참여해주세요 ⑥ 내 글이 인공 지능이 된다고? ⑦ 누구나 참여할 수 있는 글쓰기에 응모해요! ⑧ 국민 모두의 글을 모집합니다 ⑨ 전국민 글쓰기 마당에 초대합니다 ⑩ 내 글이 국어원 자료가 되는 글쓰기에 응모해요! ⑪ 전국민 작가 만들기 프로젝트 "왜 이래? 나 원고료 받는 작가야!" ⑫ 목허든 글짓기 제출하고 원고료(상금) 받고 없으면 글쓰고 원고료(상금) 받자! ⑬ 떨어진 원고도 다시보자! 내기만 하면 원고료(상금)! ⑭ AI 살찌우는 글쓰기 공모 AI 살찌우고 4차산업 앞당기자! ⑮ 글짓기 찾고 쓰고 내고 받자! ⑯ 전국민 글쓰기 공모전 ⑰ 전국민 글쓰기 모음전 ⑱ 전국민 자유 백일장 ⑲ 인공 지능을 위한 전국민 글 모음 ⑳ 로봇이 말하는 내가 쓴 글 공모전
19.12.3	1차 후보안 선정	① 너도 나도 글쓰기 운동 ② 인공 지능 씨앗 글쓰기 ③ 전 국민 글 모이 운동 ④ 글 모으기 운동 ⑤ 너도 나도 글 모이 운동
19.12.10	2차 후보안 선정	① 인공 지능이 똑똑해지는 생활 속 글 찾기 ② 인공 지능이 똑똑해지는 일상 글 모으기 ③ 인공 지능이 똑똑해지는 내 글을 만들어 보자

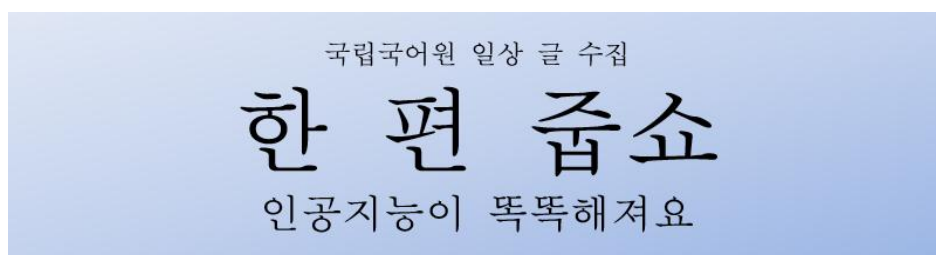
19.12.17	3차 후보안 선정	① 생활 속 보물 글 찾기 ② 인공 지능을 똑똑하게 하는 일상 글 공개수배 ③ 보통날의 보물글찾기(당신의 글로 인공 지능이 똑똑해집니다) ④ 인공 지능이 똑똑해지는 국립국어원 일상 글 수집 ⑤ 국립국어원 일상 글 공개 수배
19.12.24	공모 제목 선정	국립국어원 일상 글 수집 한 편 줄쇼 인공 지능이 똑똑해져요

2.1.4 ‘한 편 줍쇼’ 누리집 관련 디자인

2.1.4.1 공모 제목 디자인

공모 제목의 글자꼴 등은 ‘전 국민 대상 일상 글 수집’이라는 사업의 성격에 적합하게 정하였다. 홍보용 광고지와 소책자 인쇄를 고려하여 제작하였다.

1) 초기 글자



2) 시안 1



3) 시안 2



<그림 4> ‘한 편 줍쇼’ 제목 도안

2.1.4.2 누리집 초기 화면

누리집 초기 화면의 시안도 글자와 어울리도록 여러 시안을 대상으로 검토하였으며 수정 작업을 거친 후 최종 확정하였다. 누리집 초기 화면 디자인 시에는 홍보용 광고지와 소책자에도 동일하게 들어갈 수 있도록 제작하였다.

1) 시안 1



2) 시안 2



3) 최종 누리집 초기 화면



〈그림 5〉 ‘한 편 잡쇼’ 누리집 도안

2.1.4.3 누리집 저작권 이용 허락 계약 시스템 구축

누리집에서 참여자가 저작권 이용 허락 계약을 체결할 수 있도록 전자 서명이 가능한 계약 시스템을 구축하였다. 우선 저작권 이용 허락 계약서를 작성해서 완성하였으며, 이를 누리집 시스템에서 이용할 수 있도록 만들었다.

붙임 2.

귀하가 제출하신 저작물의 사용 목적과 이용 허락 대상 권리는 아래와 같습니다.

국립국어원은 대규모 언어 자료를 구축하여, 이를 국어 연구 및 자연언어처리 기술 개발 등을 위해 사용하고자 합니다. 이에 귀하의 저작물의 활용에 대한 승인을 구하고자 합니다. 귀하의 저작물을 이용하여 구축한 말뭉치가 귀하의 저작 출판에 관한 어떠한 권리에도 손상을 입히지 않을 것임을 약속드립니다.

모토록 21세기 4차 산업 혁명 시대에 귀하의 저작물을 우리말 정보 처리 발전의 기초가 되는 국가적 언어 자료의 구축에 유용하게 활용될 수 있도록 협조하여 주시면 감사하겠습니다. 이를 위해 이용 허락에 앞서 아래 내용을 읽어 주시길 바랍니다.

- 국립국어원 일상 글 수집 '한 편 중소' 참가 본인은 제출 저작물의 제출된 모든 내용 및 문서는 본인이 작성한 저작물임을 확인합니다.
- 제출한 저작물이 제3자와 분쟁소지 가능성이 있는 저작물에 대해서는 제출하지 않습니다.
- 저작물이 제3자의 소유권, 저작권, 초상권 등 기타 제반 권리를 침해하지 않았음을 보증하여야 하며, 향후 분쟁이 발생할 경우 일체의 법적 도의적 책임은 해당 저작물을 제출한 제출자에게 있습니다.
- 타 공모전에서 시상 또는 수상 경력이 있는 작품, 표절 작품으로 판명될 경우 국립국어원 사용 용도에서 제외되며, 사례금은 지급되지 않습니다.
- 제출된 저작물은 이용허락 기간 내에서 국립국어원에서 사용하며, 제출자가 이용 허락 중지 의사를 밝히면, 의사 내용에 따라 이용 허락을 중지할 수 있습니다.
- 제출된 저작물은 개별 단어 및 문장, 텍스트에 대한 복제·변형(편집 및 삭제, 정리, 분리 등) 및 정보 추출에 쓰이며, 연구 및 기술 개발용으로 학계·연구·기관·사업체 등에 제공·배포하여 분석하는 데 쓰입니다.

국가 언어 자원(말뭉치) 구축 및 활용 저작권 이용허락 계약서.

저작자 및 저작권 이용허락자 (이하 "권리자"이라 함)와 저작권 이용자 국립국어원(이하 "이용자"이라 함)은 아래 저작물에 관한 저작권산권 이용허락과 관련하여 다음과 같이 계약을 체결한다.

다 음

제1조 (계약의 목적)
본 계약은 저작재산권 이용허락과 관련하여 권리자와 이용자 사이의 권리관계를 명확히 하는 것을 목적으로 한다.

제2조 (계약의 대상)
본 계약의 이용허락 대상이 되는 권리는 아래의 저작물(이하 "대상저작물")에 대한 저작재산권 중 당사자가 합의한 권리로 한다.

저작물: 저작자가 국립국어원의 <개인적 글쓰기 자료 수집 및 말뭉치 지식 강연회 개최> 사업 기간(2020년 1월 31일 ~ 2020년 3월 14일까지) 동안 위 사업에 제공하는 모든 개인 비출판물


저작자:
동별: ☒ 어문저작물
권리: ☒ 복제권, ☒ 전송권, ☒ 배포권, ☒ 2차적저작물작성권

※ 저작권 이용허락 대상 권리의 내용

1. 국립국어원 및 국립국어원이 발주한 용역 사업의 수행자가 대상저작물을 일정한 형식으로 전자적 기록 매체에 담아 보존하는 일
2. 국립국어원 및 국립국어원이 발주한 용역 사업의 수행자가 자모, 음절, 어휘, 어절, 구절, 문장 및 텍스트 단위의 국어 연구와 언어 정보 처리 분야에 응용하기 위해 대상저작물을 복제·변형(복사·머리말·도표·그림·각주 등의 편집 및 삭제, 언어 단위별 분리, 언어적 비언어적 정보 부각 등)하는 일
3. 국립국어원 및 국립국어원이 발주한 용역 사업의 수행자가 대상저작물 및 그 복제·변형물을 연구 및 기술 개발용으로 학계·연구기관·산업체 등이 이용할 수 있도록 제공·배포하는 일
4. 대상저작물 및 그 복제·변형물을 제공·배포받은 학계·연구기관·산업체 등이 국어 연구와 언어 정보 처리 분야 응용을 위하여 대상저작물 및 그 복제·변형물을 분석 및 처리하여 사용하는 것을 허락하는 일

□ 상기 내용을 확인하였으며, 동의합니다

<그림 6-1> 저작권 이용 허락 계약서(오프라인용)



귀하가 제출하신 저작물의 사용 목적과 이용 허락 대상 권리는 아래와 같습니다.

국립국어원은 대규모 언어 자료를 구축하여, 이를 국어 연구 및 자연언어처리 기술 개발 등을 위해 사용하고자 합니다. 이에 귀하의 저작물의 활용에 대한 승인을 구하고자 합니다. 귀하의 저작물을 이용하여 구축한 말뭉치가 귀하의 저작 출판에 관한 어떠한 권리에도 손상을 입히지 않을 것임을 약속드립니다.

모토록 21세기 4차 산업 혁명 시대에 귀하의 저작물을 우리말 정보 처리 발전의 기초가 되는 국가적 언어 자료의 구축에 유용하게 활용될 수 있도록 협조하여 주시면 감사하겠습니다. 이를 위해 이용 허락에 앞서 아래 내용을 읽어 주시길 바랍니다.

- 국립국어원 일상 글 수집 '한 편 중소' 참가 본인은 제출 저작물의 제출된 모든 내용 및 문서는 본인이 작성한 저작물임을 확인합니다.
- 제출한 저작물이 제3자와 분쟁소지 가능성이 있는 저작물에 대해서는 제출하지 않습니다.
- 저작물이 제3자의 소유권, 저작권, 초상권 등 기타 제반 권리를 침해하지 않았음을 보증하여야 하며, 향후 분쟁이 발생할 경우 일체의 법적 도의적 책임은 해당 저작물을 제출한 제출자에게 있습니다.
- 타 공모전에서 시상 또는 수상 경력이 있는 작품, 표절 작품으로 판명될 경우 국립국어원 사용 용도에서 제외되며, 사례금은 지급되지 않습니다.
- 제출된 저작물은 이용허락 기간 내에서 국립국어원에서 사용하며, 제출자가 이용 허락 중지 의사를 밝히면, 의사 내용에 따라 이용 허락을 중지하여야 합니다.
- 제출된 저작물은 개별 단어 및 문장, 텍스트에 대한 복제·변형(편집 및 삭제, 정리, 분리 등) 및 정보 추출에 쓰이며, 연구 및 기술 개발용으로 학계·연구·기관·사업체 등에 제공·배포하여 분석하는 데 쓰입니다.

국가 언어 자원(말뭉치) 구축 및 활용 저작권 이용허락 계약서

저작자 및 저작권 이용허락자 (이하 "권리자"이라 함)와 저작권 이용자 국립국어원(이하 "이용자"이라 함)은 아래 저작물에 관한 저작권산권 이용허락과 관련하여 다음과 같이 계약을 체결한다.

다 음

제1조 (계약의 목적)
본 계약은 저작재산권 이용허락과 관련하여 권리자와 이용자 사이의 권리관계를 명확히 하는 것을 목적으로 한다.

제2조 (계약의 대상)
본 계약의 이용허락 대상이 되는 권리는 아래의 저작물(이하 "대상저작물")에 대한 저작재산권 중 당사자가 합의한 권리로 한다.

저작물: 저작자가 국립국어원의 <개인적 글쓰기 자료 수집 및 말뭉치 지식 강연회 개최> 사업 기간(2020년 01월 31일 ~ 2020년 3월 14일까지) 동안 위 사업에 제공하는 모든 개인 비출판물

저작자:
동별: ☒ 어문저작물
권리: ☒ 복제권, ☒ 전송권, ☒ 배포권, ☒ 2차적저작물작성권

※ 저작권 이용허락 대상 권리의 내용

1. 국립국어원 및 국립국어원이 발주한 용역 사업의 수행자가 대상저작물을 일정한 형식으로 전자적 기록 매체에 담아 보존하는 일
2. 국립국어원 및 국립국어원이 발주한 용역 사업의 수행자가 자모, 음절, 어휘, 어절, 구절, 문장 및 텍스트 단위의 국어 연구와 언어 정보 처리 분야에 응용하기 위해 대상저작물을 복제·변형(복사·머리말·도표·그림·각주 등의 편집 및 삭제, 언어 단위별 분리, 언어적 비언어적 정보 부각 등)하는 일
3. 국립국어원 및 국립국어원이 발주한 용역 사업의 수행자가 대상저작물 및 그 복제·변형물을 연구 및 기술 개발용으로 학계·연구기관·산업체 등이 이용할 수 있도록 제공·배포하는 일
4. 대상저작물 및 그 복제·변형물을 제공·배포받은 학계·연구기관·산업체 등이 국어 연구와 언어 정보 처리 분야 응용을 위하여 대상저작물 및 그 복제·변형물을 분석 및 처리하여 사용하는 것을 허락하는 일

□ 상기 내용을 확인하였으며, 동의합니다.

<그림 6-2> 저작권 이용 허락 계약서(온라인용)

저작권 이용 허락 계약서와 관련해서는 법무법인(유미)에 의뢰해서 저자명 공개 동의 여부 등에 대한 법률 자문을 받아 진행하였다.

2.1.4.4 누리집 개인 인증 시스템 구축

전자 계약 시에는 본인 인증을 하였으며, 미성년자의 경우에는 법정 대리인의 동의 절차도 포함하였다. 저작권 이용 허락 계약 당사자가 본인의 글을 1,000자 이상 제출하고 표절 검사를 통과하면 모바일 문화 상품권을 지급하였다. 법적으로 만 18세 이하의 상품권을 지급받거나 개인 정보를 제공해 저작권 이용 허락 계약을 체결하는 행위 등을 할 때 법정 대리인(부모 등)의 동의가 법적으로 필요해 추가 인증 절차를 거치게 했다.

만 14세 미만은 나이스 아이핀 인증으로 보호자의 인적 사항을 확인하고 법정 대리인의 동의를 받은 것으로 하였다. 이를 위해 누리집 시스템에 기본적인 휴대 전화 개인 인증 장치와 함께 부모와 자녀를 확인할 수 있는 나이스 아이핀 인증 시스템을 구축하였다.

만 14세 이상부터 18세 이하의 경우에는 누리집에서 저작권 이용 허락 계약을 체결했다고 하더라도 부모가 이러한 계약 체결에 동의하는 ‘법정 대리인 동의서’ 및 법정 대리인과의 관계를 확인할 수 있는 ‘가족 관계 증명서’ 등을 필수로 제출하도록 하였다.

〈표 8〉 누리집 개인 인증 시스템

만 14세 미만	만 14세~18세
‘한 편 줍쇼’ 누리집 회원 가입 시 부모가 자녀의 나이스 아이핀 인증으로 자녀 명의로 회원 가입: 부모·자녀 관계가 인증되므로 추가 인증 서류가 필요하지 않음	‘한 편 줍쇼’ 누리집 회원 가입 시 만 14세~18세 청소년 본인이 회원 가입: 법정 대리인이 해당 권리자의 계약 체결에 동의하는 ‘법정 대리인 동의서’와 법정 대리인과의 관계를 증명하는 ‘가족 관계 증명서’ 등 제출

2.1.4.5 누리집 메뉴 설정

‘한 편 줍쇼’의 누리집 메뉴는 참여자들의 편의를 고려하여 구성하였다. [행사 소개], [참가 방법], [제출하기], [알림]은 누리집에서 바로 들어갈 수 있게 누리집 상단에 배치하였다. [참가 방법]은 참가 대상과 행사 기간, 참가 방법, 수집 내용, 참고 사항, 제출 방법 등을 안내하였다. 본 행사에서는 글 제출 시 저작권 이용 허락 계약서를 체결하는 점이 중요하므로 미리 저작권 이용 허락 계약서를 검토하도록 미리보기 기능을 제공하였다. [나의 메뉴]에서는 회원의 정보를 관리하고, 저작권 계약을 맺고 글을 작성하는 등 회원이 본인의 활동을 총괄할 수 있다. 글 작성은 누리집 상단에 [제출하기]를 배치하여 누리집 사용자의 접근이 쉽게 하였다. [알림]은 행사 기간 중요 알림 사항을 게시하였으며, ‘내 글이 비출판물 말뭉치가 된다?’ 글을 게시하여 참여자의 글이 어떻게 쓰이는지 안내하기도 하였다.

[행사 소개]

행사의 의미, 참여 대상, 누리집의 기능, 문의처, 국립국어원 국어 말뭉치 사업 소개 등

[참가 방법]

참가 대상, 수집 기간, 참가 방법, 수집 내용, 참고 사항, 제출 방법, 예시, 문의처, 주최, 주관에 대한 안내 내용 등

[나의 메뉴]

내 정보 변경, 계약하기, 글 작성하기, 회원 탈퇴

[알림]

행사 기간 동안 중요한 알림 사항 게시 계약, 법정 대리인 동의서, 문화 상품권 지급, 홍보 행사, 일부 연령대 마감 등 안내

〈그림 7〉 ‘한 편 줍쇼’ 누리집 메뉴 설정

2.1.5.2 온라인 홍보: 누리 소통망 행사, 온라인 커뮤니티

누리 소통망 행사

행사 사전 홍보를 위해 국립국어원 페이스북에서 누리 소통망 1차 행사를 진행하였다. 총 445명이 참여하여 10명의 당첨자를 선정하였다. 2차 이벤트도 페이스북에서 진행하였으며, ‘한 편 줍쇼’에 참여한 후 인증하는 방식으로 총 73명이 참여해 23명의 당첨자를 선정하였다.

국립국어원 심표, 마침표.
2월 3일 · 🌐

[홍보 행사 당첨자 발표] 기쁨은 나누면 배가 되는 법!
친구들에게 '한 편 줍쇼' 알리기 행사 당첨자 발표

'한 편 줍쇼' 홍보 행사에 참여해 주신 모든 분들께 감사드립니다.
공정한 추첨을 통해 총 열 명의 당첨자를 선정했습니다.
아래 주소에 개인 정보를 입력해 주세요.

개인 정보 입력하기 ▶ <https://c11.kr/d1ng>

* 기한 내에 경품 수령을 위한 개인 정보 수집에 회신하지 않으면 당첨이 취소됩니다.

#국립국어원 #한편줍쇼 #일상글 #댓글홍보행사 #소환홍보행사
#인공지능 #빅데이터 #말뭉치

국립국어원
기쁨은 나누면 배가 되는 법

친구들에게 한 편 줍쇼 알리기!!

당첨자 발표

안상윤	강창국
이주환	최수미
이성조	이원규
아현정	선혜신
js Jang	양현김

당첨되신 모든 분들 축하드립니다.

당첨되신 분들은 2월 20일(목)까지 개인 정보 입력 부탁드립니다.
* 기한 내에 경품 수령을 위한 개인 정보 수집에 회신하지 않으면 당첨이 취소됩니다.

‘1차 행사’

국립국어원 심표, 마침표.
2월 3일 · 🌐

[홍보 행사] 인증만 해도 선물이 가득!
말뭉치 자료 모집에 참여하고, 인증하고, 갤럭시 버즈 가져가자!

1단계. 말뭉치 자료 모집 누리집 '한 편 줍쇼'에 들어가서 참여한다.
2단계. 자신의 페이스북 담벼락에 '한 편 줍쇼' 제출 완료 사진을 올린다.(전체 공개 필수)
3단계. 페이스북 댓글로 참여한 링크를 올리면 증포 완료!

* '한 편 줍쇼' 누리집 주소 : <http://iwrite.kr/main>

이 모든 절차를 완료해 주신 분들 중 추첨을 통해 소정의 상품을 드립니다!

1등: 갤럭시 버즈(1명)
2등: 인공지능 스피커(2명)
3등: 음료 상품권(20명)

참여 기간: 2. 3.(화) ~ 2. 16.(일)
당첨자 발표: 2. 17.(월) 국립국어원 심표, 마침표, 페이스북

※ 발표 후 기한 내에 경품 수령을 위한 개인 정보 수집에 회신하지 않으면 당첨이 취소됩니다.
※ 5만원 이상의 경품의 경우 상품 금액의 22%에 해당하는 각종 공과금이 부과되며, 이는 당첨자 본인 부담으로 진행됩니다.
(입금 금액 및 관련 내용은 당첨자 개인별로 연락 후 공지됩니다.)

댓글 인증 행사 문의: (주)레인보우커뮤니케이션 전화: 070-4808-3640, 전자우편: shjeung@rainbowpr.co.kr
한 편 줍쇼 글쓰기 문의: 한 편 줍쇼 운영 사무국((주)메트릭스, 전화: 080-801-8377, 전자 우편: iwrite@metrix.co.kr)

#국립국어원 #한편줍쇼 #참여행사 #댓글홍보 #인증행사
#인공지능 #빅데이터 #말뭉치

국립국어원
인증만 해도 선물이 가득

참여하고, 인증하고, 가져가자!

참여 기간: 2. 3.(월) ~ 2. 16.(일) 당첨자 발표: 2. 17.(월)

완전 룰잡이~
난 참 애하고 인증만 했는데 선물이 나오니!

‘2차 행사’

<그림 9> 누리 소통망 이용 홍보 행사

온라인 커뮤니티 홍보

온라인 커뮤니티를 대상으로 한 홍보는 총 3차에 걸쳐 진행하였다. 1, 2차는 총 358 곳을 대상으로 하였다. 소설 17곳, 인문학 15곳, 글쓰기 1곳, 빅 데이터 7곳, 공모전 51곳, 교육/청소년/어린이 42곳, 일반인 대상 120곳, 학생 대상 35곳, 이벤트 70곳의 커뮤니티에 ‘한 편 줍쇼’를 홍보하였다.

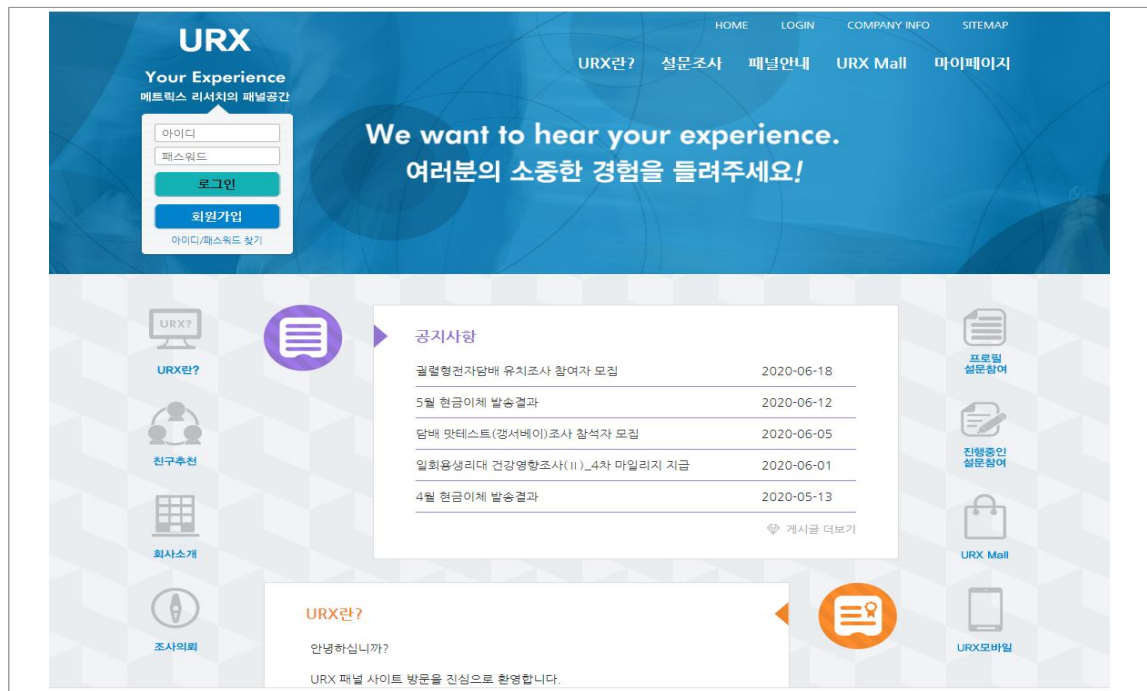
3차 홍보는 글쓰기 자료를 충분히 수집하지 못한 연령대인 초등학교와 중·고등학교, 만 60세 이상을 대상으로 진행하였다. 학부모 관련 커뮤니티 152곳과 노인 관련 온라인 커뮤니티 128곳에 대해 ‘한 편 줍쇼’를 홍보하고 참여 방법을 안내하였다.



〈그림 10〉 ‘한 편 줍쇼’ 1, 2차 온라인 홍보

(주)메트릭스 패널 대상 홍보

(주)메트릭스 패널 56,000명을 대상으로 ‘한 편 줍쇼’ 홍보 전자 우편을 발송하였으며 링크를 통해 참여를 유도하였다. 총 11회에 걸쳐 전자 우편을 발송하였다.



<그림 11> (주)메트릭스 패널 누리집

The screenshot shows a promotional page for the '한 편 줍쇼' (Pick a Side) promotion. It features a blue header with the text '국민국어원 일상 글 수집 "한 편 줍쇼" 참여 안내' (National Korean Language Institute Daily Writing Collection "Pick a Side" Participation Guide). The main content area has a light blue background and contains the following text:

안녕하십니까?
리서치 전문기관 (주)메트릭스입니다.
저희는 국립국어원과 함께 인공지능 산업 발전을 위해 일반 국민의 일상적인 글을 수집하고 있습니다.
아래 내용 및 홈페이지(<http://www.lwrite.kr/main>) 내용 참고하시어 많은 참여 바랍니다.
감사합니다.

At the bottom, there's a link: <http://www.lwrite.kr/main> 로 접속하셔서 참여 바랍니다.

<http://www.lwrite.kr/main> 로 접속하셔서 참여 바랍니다.

참여 방법

- ① 누리집(사이트) 접속 후 회원 가입하기
- ② 누리집(사이트) 상단 [제출하기]를 누르기
- ③ 참가자 연령 선택(만 14세 미만, 만 14세 이상)
- ④ 개인 정보 수집 및 저작권 이용 허락 동의서 체결
 - 전자 우편(이메일) 혹은 카카오톡 등 편한 방법으로 작성
- ⑤ 휴대 전화 본인 인증
 - 만 14세 미만은 법정 대리인(부모 등)이 휴대 전화 본인 인증
- ⑥ 개인 정보 입력
 - 이름, 생년월일, 휴대 전화 번호, 전자 우편(이메일), 비밀번호, 직업, 성별

※ 만 14세 미만인 청소년은 법정 대리인의 이름과 생년월일, 청소년의 이름과 생년월일, 성별 입력

- 만 14세 미만 청소년 '아이핀 인증' 필수

※ 아이핀: <http://www.niceipin.co.kr>에서 발급

⑦ 글 작성 및 제출하기

- [글 작성하기] 누르기
- 글감: [글감 뽑기]에서 고르거나 '자유 주제' 선택
- [작성 연도]: 제출하는 글을 처음 쓴 해
- * 작년에 쓴 글 - 2019, 이번 행사를 위해 새로 쓰는 글 - 2020
- [글 종류] 선택
- [제목]: 제목 혹은 '제목 없음' 입력
- [글쓰기]: 기존의 글을 입력하거나 새 글 작성
- 입력된 글 1,000자 이상 확인하고 최종 제출하기 누르기

※ 조중학생은 손 글씨로 쓴 글(이미지, PDF 파일)을 올리고, 글자 수 확인을 위해 글쓰기 창에도 글 입력, 글의 분량과 표절 검사는 손 글씨 파일의 글로 확인함(교정 없이 제출)

※ 본인 및 타인의 개인정보(주소, 이름, 생년월일 등)가 글에 포함되지 않도록 주의해 주십시오

감사합니다.

<그림 12> (주)메트릭스 패널 대상 발송 ‘한 편 줍쇼’ 홍보 전자 우편

2.1.6 글쓰기 자료 수집 현황

2.1.6.1 ‘한 편 줍쇼’ 누리집 글 수집 결과

글쓰기 자료 수집 결과는 다음과 같다. 만 60세 이상 321명, 만 30세 ~ 만 59세 648명, 만 19세 ~ 만 29세 2,720명, 만 16세 ~ 만 18세 1,058명, 만 13 ~ 만 15세 497명, 만 7세 ~ 만 12세 800명으로 총 6,044명의 글을 수집하였다.

〈표 9〉 연령별 ‘한 편 줍쇼’ 참여자 구성 결과

구분	제출자 수(명)	비중
만 7세 ~ 만 12세	800	13.2%
만 13세 ~ 만 15세	497	8.2%
만 16세 ~ 만 18세	1,058	17.5%
만 19세 ~ 만 29세	2,720	45.0%
만 30세 ~ 만 59세	648	10.7%
만 60세 이상	321	5.3%
전체	6,044	100.0%

2.1.6.2 미성년자 법정 대리인 동의서 및 증명서 제출

만 14세 ~ 만 18세는 법정 대리인이 본 ‘한 편 줍쇼’ 행사에 미성년자인 권리자가 참여하는 것에 동의하는 과정을 거쳤다. 법정 대리인과 미성년자의 관계를 증명하는 가족 관계 증명서를 필수로 제출하였다. 가족 관계 증명서는 법정 대리인 확인 용도로만 사용하며 최종 확인 후 폐기 처리한다. 만 14세 ~ 만 18세의 미성년자가 법정 대리인 동의서와 증빙 서류를 제출하지 않은 경우에는 제출한 글의 파일을 폐기하고, 참여 완료 시 지급하는 문화 상품권도 지급하지 않았다. 따라서 3월 12일까지 글을 제출한 만 14세 ~ 만 18세의 미성년 참여자들에게 조사원이 연락하여 필수 서류 제출을 독려하였다. 해당 연령대 참여자 1,200명 중 200명가량이 법정 대리인 동의서 등을 제출하지 않아 참여에서 배제되었다. 행사 종료 후 법정 대리인 동의서 미제출로 인한 수집

부족분은 기존 참여자에게 소개를 받아 추가로 수집하였다.

법정대리인을 맡고 있는 권리자 _____가 제출한 '국가 언어
자원(말뭉치) 구축 및 활용 저작권 이용허락 계약서'에 법정대리인인
_____은 본 계약 내용에 동의합니다.

법정대리인 : _____ (인)
성명 : _____
주민등록번호 :

--	--	--	--	--	--	--	--	--	--

 * * * * *

<그림 13> 법정 대리인 동의서

표절 검사 프로그램인 카피킬러로 접수된 글의 표절 여부를 검사하였다. 접수 순서대로 카피킬러 프로그램으로 검사하였고, 표절률에 따라 표절 검수자가 글을 확인하는 절차를 거쳤다.

〈표 10〉 표절 검사 결과

검사	제출일	표절 검사 인원	표절 인원	표절률(%)
1차	2월3일~4일	50	5	10.0
2차	2월5일	60	4	6.6
3-1차	2월6일	1,231	44	3.6
3-2차	2월6일	1,250	53	4.2
3-3차	2월6일	581	39	6.7
4차	2월7일~14일	334	12	3.6
5차	2월15일	255	7	2.7
6차	2월16일~29일	1,326	58	4.4
7차	3월1일~11일	1,109	24	2.2
전체		6,196	246	4.9

표절 검사 결과, 독후감을 작성하며 글의 일부를 인용한 경우가 가장 많은 표절 사례로 나타났다. 이러한 경우에 책의 특정 부분을 똑같이 쓸 수밖에 없으나 인용 표시를 제대로 하지 않으면 표절로 간주하여 제출 글로 인정하지 않았다.

2.1.8 원시 말뭉치 정제

2.1.8.1 원문 자료 입력(초등학생 및 오프라인 제출자 글 대상)

초등학생은 보호자가 글을 대신 입력하며 원문과 달리 맞춤법, 띄어쓰기, 표현 등이 교정할 가능성 있어 손 글씨 원문을 제출하게 하였다. 참여 시 누리집에 입력한 글은 글자 수 확인용으로만 쓰였으며, 원시 말뭉치는 전문 요원이 손 글씨 원문을 기준으로 입력하였다. 초등학생뿐만 아니라 오프라인 추가 수집 시에는 연령에 관계없이 원문 파일(손 글씨 또는 워드프로세서 작성 출력 파일)을 받았다. 총 1,059명의 참여자가 원문을 제출하였으며, 제출된 원문은 3,499편, 1,566,380개 글자 분량이었다. 입력 기간은 3주 소요되었다.

1차 수집 후 정제 작업을 하였으며, 제출한 원문이 흐릿하거나 글씨가 잘려 입력하기에 부적합하면 제출자에게 원문을 다시 전송받았다. 연락이 되지 않거나 재제출을 거부한 경우에는 수집 분량에서 배제하였다. 재요청 및 재전송, 원문 재입력에는 약 3주가 소요되었다.

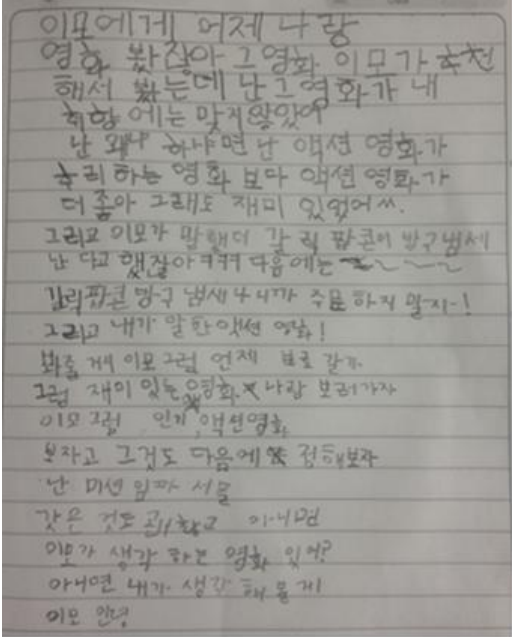
〈표 11〉 손글씨 제출자 연령별 구성

연령 구분(만 나이)	제출자 수(명)
만 7세 ~ 만 12세	778
만 13세 ~ 만 15세	62
만 16세 ~ 만 18세	131
만 30세 ~ 만 59세	6
만 60세 이상	82
전체	1,059

2.1.8.2 원문 입력 규칙

원문 입력 시에 적용한 규칙은 다음과 같다. 맞춤법, 띄어쓰기를 교정하지 않고 원문과 동일하게 입력했으며, 글씨가 불분명한 부분 기호화, 개인 정보 비식별화와 같은 마크업(markup) 처리를 했다. 다음과 같이 진행하였다.

- 원문의 본문이 잘려서 글씨가 안 보이는 경우 해당 글은 폐기함
- 본인의 손으로만 쓴 글을 입력 대상으로 간주(손 글씨 이외의 글은 옮기지 않음. 예를 들어 일기장의 “날씨”와 같은 틀은 입력 대상 아님)
- 원문 그대로 옮김: 띄어쓰기, 맞춤법 오류 등을 교정하지 않고 그대로 옮김
- 글씨 식별이 불확실한 부분은 <unclear></unclear> 태깅
- 고쳐 쓰기 관련: 취소선, 엑스 등으로 글쓴이가 지운 부분은 입력하지 않고, 취소하지 않고 문장을 고쳐 쓴 경우에는 <trunc></trunc> 태깅
- 단락 태깅: <p></p> 태깅. 줄의 오른쪽 끝에 여백이 없으면 단락이 이어지는 것으로 간주함. 이에 반해 줄의 오른쪽에 후행 어절을 적을 여백이 있으면 글쓴이가 의도적으로 문단을 분리한 것으로 간주하여 새로운 단락으로 처리함.

	<p><p>이모에게 어제 나랑</p> <p>영화 봤잖아 그 영화 이모가 추천해서 봤는데 난 그 영화가 내 취향에는 맞지않았어</p> <p>난 왜냐 하나면 난 액션 영화가 좋아하는 영화 보다 액션 영화가 더 좋아 그래도 재미 있었어^^.</p> <p>그리고 이모가 말했더 갈릭 팝콘이 방구냄새 난 다고 했잖아 ㅋㅋ 다음에는 <unclear>~~~~</unclear> 갈릭팝콘 방구 냄새나니까 주문하지 말자!</p> <p>그리고 내가 말한액션 영화!</p> <p>봐줄 거 이모 그럼 언제 보러 갈까</p> <p>그럼 재미있는 영화 나랑 보러가자</p> <p>이모 그럼 인기 액션영화</p> <p>보자고 그것도 다음에 꼭 정해보자</p> <p>난 미션 임파서블</p> <p>같은 것도 괜찮고 아니면</p> <p>이모가 생각 하는 영화 있어?</p> <p>아니면 내가 생각 해 볼 게</p> <p>이모 안녕</p></p>
---	--

〈그림 15〉 원문 입력 예시: 수원동신초등학교 신재혁

○ 개인 정보 비식별화

이름: &name1&, &name2&, ...

주민 등록 번호: <anon type="social-security-num" />

전화번호: <anon type="tel-num" />

2.1.8.3 자유 글감 ‘글감’ 분석

글쓴이가 글 제출 시 ‘자유 글감’을 선택한 경우에는 사업 수행자가 고용한 전문 요원이 글을 읽고 글감 부분(subclass)에 “null_글감” 형태로 입력하였다. 따라서 글쓴이가 입력한 글감은 “글감”으로 입력되고, 전문 요원이 입력한 글감은 “null_글감”으로 입력되어 있다.

온라인으로 수집한 약 7,700여 편과 초등학생 및 오프라인 수집(사이트 외 수집) 대상자들의 약 3,300여 편 모두 글감 선정 작업을 진행하였다. 글감 작업 1차에서는 맞춤법을 글쓴이가 틀렸을 경우에는 수정하지 않고 글감 작업을 진행했으나, 2차 정제 작업 시 맞춤법과 띄어쓰기를 교정하였다.

 글 정보

NO	글감	작성연도	글종류	제목
1	자유주제	2020	수필	달리는 150 버스 오른쪽 세번째 노약자석

→ 자유 주제 또는 자유 글감은 SJML에 “null_글감”으로 입력

<그림 16> 전문 요원 글감 입력 대상 글 예시

2.1.8.4 XML 문서 구조와 형식

원시 말뭉치 파일의 확장자는 SJML이며, 문자 인코딩은 유니코드(UTF-8)이다. 기본 구조와 형식은 다음과 같다.

```
<?xml version="1.0" encoding="UTF-8"?>
<SJML>
  <header>
    <fileInfo>
      <fileId>WDRW1900100000</fileId>
      <annoLevel>원시</annoLevel>
      <category>비출판물 > 글 종류</category>
    </fileInfo>
    <sourceInfo>
      <title>제목</title>
      <author id="P00000" age="나이" occupation="직업" sex="성별" submission="
제출 방법" handwriting="손글씨 원문 유무">저작자 이름</author>
    </sourceInfo>
  </header>
  <text date="20200000" subclass="(null_)"글감">
    <p>.....</p>
  </text>
</SJML>
```

〈그림 17〉 XML 기본 구조

〈표 12〉 메타 정보 구조

<fileInfo>	<fileId>	파일의 고유 식별자(말뭉치 파일명)
	<annoLevel>	주석 수준: 원시
	<category>	비출판물 > 글 종류
<sourceInfo>	<title>	글의 제목
	<author>	글 작성자(저작권자) 정보: ID, 나이, 직업, 성별, 제출 방법, 손 글씨 원문 여부, 저작자 이름(글쓴이가 비공개를 원하는 경우에는 개인글 작성자로 기재됨)
<text>	<text date>	글 작성 날짜(글 제출일이 아님)
	<subclass>	글감

2.1.8.5 말뭉치 파일명 부여 형식

개인적 글쓰기 자료는 문어 말뭉치 중 비출판물 말뭉치에 속하며 말뭉치 파일명 부여는 국립국어원 말뭉치 파일명 부여 기준에 따랐다. 말뭉치 유형 구분, 매체 및 장르 분류, 분석 층위 구분, 구축년도, 8자리 일련번호로 구성되어 있다. WDRW190XXYYYYYY가 기본 형식으로 뒤의 5자리(YYYYYY)는 글 작성자 아이디이며, XX는 한 작성자가 제출한 글의 수를 나타내는 것으로 지정하였다.

〈표 13〉 XML 정제 사항

1	2	3	4	5	6	7	~	14
말뭉치 유형 구분	매체 및 장르 분류	분석 층위 구분	구축년도	8자리 일련번호				
W: 문어 말뭉치	D: 비출판물	OR: 원문 자료 RW: 원시	19	#####				

```

<?xml version="1.0" encoding="UTF-8"?>
<SJML>
  <header>
    <fileInfo>
      <fileId>WDRW1900520500</fileId>
      <annoLevel>원시</annoLevel>
      <category>비출판물 > 편지글</category>
    </fileInfo>
    <sourceInfo>
      <title>제목없음</title>
      <author id="P20500" age="9" occupation="초등학생" sex="M" submission="오프라인"
handwriting="Yes">신재혁</author>
    </sourceInfo>
  </header>
  <text date="20190000" subclass="영화">
    <p>이모에게 어제 나랑</p>
    <p>영화 봤잖아 그 영화 이모가 추천해서 봤는데 난 그 영화가 내 취향에는 맞지않았어</p>
    <p>난 왜냐 하나면 난 액션 영화가 추리하는 영화 보다 액션 영화가</p>
    <p>더 좋아 그래도 재미 있었어~.</p>
    <p>그리고 이모가 말했더 갈릭 팝콘이 방구냄새 난 다고 했잖아 ㅋㅋㅋ 다음에는
<unclear>~~~~</unclear> 갈릭팝콘방구 냄새나니까 주문하지 말자!</p>
    <p>그리고 내가 말한액션 영화!</p>
    <p>봐줄 거지 이모 그럼 언제 보로 갈까</p>
    <p>그럼 재미있는 영화 나랑 보러가자</p>
    <p>이모 그럼 인기 ,액션영화</p>
    <p>보자고 그것도 다음에 정해보자</p>
    <p>난 미션 임파 서블</p>
    <p>жат은 것도 괜찮고 아니면</p>
    <p>이모가 생각 하는 영화 있어?</p>
    <p>아니면 내가 생각 해 볼 게</p>
    <p>이모 안녕</p>
  </text>
</SJML>

```

〈그림 18〉 비출판물 말뭉치 XML 예시

2.2 말뭉치 지식 강연회 개최

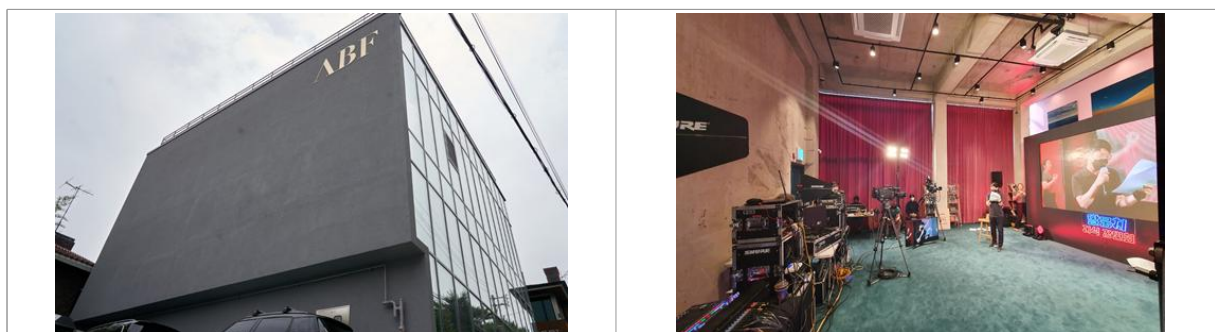
2.2.1 말뭉치 지식 강연회 행사 개요

국어 정보화에 대한 국민의 관심을 높이고 우리말 자원을 활용한 연구 및 산업 환경 조성에 기여하고자 2020년 6월 18일(목) ‘말뭉치 지식 강연회’를 개최하였다. 2020년 2월 대면 강연회로 개최할 계획이었으나, 코로나바이러스감염증-19로 인해 한 차례 연기해 온라인 방식으로 개최하였다.



<그림 19> 말뭉치 지식 강연회 광고지

온라인 생중계는 6월 18일(목) 14시 20분부터 18시 15분까지 강남구에 위치한 ‘ABF 스튜디오’에서 진행했으며, 온라인 강연회 플랫폼 ‘콩콩(CongKong)’을 통해 송출했다.



<그림 20> 말뚝치 지식 강연회 행사 장소

<표 14> 강연회 진행 일정 개요 시간

구분	시간		내용
입장	14:20~14:30	10분	참가자 입장 대기
공식 행사	14:30~14:35	5분	개회
	14:35~14:40	5분	축사 / 개회사
강연	14:40~15:25	45분	민규동 감독, 주성철 평론가 영화 속 사례
	15:25~15:35	10분	휴식 시간(전시 영상 송출) ‘성.시.경’ 팀
	15:35~15:40	05분	휴식 시간(문제 풀이)
	15:40~16:25	45분	남세동 대표 - 자연어 처리 1
	16:25~17:10	45분	이경일 대표 - 자연어 처리 2
	17:10~17:20	10분	휴식 시간(전시 영상 송출) ‘에이아이인(이하 AI人)’ 팀
	17:20~17:25	5분	휴식 시간(문제 풀이)
	17:25~18:10	45분	송길영 부사장 - 4차 산업
폐회	18:10~18:15	5분	경품 추첨(문제 풀이 2건)
	18:15~	-	참가자 기념품 안내

〈표 15〉 강연회 진행 세부 일정

구분	시간		내용
설치 / 시스템 사전 연습	6월 17일(수) 13:00~18:00	-	지식 강연회 촬영장 구성 - 시스템: 음향, 영상, 중계, 조명 - 제작물: 현수막, 엑스배너, 연사 대기실, 강연회 제목 네온등 시스템 사전 연습 - 영상 및 발표 자료, 음향 시스템 테스트 등
전체 사전 연습	6월 17일(목) 10:00~10:30	30분	제작진 회의 및 시스템 사전 연습 - 음향, 영상, 중계, 조명
	10:30~11:00	30분	1차 사전 연습 - 공식 행사 사전 연습 - 지식 강연회 강연(4개) 사전 연습 진행 - 발표 자료 및 송출 화면 확인
	11:00~11:30	30분	2차 사전 연습 - 공식 행사 사전 연습 - 지식 강연회 강연(4개) 실제 식순과 동일한 순서로 사전 연습 진행
	13:30~14:00	30분	최종 사전 연습 - 시스템(음향, 영상, 중계, 조명) 포함 영상, 발표 자료 확인 사회자 대본 연습 - 아나운서 현장 미팅 및 대본 연습 진행
	14:00~14:30	20분	모든 제작진 정위치 및 행사 진행 - 각 제작진 위치별 정위치 - 강연 참여자 온라인 플랫폼 입장 시작 - 행사 시작 관련 안내문 송출
사전 고지	14:20~14:30	10분	행사 안내 - 문소리 아나운서
개회 선언	14:30~14:35	5분	개회 선언 - 문소리 아나운서
축사 / 개회사	14:35~14:40	5분	내빈 개회사 및 축사 영상 소개 - 축사: 문화체육관광부 제1차관 오영우 - 개회사: 국립국어원장 소강춘
강연	14:40~15:25	45분	강연1_말대꾸를 시작한 인공 지능 민규동 감독, 주성철 평론가
휴식 시간	15:25~15:35	10분	영상_인공 지능 이야기 창작 보조 기술 ‘성.시.경’ 팀의 제작 영상(3분) 송출
	15:35~15:40	5분	문제 풀이 - 민규동 감독, 주성철 평론가 강연 관련 문제

구분	시간		내용
강연	15:40~16:25	45분	강연2_인공 지능에 말뚱치를 더하다 남세동 보이저엑스 대표
	16:25~17:10	45분	강연3_인공 지능을 완성시키는 언어 빅 데이터 이경일 솔트룩스 대표
휴식 시간	17:10~17:20	10분	영상_‘AI人’ 인공 지능 기반 한국어 시 생성 기술 소개(4분 5초) 송출
	17:20~17:25	05분	문제 풀이 - 남세동, 이경일 대표 강연 관련 자연어 처리 분야 문제
강연	17:25~18:10	45분	강연4_인공 지능에 말뚱치를 더하다 송길영 다음소프트 부사장
폐회	18:10~18:15	5분	경품 추첨 문제 풀이 - 말뚱치 관련 문제(2문제)
	18:15~	-	폐회 참가자 기념품 안내

2.2.2 사전 홍보

말뭉치 지식 강연회를 성공적으로 개최하기 위해 온·오프라인 홍보를 진행했다. 주로 유관 기관 종사자 및 관련 기업 관계자를 대상으로 홍보했으며, 관련 학과 학생 및 유관 분야 취업 희망자 위주로 강연회 참가를 유도했다.

2.2.2.1 오프라인 홍보

국민 접점 홍보를 위해 공공 기관 보유 매체 및 사업 수행자 보유 네트워크를 활용하여, 강연회 홍보 영상을 오프라인 매체에서 1월 20일(월)부터 2월 7일(금)까지 약 3주간 비예산으로 송출했다.

〈표 16〉 비예산 홍보_공공 기관 보유 매체 송출

지역		보유 매체	지역		보유 매체
경기도	광주시청	아이피 티브이, 버스 정보 시스템 206대	세종특별자치시		버스 정보 시스템
	광명시청	버스 정보 시스템, 아이피 티브이 25곳, 전광판 2곳	충청남도	아산시청	전광판, 미디어보드
	안양시청	인터넷 방송국, 버스 정보 시스템, 미디어보드		천안시청	버스 정보 시스템
	양주시청	전광판, 전철역, 마트, 아이피 티브이 29곳	충청북도	청주시청	버스 정보 시스템
	연천군청	누리 소통망	강원도	원주시청	버스 정보 시스템
	수원시청	버스 정보 시스템	경상북도	구미시청	전광판
	양평군청	버스 정보 시스템 41개소	경상남도	밀양시청	버스 정보 시스템
	안산시청	버스 정보 시스템		창원시청	버스 정보 시스템
	용인시청	버스 정보 시스템		김해시청	버스 정보 시스템
	인천시청	버스 정보 시스템	서울시청		전광판
	동두천시청	버스 정보 시스템	인천시청		전광판
	평택시청	버스 정보 시스템, 블로그	울산시청		버스 정보 시스템
	고양시청	버스 정보 시스템			
	시흥시청	창업지원센터			



〈그림 21〉 버스 정보 시스템 홍보 모니터링 캡처(평택시청 소통홍보관실 제공)

2.2.2.2 온라인 홍보(카드 뉴스)

말뭉치에 대한 정보 제공 및 말뭉치 지식 강연회 소개를 위해 누리 소통망을 활용했다. 말뭉치 관련 핵심 내용을 압축하여 ‘말뭉치’와 같은 생소한 용어를 효과적으로 노출하고 행사를 알릴 수 있도록 홍보물을 제작했다.

1차 카드 뉴스는 1월 22일(수) 국립국어원 공식 페이스북에 공개했다. “알쏭싹잡” 프로그램 형식을 활용하여 말뭉치 활용 방법과 말뭉치 지식 강연회를 소개했다. 좋아요 13개, 공유 1회를 기록했다(2020.07.13. 기준).



〈그림 22〉 1차 카드 뉴스

2차 카드 뉴스는 1월 30일(목) 국립국어원 공식 페이스북에 공개했다. 말뭉치를 활용한 기술을 소개하고, 개인적 글쓰기 자료 수집 행사인 “한 편 줍쇼” 정보를 알렸다. 좋아요 20개, 공유 8회를 기록했다(2020.07.13. 기준).



<그림 23> 2차 카드 뉴스

3차 카드 뉴스는 6월 12일(금) 국립국어원 공식 페이스북에 공개했다. 2월에 연기되었던 말뚝치 지식 강연회가 다시 돌아왔다는 내용으로 구성했다. 강연회 참석 유도를 위해 사전 등록이 가능한 온라인 링크를 본문에 포함했다. 좋아요 27개, 댓글 5개, 공유 8회를 기록했다(2020.07.13. 기준).



<그림 24> 3차 카드 뉴스

2.2.2.3 온라인 홍보(유관 기관 대상)

말뭉치 지식 강연회 참여 유도를 위해 언어, 인공지능, 빅 데이터 등 관련 기업 및 대학교에 연락을 취해 온라인 홍보 협조를 요청했다. 1차로 1월, 2차로 6월에 진행하였다. 특히 학생들을 대상으로는 대학교 및 학과 누리집에서 강연회 정보를 확인할 수 있도록 했다.

〈표 17〉 온라인 홍보_관련 업체 연락 목록

번호	구분	기업명	번호	구분	기업명
1	인공 지능	주식회사 솔루게이트	21	빅 데이터	텐스페이스
2		데이타플로	22		주식회사 스위트케이
3		주식회사 아스크스로리	23		주식회사 커넥트닷
4		아임클라우드	24	인공 지능	심심이(주)
5		주식회사 인퍼닉	25		(주) 넷솔루션
6		(주)뉴로컴즈	26		노마드커넥션
7		다크트레이 코리아	27		모비코
8		머니브레인	28		아르군
9		주식회사 딥픽셀	29		주식회사 마키나락스
10		주식회사 레인보우브레인	30		(주)미디어그룹사람과숲
11		주식회사 마인즈랩	31		셀렉트스타
12		(주) 케이로봇	32		주식회사 딥네츄럴
13		주식회사 알체라	33		(주)데브언리밋
14		(주) 와이즈스톤	34		이스트소프트
15	빅 데이터	deep visions	35		테스트웍스
16		(주) 두잉랩	36		주식회사 키니엔파트너스
17		딥서치	37		엠아이큐브솔루션
18		주식회사 스타핑	38		아트케이오스모스
19		메가존(주)	39		(주)아크릴
20		몬드리안에이아이주식회사	-		-

〈표 18〉 온라인 홍보_대학 연락 목록

번호	구분	지역	학교명
1	국어국문학과	서울특별시	건국대학교(서울캠퍼스)
2		서울특별시	경희대학교(본교-서울캠퍼스)
3		서울특별시	고려대학교(본교)
4		서울특별시	광운대학교(본교)
5		서울특별시	국민대학교(본교)
6		서울특별시	덕성여자대학교(본교)
7		서울특별시	동국대학교(서울캠퍼스)
8		서울특별시	동덕여자대학교(본교)
9		서울특별시	명지대학교(인문캠퍼스)
10		서울특별시	삼육대학교(본교)
11		인천광역시	인천대학교(본교)
12		경기도	가톨릭대학교(본교)
13		경기도	강남대학교(본교)
14		경기도	경기대학교(본교)
15		경기도	단국대학교(죽전캠퍼스)
16	문예창작과	서울특별시	명지대학교(인문캠퍼스)
17		서울특별시	서울과학기술대학교(본교)
18		서울특별시	서울디지털대학교
19		서울특별시	서울디지털대학교
20		서울특별시	서울사이버대학교
21		서울특별시	송실대학교(본교)
22		서울특별시	송실사이버대학교
23		서울특별시	추계예술대학교(본교)
24		서울특별시	한국예술종합학교
25		경기도	경기대학교(본교)
26		경기도	대진대학교(본교)
27		경기도	중앙대학교(안성캠퍼스)
28	문헌정보학과	서울특별시	상명대학교(서울캠퍼스)
29		서울특별시	상명대학교(서울캠퍼스)
30		서울특별시	서울여자대학교(본교)
31		서울특별시	성균관대학교(본교)
32		서울특별시	숙명여자대학교(본교)
33		서울특별시	연세대학교(신촌캠퍼스)
34		서울특별시	이화여자대학교(본교)
35		서울특별시	이화여자대학교(본교)
36		서울특별시	중앙대학교(서울캠퍼스)
37		서울특별시	한성대학교(본교)
38	문화콘텐츠학과	서울특별시	광운대학교(본교)
39		서울특별시	글로벌사이버대학교
40		서울특별시	상명대학교(서울캠퍼스)
41		서울특별시	상명대학교(서울캠퍼스)
42		서울특별시	서강대학교(본교)
43		서울특별시	서울과학기술대학교(본교)
44		서울특별시	서울디지털대학교
45		서울특별시	서울사이버대학교
46		서울특별시	성신여자대학교(본교)
47		서울특별시	성신여자대학교(본교)

번호	구분	지역	학교명
48	한국어학과	서울특별시	경희대학교(본교-서울캠퍼스)
49		서울특별시	경희사이버대학교
50		서울특별시	경희사이버대학교
51		서울특별시	고려사이버대학교
52		서울특별시	국민대학교(본교)
53		서울특별시	디지털서울문화예술대학교
54		서울특별시	사이버한국외국어대학교
55		서울특별시	서강대학교(본교)
56		서울특별시	서울사이버대학교
57		서울특별시	세종사이버대학교
58	응용학과	서울특별시	건국대학교(서울캠퍼스)
59		서울특별시	경희사이버대학교
60		서울특별시	고려사이버대학교
61		서울특별시	광운대학교(본교)
62		서울특별시	상명대학교(서울캠퍼스)
63		서울특별시	서울여자대학교(본교)
64		서울특별시	한국방송통신대학교
65		인천광역시	인하대학교(본교)
66		경기도	가천대학교(글로벌캠퍼스)
67		경기도	대진대학교(본교)
68	미디어학과	서울특별시	건국대학교(서울캠퍼스)
69		서울특별시	건국대학교(서울캠퍼스)
70		서울특별시	경기대학교(서울캠퍼스)
71		서울특별시	경희사이버대학교
72		서울특별시	고려대학교(본교)
73		서울특별시	국민대학교(본교)
74		서울특별시	덕성여자대학교(본교)
75		서울특별시	동국대학교(서울캠퍼스)
76		서울특별시	서강대학교(본교)
77		서울특별시	서강대학교(본교)
78	정보통신학과	서울특별시	건국대학교(서울캠퍼스)
79		서울특별시	경희사이버대학교
80		서울특별시	고려대학교(본교)
81		서울특별시	광운대학교(본교)
82		서울특별시	광운대학교(본교)
83		서울특별시	광운대학교(본교)
84		서울특별시	동국대학교(서울캠퍼스)
85		서울특별시	서울과학기술대학교(본교)
86		서울특별시	성공회대학교(본교)
87		서울특별시	성공회대학교(본교)

번호	구분	지역	학교명
87	응용소프트웨어공학과	서울특별시	경희대학교(본교-서울캠퍼스)
88		서울특별시	경희사이버대학교
89		서울특별시	광운대학교(본교)
90		서울특별시	광운대학교(본교)
91		서울특별시	국민대학교(본교)
92		서울특별시	명지대학교(인문캠퍼스)
93		서울특별시	서강대학교(본교)
94		서울특별시	서강대학교(본교)
95		서울특별시	서울사이버대학교
96		서울특별시	서울여자대학교(본교)
97	디지털콘텐츠학과	서울특별시	경희대학교(본교-서울캠퍼스)
98		서울특별시	명지대학교(인문캠퍼스)
99		서울특별시	성공회대학교(본교)
100		서울특별시	성균관대학교(본교)
101		서울특별시	세종대학교(본교)
102		서울특별시	한국외국어대학교(본교)
103		서울특별시	한국외국어대학교(본교)
104		경기도	신경대학교(본교)
105		경기도	한신대학교(본교)



<그림 25> 온라인 홍보_유관 기관 연락 결과

2.2.2.4 온라인 홍보(커뮤니티 배너 광고)

온라인 포털 대표 취업 커뮤니티인 ‘스펙업(회원 수 1,939,425명)’에서 대학·대학원생 대상으로 광고를 집행했다. 첫 화면에 배너 광고를 걸어 20대~30대 대학·대학원생을 겨냥해 홍보를 진행했다. 또한 스펙업 페이스북(팔로워 395,484명)에 홍보물을 게시하고 스펙업 카카오톡 채널 팔로워를 대상으로 카카오톡 메시지를 발송하여 다수의 인원에게 한 번에 홍보할 수 있었다.

네이버 커뮤니티 배너 광고는 2020년 1월 20일(월)부터 2월 4일(화)까지 진행했으며, 1월 20일(월) 페이스북 담벼락에도 게시했다. 카카오톡 메시지는 1월 29일(수) 일괄 전송했다.



〈그림 26〉스펙업 배너 광고, 페이스북 게시물, 카카오톡 메시지 발송

2.2.2.5 언론 홍보

말뚱치 지식 강연회 개최 보도자료 초안 작성을 지원하여 원활한 언론 홍보를 보조했다.

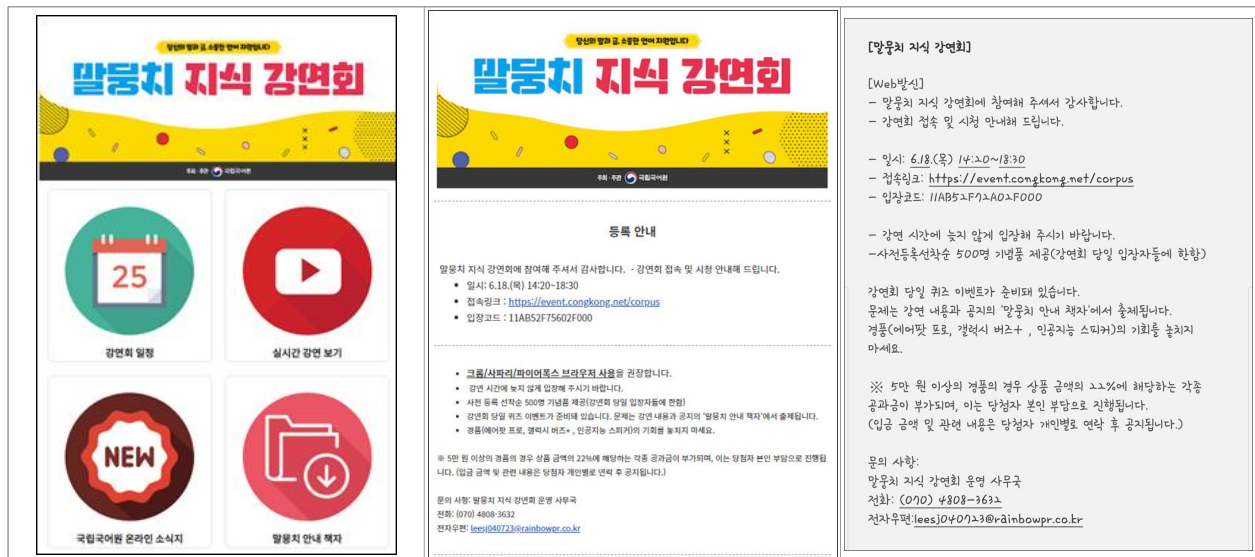


〈그림 27〉1차, 2차 보도자료

2.2.3 행사 운영

2.2.3.1 온라인 강연회 사전 접수 신청

본 행사는 최종적으로 온라인상에서 개최했으며, 강연 자료 유출을 막기 위해 특정 외부 송출 플랫폼, 콩콩(CongKong)을 활용했다. 이에 신청자 개개인에게 개별 코드를 발송하여 입장할 수 있도록 했고, 2월 오프라인 강연 사전 접수자를 대상으로 먼저 참여 신청을 받았다.



〈그림 28〉 누리집 사전 접수 화면(왼쪽), 개별 입장 코드 발송(전자 우편, 문자)

2.2.3.2 사전 준비

말뚱치 지식 강연회 참가자들을 위해 기념품을 총 530묶음 준비했다. 말뚱치 책자, 휴대용 선풍기, 카드형 유에스비(32GB)로 구성했다.

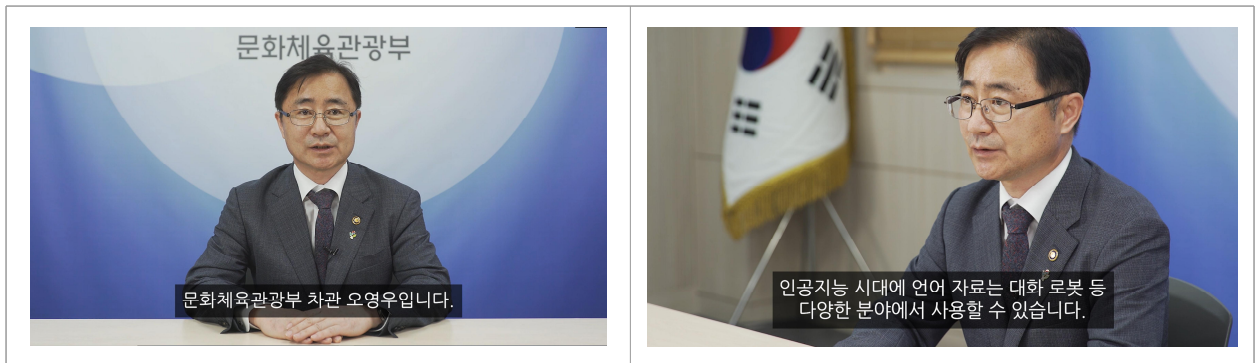
또한 온라인 강연 중간 이탈 방지를 위해 휴식 시간에는 영상 송출과 함께 문제 풀이 행사를 준비했다. 경품은 총 4개로 인공 지능 스피커(2개), 갤럭시 버즈 플러스(1개), 에어 팟 프로(1개)로 구성했다.

강연회 후기 작성 행사를 위한 경품으로는 갤럭시 워치 액티브2(1개)를 준비했다.



〈그림 29〉 기념품 구성(왼쪽), 경품 구성(오른쪽)

강연회를 온라인으로 진행하며, 축하 및 개회사 또한 영상으로 대체했다. 오영우 문화체육관광부 제1차관이 축하를 맡았고, 소강춘 국립국어원장이 개회사를 하였다. 축하와 개회사 녹화 영상은 강연회 당일 온라인 플랫폼에서 송출했다.



〈그림 30〉 문화체육관광부 제1차관 오영우 축하



〈그림 31〉 국립국어원장 소강춘 개회사

오프라인 강연회에서 전시할 예정이었던 한국콘텐츠진흥원의 인재 양성 교육 프로그램 ‘콘텐츠임팩트 창의랩’ 교육생(‘성.시.경’ 팀, ‘AI人’ 팀)의 작품과 천계영 작가가 자료는 온라인 강연회에 영상을 송출하는 것으로 대체했다. 강연회 휴식 시간에 영상을 송출하여 다채로운 행사가 되도록 구성하였다. ‘성.시.경’ 팀은 인공 지능을 활용하여 시나리오를 작성하고 제작한 웹 드라마의 제작 영상을 송출했으며, ‘AI人’ 팀은 시를 제작하는 프로그램을 설명하는 영상을 내보냈다. 천계영 작가의 영상은 음성 인식으로 만화 작업을 하는 내용을 다룬 것이다.

〈그림 32〉 ‘성.시.경’ 팀의 인공 지능 이야기 창작 보조 프로그램(웹 드라마 ‘블라인드’)

〈그림 33〉 ‘AI人’ 인공 지능 기반 한국어 시 생성 프로그램

본 행사는 강남구 소재 에이비에프(ABF) 스튜디오에서 촬영 및 송출을 진행했다. 강연회 전날인 6월 17일(화)에는 무대 및 물자를 현장에 준비하고, 이동 동선을 확인했다.



〈그림 34〉 행사장 동선 등

6월 17일(화)에 사전 연습을 진행하며, 현장 담당자들과 함께 원활한 행사 운영을 위해 업무 분장을 진행했다.

〈표 19〉 강연회 행사 당일 업무 구분

구분	업무내용	인원수	비고
총괄 책임	<ul style="list-style-type: none"> 행사 운영 총괄 사회자 및 진행 요원 관리 	1	레인보우
진행 팀장	<ul style="list-style-type: none"> 행사 운영 담당 행사 시스템 및 동선 안내 진행 요원 업무 관리 	1	레인보우
현장 운영	<ul style="list-style-type: none"> 촬영장 장소 관리 	3	레인보우
진행 요원	<ul style="list-style-type: none"> 주차, 강연자 의전 및 동선 안내 체온 검사 	4	진행 요원

강연회 당일엔 코로나19 대응을 위해 현장 입구에서 반드시 손 소독제를 사용하게 했고, 마스크 착용과 체온을 확인하였다. 현장에서는 강연자를 제외한 모든 사람들은 마스크를 필수로 착용했다.



〈그림 35〉 현장 운영

2.2.3.3 행사 운영

행사는 모두 네 개의 강연으로 구성되었다. 첫 번째 강연 ‘말대꾸를 시작한 인공지능’은 민규동 감독과 주성철 평론가가 대담 형식으로 진행하며 영화 속에 나타난 말뭉치 관련 인공지능 기술을 이야기했다. 민규동 감독은 ‘내 아내의 모든 것’, ‘여고괴담 두 번째 이야기’ 등을 연출하였으며, 주성철 평론가는 전 씨네21 편집장이다.



〈그림 36〉 민규동 감독, 주성철 평론가

두 번째 강연 ‘인공 지능에 말뚱치를 더하다’에서는 남세동 보이저엑스 대표가 인공 지능 시대를 앞당긴 주요 기술과 기계가 언어를 처리하는 전반적인 내용을 소개했다. 남세동 대표는 세이클럽(sayclub) 1인 개발자로, 비612(B612), 라인(LINE) 등에 근무하며 컴퓨터 프로그래밍에 대한 경험이 다양한 인물이다.



<그림 37> 남세동 대표

이어서 이경일 솔트룩스 대표가 ‘인공 지능을 완성시키는 언어 빅 데이터’ 라는 주제로 인공 지능과 빅 데이터 기술을 통한 자연어 학습과 말뭉치 쓰임에 대해 강연하였다.



<그림 38> 이경일 대표

마지막으로 송길영 다음소프트 부사장이 ‘나! 너... 우리?’라는 제목으로 다양한 언어 빅 데이터를 활용, 분석하여 얻은 우리 삶의 모습과 변화 양상에 대하여 이야기 했다. 특히 ‘코로나’라는 핵심어로 자료를 분석한 결과를 공유하며 코로나로 인한 사회의 변화와 앞으로의 발전 방향을 설파하였다.



〈그림 39〉 송길영 부사장

강연 중간에는 청중의 참여를 유도하기 위해 문제 풀이 경품 행사를 실시했다. 1번과 2번 문제 정답자에게는 인공 지능 스피커(카카오 미니)를 주었고, 3번 문제는 갤럭시 버즈 플러스, 4번 문제 경품으로는 애플 에어팟 프로를 제공했다.

현장 송출 오류로 인해 실시간 댓글 작성이 불가하여, 정답은 사후 페이스북 게시물 댓글과 전자 우편으로 받았다. 중복 참여자를 거르고 각 문제당 한 명씩 당첨자를 추첨했다. 1번 문제 275명, 2번 문제 250명, 3번 문제 271명, 4번 문제 267명이 참여했으며, 1번 문제는 박종해 님, 2번 문제는 최형석 님, 3번 문제는 박분선 님, 4번 문제는 김도현 님이 당첨되었다.

<p>1번 문제</p> <p>스스로 판단하고 감정도 가지며 인간을 위협하기도 하는 컴퓨터, HAL 9000이 등장하는 스탠리 큐브릭 감독의 영화 제목은 무엇일까요?</p>	<p>정답</p> <p>2001: 스페이스 오디세이</p>
<p>2번 문제</p> <p>여러 언어 간의 비교를 위해 특별히 형태화된 다중언어 말뭉치를 무엇이라고 할까요?</p>	<p>정답</p> <p>병렬 말뭉치</p>
<p>3번 문제</p> <p>언어 자료에 컴퓨터가 읽을 수 있도록 자료의 종류나 제목, 작성자, 문단이나 문장 경계와 같은 '메타 정보'를 입력해 만든 기본 말뭉치는 무엇일까요?</p>	<p>정답</p> <p>원시 말뭉치</p>
<p>4번 문제</p> <p>하나의 글 안에서 같은 대상에 대해 서로 다르게 표현한 말들을 찾아 연결한 말뭉치는 무엇일까요?</p>	<p>정답</p> <p>상호참조 해결 말뭉치</p>

<그림 40> 문제 풀이 행사 내용

2.2.4 사후 관리: 후기 작성 행사, 강연 영상 제작, 설문 조사

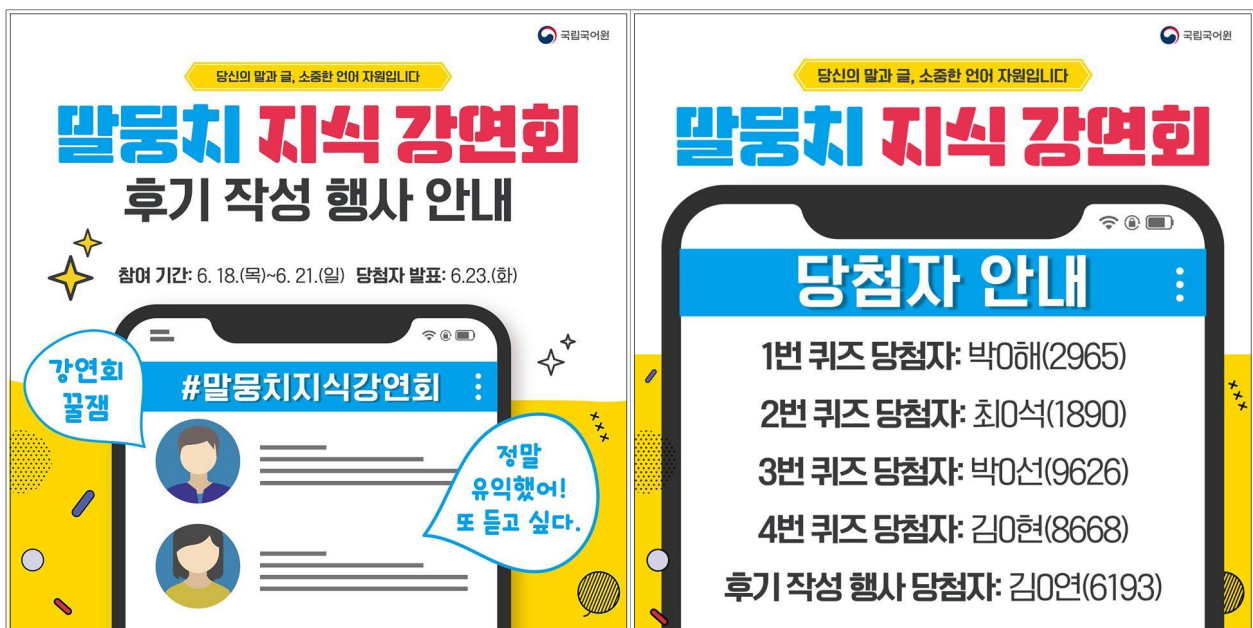
말뚱치 지식 강연회 종료 후 참가자를 대상으로 온라인 후기를 남기는 홍보 행사를 진행했다.

또한 강연 영상을 국립국어원 누리집과 누리 소통망에 게재할 수 있도록 제작하였다. 일부 강연 영상은 문화체육관광부 아이나루 공개용으로 배포하고자 한다.

2.2.4.1 후기 작성 행사

말뚱치 지식 강연회 참가자를 대상으로 누리 소통망에 후기를 남기는 행사를 진행했고, 좋아요 64개, 댓글 569개, 공유 7회를 기록했다.

강연회 당일 쉬는 시간마다 진행한 문제 풀이 행사 당첨자(4명)와 함께 후기 행사 당첨자(1명)를 추첨하여 총 5명의 당첨자를 선정하였다.



〈그림 41〉 말뚱치 지식 강연회 후기 작성 행사

말뚱치 지식 강연회 후기는 전자 우편 및 댓글, 참가자 개인 누리 소통망 게시물 등으로 다양하게 작성되었다.

후기 내용은 주로 온라인 행사 플랫폼의 시스템 오류로 초반에 접속이 어려웠지만, 전반적으로 강연의 구성이 알차고 좋았다는 의견이었다. 또한 온라인 강의로 진행되다 보니 실제로 방문하지 않아도 좋은 강의를 지방 및 해외에서 들을 수 있어서 좋았다는

의견도 있었다.

아이들과 함께 시청한 임현숙님은 말뚝치 지식 강연회에 대하여 다음과 같이 언급했다.

#말뚝치지식강연회

말뚝치 지식 강연회 후기

코로나로 강연회를 온라인으로 열어 집에서 아이들과 함께 유익한 강연을 들었습니다. 네 분의 강연자가 나오셔서 네 시간 가량 강연을 해 주셨습니다. 각 분야의 전문가, 대표님들이 생소하고 익숙치 않았던 AI에 대해 체계적이고 자세한 설명을 해 주셔서 무척 흥미로웠습니다. 예전에 재미있게 보았던 영화를 통해서도 설명해주어 좋았고 북한어 말뚝치가 정리되어 있다는 정보를 알게 되어 유익했습니다. AI는 컴퓨터 연산 방식인데 생각보다 정보를 응용하는 능력이 탁월해서 놀랐습니다. 자료를 비슷한 것끼리 분류해서 우리가 명령어를 주면 관련어를 찾는다는 사실도 알았습니다. AI 관련 기술이 생각보다 많이 진보되어 있다는 사실도 느꼈습니다. 재미있고 유익한 강연을 온라인으로 참여할 수 있어 너무 좋았습니다(임현숙, 2020).

한글문화연대의 우리말 가꿈이 18기로 활동한 이혜림님은 강연의 전체적인 내용을 후기로 담았다.

안녕하세요,

이번엔 강연회 후기를 적어보려고 합니다!

우리말가꿈이로 활동하면서 국립국어원에 더욱 더 관심을 갖게 되었는데요.

알고 보니, 국립국어원에서 '심표 마침표'라는 소식지를 운영하고 있더라고요!

그래서 구독해놓고 보다가, 좋은 행사를 발견하게 되어 참여까지 해보았습니다.

순서는 아래와 같습니다.

1. 민규동 감독과 주성철 평론가 - '말대꾸를 시작한 인공지능' (영화 속 말뚝치 관련 인공지능)
2. 남세동 보이저엑스 대표 - '인공지능에 말뚝치를 더하다' (인공지능 시대를 앞당긴 주요 기술과 컴퓨터가 언어를 처리하는 전반적인 내용 소개)
3. 이경일 솔트룩스 대표 - '인공지능을 완성시키는 언어 빅데이터' (컴퓨터에서 언어자료를 처리하기 위해 다루어 왔던 여러 방법과 현재 활용하고 있는 기술, 그리고 이에 대한 전망)
4. 송길영 다음 소프트웨어 부사장 - '나!, 너.. 우리?' (다양한 언어 빅데이터 활용, 분석해 얻은 우리 삶의 모습과 변화 양상에 대해 이야기)

4분의 강연자께서 다양한 주제로 강연을 진행해주셨어요.

개최식부터 폐회식까지 모두 참여하였습니다!

저는 강연을 들으면서, 평소 관심있었던 '빅데이터'분야에 대해 더 알아갈 수 있었던 것 같습니다.

영화 속 인공지능이라는 강연에서 인상 깊었던 건

직접 말을 하면 인공지능이 인식해 그래픽 기술을 시연해준다는 점이었어요.

인공지능이 이제는 영화에서도 활약하고 있다는게 놀라웠습니다.

(헤이 ~ 기가지니~ 카카오~ 를 넘어서는 ... ㅎㅎㅎ)

언어 빅데이터에 우리의 삶이 있다는 강연을 통해서는

앞으로 언어 생활 습관을 올바르게 (?) 가져야겠다는 생각을 했어요.

되도록이면 긍정적인 말들로 살아야겠다는 생각과 함께 자기성찰을 할 수 있었던 것 같네요... ㅎㅎㅎ

데이터가 중요해진 인공 지능에서는
앞으로 정말 데이터가 하나의 자원으로서 그 중요성이 커질 것으로 전망됨을 다시 한번 더 느낄 수 있었어요.
(데이터 관련 공부 열심히 해야겠습니다 !!!!! 데이터 과학 관심 있는 만큼 더 열심히 프로그램을 익혀 실습해보도록 노력하겠습니다)

언어 빅데이터 말뭉치란에서는
퀴즈에서 나왔던 병렬 말뭉치, 원시 말뭉치, 상호 참조 해결 말뭉치 등등 다양한 말뭉치들의 유형을 살펴볼 수 있었고, 그에 따른
연구 결과들을 알 수 있었던 것 같습니다.

+ 인하대 연구원분들의 내용도 인상 깊었습니다. 다양한 연구를 많이 하고 계신다는 것을 알게 되었어요.
국립국어원에서도 많은 노력을 하고 계시는 것도 알게 되었구요!

전반적으로 유익한 강연이었기에 만족했습니다!
하지만, 시스템 상의 오류가 계속 있었는데 $\pi\pi$ 그 점이 아쉬웠습니다.
(크롬 자주 새로고침 되더라구요... 자주 끊기구...)

아무쪼록 비대면으로 이뤄져야 하는 상황 속에서도
이렇게 마련해 진행해주셔서 고맙습니다!
4시간 정도의 긴 강연이었음에도 불구하고, 끝까지 잘 지켜주셨어요!

다음에 또 기회가 있다면, 신청해서 듣고 싶어요 ㅎㅎ
(아침, 그리고 더 많은 홍보가 이뤄졌으면 하는 바람입니다.... 많이 알려주세요....)

앞으로도 관심 갖고 보도록 하겠습니다!
후기 이상 마칠게요!!
-끝-(이혜림, 2020)

위니 최(Winny Choi) 님은 후기만으로 전체 강연 내용을 알 수 있을 만큼 강연마다
느낀 점을 후기로 작성하였다.

“말뭉치란?

언어 연구를 위해 텍스트를 컴퓨터가 읽을 수 있는 형태로 모아 놓은 언어 자료. 매체, 시간, 공간, 주
석 단계 등의 기준에 따라 다양한 종류가 있다.

국립국어원 표준국어대사전“

내가 말뭉치라는 개념을 처음 접한 건, 영단어장 빅보카를 만났을 때였다.

그 당시 사용 빈도수에 따른 단어의 나열이라는 과학적인 접근법이 평소 영단어의 넘사벽 앞에 좌절하
던 내게는 신선한 충격이었다.

'한국어 단어장 역시 사용빈도에 따라 단어 순위를 매긴다면 한국어를 공부하는 사람들에게 도움이 될
텐데... 언젠가 내가 그런 단어장을 만들 수 있을까?'라는 막연한 생각이 잠시 잠깐 나를 스쳐지났을
뿐이었다.

구글이 그랬던 것처럼, 우리의 포털사이트들도 한국어 말뭉치 데이터를 충분히 모으고 있을까?라는 의
심 어린 질문을 뒤로하고, 나의 몸과 마음은 그냥 다시 일상으로 돌아왔다.

그 후로 몇 년, 국립국어원 주최로 말뚱치 지식 강연회가 열린다는 소식을 들었을 때, 잊고 있던 '말뚱치'에 대한 기억이 떠올랐다. 그리고 호기심에 참여했던 강연회는 내 상상을 훨씬 뛰어넘는 새로운 충격이었다.

“말대꾸를 시작한 인공지능 (영화 속에 나타난 인공지능 기술)
by 민규동 감동, 주성철 평론가”

AI는 얼마나 우리 삶에 가까이 와 있는 걸까?

부르지 않아도 시도 때도 없이 등장해서는 자기 하고 싶은 말만 하는 SIRI 덕에 실생활에서의 AI에 대한 기대치는 별로 높지 않았다. 이세돌과의 바둑 경기, 퀴즈 프로 우승... 그 어디쯤에 자리하고 있던 AI에 대한 내 생각이 첫 강연부터 산산조각이 나버렸다. 퇴행성 관절염으로 마우스 클릭조차 힘이 든다는 천계영 작가가 단지 몇 마디 명령어로 웹툰을 그리는 영상을 시작으로 소설과 시나리오, 시를 쓰는 AI에 관한 영상은 충격이라는 말로도 표현할 수 없을 만큼 충격적이었다. 여러가지 AI에 관한 영화도 소개되었는데, 그중 Her라는 영화의 AI 비서에게 의존하고 AI 비서 없이는 생활이 유지되지 않는다는 내용은 스마트폰 중독이 심각한 요즘 우리의 모습을 투영한 머지않은 미래의 모습을 보여주는 것 같았다.

“인공 지능에 말뚱치를 더하다.
(스스로 학습하는 컴퓨터가 만드는 새로운 세상)

사랑 - 이별 = 추억

남세동 보이저엑스 대표”

두 번째 강연에서는 컴퓨터가 어떻게 말뚱치로 딥러닝을 하는지에 대한 과정을 쉽게 알 수 있었다. 인간과는 달리 무수한 말뚱치를 비슷한 관련 친구어로 묶어 숫자로 그 위치를 저장하는 컴퓨터에게 사랑과 이별에 대한 공식을 물었을 때, 그 대답은 내 마음을 톡 하고 터뜨려 버리기 충분했다. 어쩌면 인간보다도 더 인간적인 냄새를 풍기는 컴퓨터가 내 마음속으로 훅 들어와 버렸다.

“인공 지능을 완성시키는 언어 빅데이터
(언어 빅데이터인 말뚱치를 다루는 방법)

이경일 솔트룩스 대표”

세 번째 강연에서는 실제 한국어 말뚱치로 딥러닝을 통해 만들어진 캐릭터들을 볼 수 있었는데, 그중 어린이 유튜브 말뚱치로 딥러닝을 시킨 어린이 캐릭터와 일반 성인 유튜브 말뚱치로 딥러닝을 시킨 어린이 캐릭터의 대화는 인간이 어떤 방식으로 컴퓨터를 다뤄야 하는지에 대한 생각을 하게 했다. 마치 어린아이를 교육하는 부모의 마음이라고 해야 할까... 결국, SF 영화에 등장하는 나쁜 AI 역할도 결국 인간의 탐욕이 부르는 결과가 아닐까 하는 생각에 마음이 불편해졌다. '홍익인간'의 정신이야말로 지금 현 인류에게 더욱 절실한 게 아닐까?

“나! 너... 우리? (언어 빅데이터로 분석한 우리 삶의 변화)
가치관의 액상화

송길영 다음소프트 부사장”

네 번째 강연에서는 포털에 등장하는 언어 빅데이터를 통해 코로나 이후 우리의 삶을 조명해 주셨다. 그중 몇몇 부분은 평소 생각해 보지 않았던 내용이라 신선하게 다가왔다. 한국이 코로나 사태에도 재기가 없었던 것은, 그전에 이미 우리 생활에 존재했던 빠른 배송 시스템으로 인해 불안함을 느끼지 않았던 것이 하나의 이유라는 것. 오늘 배송하면 내일 새벽 배송으로 내가 물건을 받을 수 있다는 확실함이 우리 안에 불안을 잠식시킨다는 시각은 그간 수준 높은 국민성으로 비춰졌던 것보다 더 수긍이 갔다. 코로나로 인해 시도되었던 여러 가지 대응책이 실시간 수정되고 발전될 수 있었던 것에는 실

시간 댓글이라는 언어자원 즉 말뭉치가 있었기 때문이며, 따라서 언어자원은 굉장히 소중하게 다뤄져야 한다는 것. 소셜 디바이드를 넘어 디지털 디바이드가 앞으로 우리 세대를 어떻게 나눌지에 대한 시각 등등....

여러 가지 부분에서 깊게 생각해 볼 질문을 던져주신 것 같다.

단순히 '말뭉치 지식'이라는 호기심에 들었던 강연 치고는 그 내용이 너무나 좋았다. 인간이 이 지구를 마치 자기 것인 양 사용하고 있는 것이 당연한 것이 아니라는 생각, 그리고 호모사피엔스가 이 지구를 거의 장악 한 것처럼 AI가 언젠가 인간을 밀치고 지구를 장악할 시기가 올 수도 있다는 생각과 더불어 AI를 어떻게 다루고 함께 살아갈지에 대한 고민을 진지하게 해야 할 시기가 이미 우리 곁에 와 있음을 체감한 시간이었다. 좀 더 많은 사람들이 함께 할 수 있도록 강연 영상이 공개되었으면 하고 바라본다(최은희, 2020).

강성화(Sh Kang) 님은 강연회에 대해 다음과 같이 언급했다.

지난 2월 코로나19로 인해 연기되었던 말뭉치 지식 강연회가 6월 18일, 약 4시간 동안 진행이 되었다. 말뭉치가 인공 지능 기술에 어떻게 활용되고 있는지, 말뭉치를 통해 어떻게 사회를 들여다 볼 수 있는지를 알 수 있었던 시간.

지금까지 한 번도 관심을 가져 보지 못했던 분야이기에 내용도, 용어도 무척이나 낯설었지만 그렇기에 배우는 즐거움을 느낄 수 있었고 새로운 지식들을 쌓을 수도 있는 시간이었다. 기대보다 훨씬 더 흥미로운 내용에 하나라도 놓칠까 점심 굶으며 집중했던 것이 아깝지 않았던 강연회. 이런 강연회가 자주 열려 보다 많은 이들이 우리말에 대한 관심을 가질 수 있었으면, 소중한 우리말을 지키는 것이 얼마나 중요한 것인지를 알 수 있었으면 좋겠다는 생각을 해 본다. 또한 빠르게 변화하는 사회에 발맞추어 언어를 이해하려는 끊임없는 시도와 노력도 필요하다는 것을 깨닫는다.

온라인으로 진행된 강연회이기에 준비 과정이 결코 쉽지 않았을 것이라고 짐작이 된다. 이런 좋은 강연회를 위해 애써 주신 모든 분들께 고생하셨다는 말씀, 전하고 싶다(강성화, 2020).

2.2.4.2 말뭉치 지식 강연회 만족도·개선 사항 설문 조사

강연회 종료 후 더 나은 말뭉치 지식 강연회를 위해 만족도 및 개선 사항에 대한 설문 조사를 진행했다. 조사는 말뭉치 지식 강연회 참여자 대상으로 구글(Google) 온라인 설문지를 발송했으며, 2020년 7월 2일(목)부터 2020년 7월 9일(목)까지 일주일간 총 253명의 참여자들이 응답하였다.

설문 조사 결과, 강연회 생방송 초반 발생한 송출 문제로 인해 낮은 만족도를 표한 응답자가 있었으나 응답자(253명) 중 절반이 넘는 인원(191명)이 강연회에 대해 높은 만족도를 보인 것으로 나타났다.

설문 조사 응답자의 약 44%의 응답자가 민규동 감독, 주성철 평론가와 송길영 부사장의 강연을 가장 기대했다고 답변하였지만, 업무나 학업에 가장 도움이 된 강연으

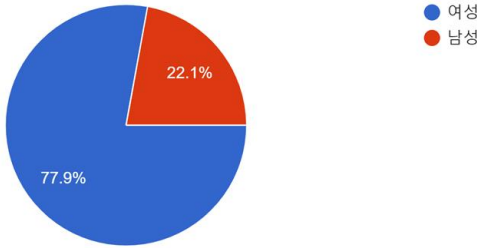
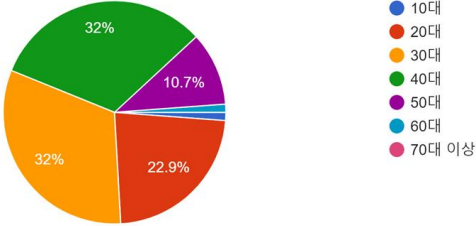
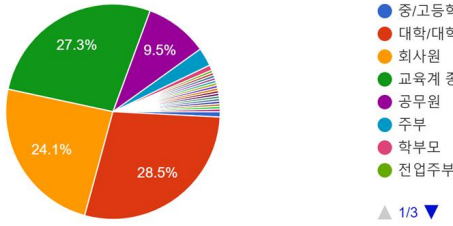
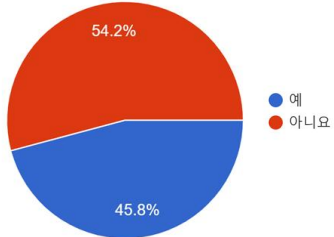
로는 48%의 응답자가 이경일 대표의 강연을 꼽았다. 이는 말뚝치에 대한 쉬운 강의보다는 전문적인 강의를 원하는 참가자들의 의견이 반영된 것이라 할 수 있다. 따라서 앞으로 말뚝치 지식 강연회를 개최할 때에는 말뚝치 관련 연구나 기술에 대해 보다 전문적인 내용을 다루어도 좋을 것으로 판단된다.

설문 조사 대상 기본 정보에 따르면 강연회 참가자들 중 대학/대학원생 및 회사원, 교육계 종사자(202명)가 주로 강연회에 참석한 것으로 나타났다. 이러한 결과는 평일 낮 시간, 관련 분야 종사자가 아님에도 불구하고 충분히 관심이 있을 만한 강연회였다는 것을 의미한다고 볼 수 있다.

<표 20> 강연회 만족도 및 개선 사항 설문 조사

구분	표	데이터(명)	
1	<p>본 강연회는 어떤 경로로 알게 되셨나요?</p> <p> ■ 국립국어원 공식 홈페이지 ■ 국립국어원 공식 SNS(페이스북 등) ■ 대학/대학원 홈페이지 공지사항 ■ 네이버 커뮤니티 배너 광고 ■ 친구 및 주변 지인 추천 ■ 기사 및 보도자료 ■ 기타 </p>	국립국어원 공식 누리집	41
		국립국어원 공식 누리 소통망 (페이스북 등)	33
		대학/대학원 누리집 공지사항	42
		네이버 커뮤니티 배너 광고	7
		친구 및 주변 지인 추천	85
		기사 및 보도자료	18
		기타	27
2	<p>강연을 듣기 전 가장 기대했던 강연을 선택해 주세요.(복수 선택 가능)</p>	첫 번째 강연) 민규동 감독, 주성철 평론가	112
		두 번째 강연) 남세동 보이저엑스 대표	72
		세 번째 강연) 이경일 솔트룩스 대표	100
		네 번째 강연) 송길영 다음소프트 부사장	116
3	<p>강연을 들은 후 업무나 학업 등에 가장 도움이 된 강연을 선택해 주세요.(복수 선택 가능)</p>	첫 번째 강연) 민규동 감독, 주성철 평론가	62
		두 번째 강연) 남세동 보이저엑스 대표	103
		세 번째 강연) 이경일 솔트룩스 대표	123
		네 번째 강연) 송길영 다음소프트 부사장	102
4	<p>어떤 방식의 강연회를 선호하시는지 선택해 주세요.</p> <p> ■ 온라인 강연회 ■ 오프라인 강연회 ■ 온라인과 오프라인이 결합된 형태의 강연회(현장 강의 참석과 온라인 생중계 동시 진행) </p>	온라인 강연회	85
		오프라인 강연회	40
		온·오프라인 통합 강연회	128
5	<p>강연회에 대한 전반적인 만족도는 어느 정도인가요?</p>	1 별로 만족하지 않음	5
		2 만족하지 않음	14
		3 보통	40
		4 만족	125
		5 아주 만족	66

<표 21> 강연회 만족도 및 개선 사항 설문 조사 대상 기본 정보

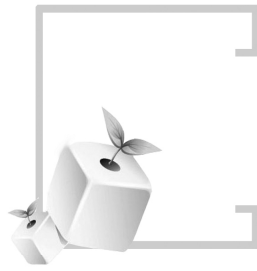
구분	표	데이터(명)	
1	<p>귀하의 성별을 선택해주세요.</p>  <p>● 여성 ● 남성</p>	여성	197
		남성	56
2	<p>귀하의 연령대를 선택해 주세요.(만 나이를 기준으로 선택해 주세요.)</p>  <p>● 10대 ● 20대 ● 30대 ● 40대 ● 50대 ● 60대 ● 70대 이상</p>	10대	3
		20대	58
		30대	81
		40대	81
		50대	27
		60대	3
		70대 이상	0
3	<p>귀하의 직업을 선택해 주세요.</p>  <p>● 중/고등학생 ● 대학/대학원생 ● 회사원 ● 교육계 종사자 ● 공무원 ● 주부 ● 학부모 ● 전업주부</p> <p>▲ 1/3 ▼</p>	중/고등학생	2
		대학/대학원생	72
		회사원	61
		교육계 종사자	69
		공무원	24
		기타(자영업, 학부모 등)	25
4	<p>본 강연회 관련 분야에 종사하시나요?(자연어 처리, 언어학, 인공지능 등)</p>  <p>● 예 ● 아니요</p>	예	137
		아니요	116

2.2.4.3 강연 영상 제작

강연회 현장 요약 영상 및 생중계 송출 영상, 휴식 시간에 송출한 전시 영상을 국립국어원 소장용 혹은 배포용으로 제작했다. 현장 요약 영상은 국립국어원 공식 누리집 및 누리 소통망에 게시할 예정이며, 강연 영상은 저작권 문제로 인해 문화체육관광부 아이나루 공개와 내부 소장용으로만 사용할 수 있다.

〈표 22〉 강연회 결과 동영상 목록

구분	콘텐츠	배포처	비고
1	현장 요약 영상(1분 30초)	공식 누리집 및 누리 소통망 (유튜브, 페이스북 등)	국어원 자료로 활용 가능
2	현장 요약 영상(4분)		
3	강연회 전체 송출 영상	내부 소장용	저작권 문제로 외부 송출 불가
4	전체 촬영본		
5	남세동 생중계 송출 영상		
6	민규동, 주성철 생중계 송출 영상	아이나루	강연 자료 저작권 문제로 강연자 근접 촬영본만 아이나루 공개 가능
7	이경일 생중계 송출 영상	아이나루, 유튜브	국어원 자료로 활용 가능
8	송길영 생중계 송출 영상	배포불가	저작권 문제로 영상 녹화 및 편집 금지
9	천계영 음성 인식 활용 작품 제작 영상		유튜브 링크 활용 가능
10	‘성.시.경’ 팀 관련 영상	아이나루	한국콘텐츠진흥원 로고 삽입 관련 협의 중
11	‘AI人’ 팀 시 생성 시연 영상		



제 3 장

사업 수행 결과



3.1 개인적 글쓰기 자료 수집 결과

3.1.1 연령 및 직업별 수집 현황

글쓰기 자료 수집 과업 ‘한 편 줍쇼’에 참여해서 글을 제출한 참가자는 총 6,044명으로 집계되었다. 당초 목표로 했던 6,000명 이상을 수집하였다.

〈표 23〉 연령별 ‘한 편 줍쇼’ 참여자 구성

구분	제출자 수(명)	비중
만 7세 ~ 만 12세	800	13.2%
만 13세 ~ 만 15세	497	8.2%
만 16세 ~ 만 18세	1,058	17.5%
만 19세 ~ 만 29세	2,720	45.0%
만 30세 ~ 만 59세	648	10.7%
만 60세 이상	321	5.3%
전체	6,044	100.0%

참가자들의 직업은 대학생/대학원생이 29.6%로 가장 많이 차지하였으며, 고등학생 13.2%, 초등학생 13.0%, 중학생 8.0% 순이다. 성인은 직장인/전문직 15.5%, 무직/기타 13.9%, 전업주부 4.3%, 자영업자가 2.5%를 차지하는 것으로 나타났다.

〈표 24〉 직업별 ‘한 편 줍쇼’ 참여자 구성

구분	제출자 수(명)	비중
초등학생	786	13.0%
중학생	482	8.0%
고등학생	797	13.2%
대학생/대학원생	1,790	29.6%
직장인/전문직	936	15.5%
전업주부	260	4.3%
자영업자	154	2.5%
무직/기타	839	13.9%
전체	6,044	100.0%

총 제출 글 수는 10,919편으로 집계되었다. 1인당 약 1.80편을 작성해서 제출했으며, 총 글자 수는 9,250,097자에 달하는 것으로 산출되었다.

〈표 25〉 연령별 글 수집 결과

구분	글자 수(자)	수집 글(편)	비중
만 7세 ~ 만 12세	1,063,836	3,119	28.6%
만 13세 ~ 만 15세	765,639	854	7.8%
만 16세 ~ 만 18세	1,754,297	1,688	15.5%
만 19세 ~ 만 29세	4,325,375	3,926	36.0%
만 30세 ~ 만 59세	899,281	862	7.9%
만 60세 이상	441,669	470	4.3%
전체	9,250,097	10,919	100.0%

3.2 말뚱치 지식 강연회 결과

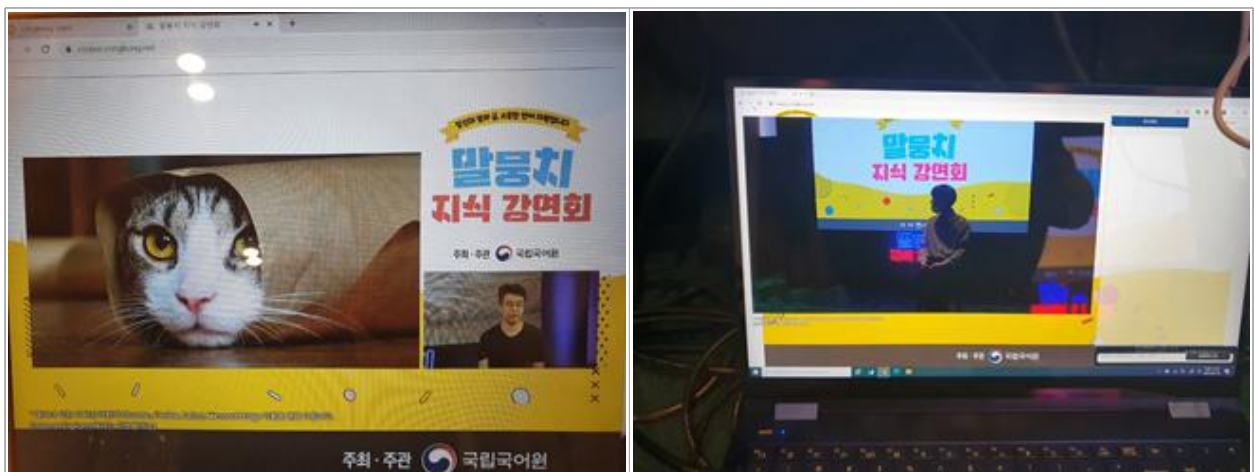
말뚱치 지식 강연회 결과는 송출, 강연, 문제 풀이 행사 부분으로 총 3개 분야로 나누어 평가할 수 있다.

3.2.1 말뚱치 지식 강연회 종합 평가

3.2.1.1 운영 결과(송출 부문)

강연자들의 강연 자료 저작권 문제로 인해 외부에 강연이 유출이 되지 않도록 외부의 접속이 제한된 플랫폼 콩콩(CongKong)에서 강연회 영상을 송출했다. 참가자 입장을 관리할 수 있도록 참가자마다 개별 입장 코드를 부여했다. 온라인으로 진행되어 서울 외 지방 및 외국 거주자도 참여할 수 있어 긍정적인 호응을 얻었다.

다만 댓글 오류와 시스템 불안정 현상이 발생¹⁾하여, 강연 참여자들이 불편을 겪었다. 온라인 강연회 개최에서 안정적인 시스템 운영은 기본이기에 시스템 사전 연습 포함 사전 연습을 두 차례 진행하였음에도 문제가 발생하였다.



〈그림 42〉 말뚱치 지식 강연회 송출 화면

1) 14:20 경부터 간헐적인 디비 연결 불량 문제 발생 (생중계 송출 간 새로 고침 현상). 이후 한 시간가량 디비 연결 수 조정 및 증설을 통해 문제 해결을 꾀하였으나 계속 불안정하였다. 웨비나 상세 사항 댓글 최신화 자동 적용 로직 및 게시판 연결 로직 등을 제거하고 배포하여 오류 상황을 회피하였다. 콩콩 서비스 내부적 디비 로직 오류로 판단된다.

3.2.1.2 운영 결과(강연 부문)

강연의 경우, 알차고 유익한 내용으로 구성되어 있었고 평소 접하기 어려웠던 내용을 쉽게 접할 수 있어서 좋았다는 반응이었다. 다만 전문적인 내용을 다룰 때 다소 이해하기 어려웠다는 평도 일부 있었다. 이에 강연자 순서를 조정하거나 강연자를 적절하게 배치하여 강연의 난이도를 조절하는 것도 좋을 것으로 판단된다.

3.2.1.3 운영 결과(문제 풀이 행사 부문)

문제 풀이 행사는 송출 플랫폼 업체에서 제공하는 실시간 댓글 서비스를 통해 진행할 예정이었으나, 송출 과정에 발생한 오류로 인해 접수 방법을 강연회 중간 변경했다. 공정한 경품 추첨을 위해 문제 풀이 행사에 참여할 수 있는 시간을 늘렸으며, 정답은 국립국어원 페이스북 담벼락 댓글과 말뭉치 지식 강연회 운영 사무국 전자 우편으로 받았다.

시스템 오류에 빠르게 대처했다는 반응도 있었지만, 누리 소통망을 사용하지 않는 참가자들에게 배려가 필요했다는 반응도 있었다. 온라인 행사 운영 시에는 시스템 오류에 대비하여 참여자 전원이 경품 행사에 참여할 수 있는 별도 방안을 미리 준비할 필요가 있다.

3.2.1.4 실시간 접속 결과

말뭉치 지식 강연회 사전 신청자는 총 652명이었으나 접속 여부 확인 결과, 강연회에 참석한 사람은 총 524명²⁾으로 나타났다. 강연 당 참석자는 평균 120~130명 내외였으며, 강연 시작 전 공식 행사 참석자가 223명으로 가장 많은 것으로 나타났다.

2) 최소 로그인 1회 기준

〈표 26〉 온라인 강연회 평균 접속자 결과

구분	평균 접속자 현황(명)	시간	내용
입장	87	14:20~14:30	참가자 입장 대기
공식 행사	223	14:30~14:35	개회
		14:35~14:40	축사 / 개회사
강연	121	14:40~15:25	민규동 감독, 주성철 평론가 - 말뭉치 사례
	134	15:25~15:35	휴식 시간(전시 영상 송출) - 성.시.경 팀
		15:35~15:40	휴식 시간(문제 풀이 1건)
	110	15:40~16:25	남세동 대표 - 자연어 처리 1
		16:25~17:10	이정일 대표 - 자연어 처리 2
	141	17:10~17:20	휴식 시간(전시 영상 송출) - AI人 팀
		17:20~17:25	휴식 시간(문제 풀이 1건)
	151	17:25~18:10	송길영 부사장 - 4차 산업
폐회	140	18:10~18:15	경품 추첨(문제 풀이 2건)
		18:15~	참가자 기념품 안내
평균 시간당 접속자(명)		120~130 내외	
최종 입장 기록 인원(명)		524	

참고문헌

- 강성화(2020년 6월 19일). 말뚱치 지식 강연회(feat. 국립국어원) 후기. 2020년 6월 19일 수정, <https://m.blog.naver.com/nangfaforu/222006414835>
- 이혜림(2020년 6월 20일). [국립국어원 행사] 말뚱치 지식 강연회 후기. 2020년 6월 20일 수정, <https://blog.naver.com/luvme121/222007036638>
- 임현숙(2020년 6월 21일). #말뚱치지식강연회. 2020년 6월 21일 수정, <https://www.facebook.com/100050373859759/posts/155436262812168/>
- 최은희(2020년 6월 21일). 말뚱치 지식 강연회 후기. 2020년 6월 21일 수정, <https://blog.naver.com/winniec/222008092362>

부록



1. 개인적 글쓰기 자료 수집 저작권 이용 허락계약서

귀하가 제출하신 저작물의 사용 목적과 이용 허락 대상 권리는 아래와 같습니다.

국립국어원은 대규모 언어 자료를 구축하여, 이를 국어 연구 및 자연언어처리 기술 개발 등을 위해 사용하고자 합니다. 이에 귀하의 저작물의 활용에 대한 승인을 구하고자 합니다. 귀하의 저작물을 이용하여 구축한 말뭉치가 귀하의 저작·출판에 관한 어떠한 권리에도 손상을 입히지 않을 것임을 약속드립니다.

모쪼록 21세기 4차 산업 혁명 시대에 귀하의 저작물을 우리말 정보 처리 발전의 기초가 되는 국가적 언어 자원의 구축에 유용하게 활용할 수 있도록 협조하여 주시면 감사하겠습니다. 이를 위해 이용 허락에 앞서 아래 내용을 읽어 주시길 바랍니다.

- 국립국어원 일상 글 수집 ‘한 편 줍쇼’ 참가 본인은 제출 저작물의 제출된 모든 내용 및 문서는 본인이 작성한 저작물임을 확인합니다.
- 제출한 저작물이 제3자와 분쟁소지 가능성이 있는 저작물에 대해서는 제출하지 않습니다.
- 저작물이 제3자의 소유권, 저작권, 초상권 등 기타 제반 권리를 침해하지 않았음을 보증하여야 하며, 향후 분쟁이 발생할 경우 일체의 법적·도의적 책임은 해당 저작물을 제출한 제출자에게 있습니다.
- 타 공모전에서 시상 또는 수상 경력이 있는 작품, 표절 작품으로 판명될 경우 국립국어원 사용 용도에서 제외되며, 사례금은 지급되지 않습니다.
- 제출된 저작물은 이용허락 기간 내에서 국립국어원에서 사용하며, 제출자가 이용 허락 중지 의사를 밝히면, 의사 내용에 따라 이용 허락을 중지하여야 합니다.
- 제출된 저작물은 개별 단어 및 문장, 텍스트에 대한 복제·변형(편집 및 삭제, 정리, 분리 등) 및 정보 추출에 쓰이며, 연구 및 기술 개발용으로 학계·연구 기관·사업체 등에 제공·배포하여 분석하는 데 쓰입니다.

국가 언어 자원(말뭉치) 구축 및 활용 저작권 이용허락 계약서

저작자 및 저작권 이용허락지 홍길동 (이하 “권리지”이라 함)와 저작권 이용자 국립국어원(이하 “이용자”이라 함)은 아래 저작물에 관한 저작권 이용허락과 관련하여 다음과 같이 계약을 체결한다.

다 음

제1조 (계약의 목적)

본 계약은 저작권 이용허락과 관련하여 권리와 이용자 사이의 권리관계를 명확히 하는 것을 목적으로 한다.

제2조 (계약의 대상)

본 계약의 이용허락 대상이 되는 권리는 아래의 저작물(이하 “대상저작물”)에 대한 저작권 중 당사자가 합의한 권리로 한다.

저작물: 저작자가 국립국어원의 <개인적 글쓰기 자료 수집 및 말뭉치 지식 강연회 개최> 사업 기간(2020년 01월 31일 ~ 2020년 3월 14일까지) 동안 위 사업에 제공하는 모든 개인 비출판물

저작자: 홍길동

종별: ☒ 어문저작물

권리: ☒ 복제권, ☒ 전송권, ☒ 배포권, ☒ 2차적저작물작성권

※ 저작권 이용허락 대상 권리의 내용

1. 국립국어원 및 국립국어원이 발주한 용역 사업의 수행자가 대상저작물을 일정한 형식으로 전자적 기록 매체에 담아 보존하는 일
2. 국립국어원 및 국립국어원이 발주한 용역 사업의 수행자가 자모, 음절, 어휘, 어절, 구절, 문장 및 텍스트 단위의 국어 연구와 언어 정보 처리 분야에 응용하기 위해 대상저작물을 복제·변형(목차·머리말·도표·그림·각주 등의 편집 및 삭제, 언어 단위별 분리, 언어적·비언어적 정보 부착 등)하는 일
3. 국립국어원 및 국립국어원이 발주한 용역 사업의 수행자가 대상저작물 및 그 복제·변형물을 연구 및 기술 개발용으로 학계·연구기관·산업체 등이 이용할 수 있도록 제공·배포하는 일
4. 대상저작물 및 그 복제·변형물을 제공·배포받은 학계·연구기관·산업체 등이 국어 연구와 언어 정보 처리 분야 응용을 위하여 대상저작물 및 그 복제·변형물을 분석 및 처리하여 사용하는 것을 허락하는 일

☐ 상기 내용을 확인하였으며, 동의합니다.

제3조 (이용허락 기간)

대상저작물의 이용허락 기간은 계약 체결일부터 2040. 12. 31.까지로 한다. 최초 계약 체결 및 자동 갱신에 따른 이용허락 기간의 종료 시점까지 관리자가 이용허락을 중지하고자 하는 의사를 밝히지 아니하면 이용허락이 5년 단위로 자동 갱신되어 이용허락 내용이 유지되며, 관리자가 이용허락 중지 의사를 밝히면 그 의사 내용에 따라 이용허락을 중지하여야 한다.

제4조 (관리자의 의무)

(1) 관리자는 이용자에게 대상저작물에 관하여 본 계약서 제2조에 따른 저작권산을 이용할 권리를 제3조의 기간 동안 비독점적으로 허락한다.

(2) 관리자는 이용자에게 사업 기간 내 대상저작물의 이용을 위해 필요한 상당한 자료를 인도하여야 한다.

(3) 관리자는 대상저작물에 제3자의 이용허락권, 질권 등이 존재하는 경우, 이용자에게 그 사실을 사전에 알려야 한다.

(4) 관리자는 대상저작물의 저작권산권 전부 또는 일부를 제3자에게 양도하거나 이에 대하여 질권을 설정하고자 하는 경우, 사전에 이용자에게 이 사실을 통보하여야 한다.

(5) 관리자가 저작물을 제공할 시 타인의 권리를 침해하여서는 아니 된다. 이와 관련하여 발생하는 문제에 대해서는 관리자가 그 책임을 부담하며, 국립국어원은 그 책임으로부터 면책된다.

☐ 상기 내용을 확인하였으며, 동의합니다.

제5조 (이용자의 권리 및 의무)

(1) 이용자는 대상저작물을 제3조의 이용허락 기간 동안 제2조의 이용 허락을 받은 범위 내에서 비독점적으로 자유롭게 이용할 수 있다.

(2) 이용료는 설정하지 아니한다.

(3) 관리자가 대상저작물의 저작명 공개를 원하지 않는 경우 이용자는 대상저작물에 관리자의 성명을 표시하지 않는다.

(4) 이용자는 대상저작물을 이용함에 있어서 저작인격권을 침해하지 아니한다. 다만, 제2조에 따른 목적에 한하여 제2조에 따른 변형을 할 수 있으며, 대상저작물의 본질적인 내용을 변경하지 않는 범위 내에서 권리자에게 그 사실을 사전에 고지한 후 사소한 수정 및 편집을 할 수 있다.

제6조 (확인 및 보증)

(1) 권리자는 이용자에게 다음 각 호의 사항을 확인하고 보증한다.

1. 대상저작물의 저작권이용허락을 체결하는 데 필요한 권리 및 권한을 적법하게 보유하고 있다는 것
2. 대상저작물의 내용이 제3자의 저작권, 상표권, 인격권을 비롯한 일체의 권리를 침해하지 아니한다는 것
3. 대상저작물에 대하여 이용자에게 사전에 알린 제3자의 권리 외에는 이용자의 이용을 제한할 수 있는 부담이 더 이상 존재하지 아니한다는 것

(2) 이용자는 권리자에게 다음 각호의 사항을 확인하고 보증한다.

1. 대상저작물에 적용된 이용허락 조건에 의해서만 대상저작물 재이용을 허락할 것
2. 대상저작물을 제3자의 명예권을 비롯한 인격적 권리를 침해하는 방식으로 이용하지 아니할 것

☐ 상기 내용을 확인하였으며, 동의합니다.

제7조 (계약내용의 변경)

본 계약 내용 중 일부를 변경할 필요가 있는 경우에는 권리자와 이용자의 서면합의에 의하여 변경할 수 있으며, 그 서면합의에서 달리 정함이 없는 한, 변경된 사항은 그 다음날부터 효력을 가진다.

제8조 (계약의 해지)

(1) 당사자는 천재지변 또는 기타 불가항력으로 계약을 유지할 수 없는 경우에 본 계약을 해지할 수 있다.

(2) 당사자는 상대방이 정당한 이유 없이 본 계약을 위반하는 경우에 상당한 기간을 정하여 상대방에게 그 시정을 최고하고, 상대방이 그 기간이 지나도록 이행하지 아니하는 경우에는 계약을 해지할 수 있다. 다만, 상대방이 명백한 시정 거부 의사를 표시하였거나 위반 사항의 성격상 시정이 불가능하다는 것이 명백히 인정되는 경우에는 위와 같은 최고 없이 계약을 해지할 수 있다.

(3) 본 계약에 대한 해지권의 행사는 상대방에 대한 손해배상청구권 행사에 영향을 미치지 아니한다.

제9조 (손해배상)

당사자가 정당한 이유 없이 본 계약을 위반하는 경우, 그로 인하여 상대방에게 발생한 모든 손해를 배상할 책임이 있다. 다만, 제8조 1항의 사유로 본 계약을 이행하지 못한 경우에는 손해배상책임을 면한다.

제10조 (비용의 부담)

계약 체결에 따른 비용은 이용자가 전부 부담한다.

제11조 (분쟁해결)

(1) 본 계약에서 발생하는 모든 분쟁은 권리와 이용자가 상호 원만한 합의에 이르도록 노력하여야 하며, 분쟁이 원만히 해결되지 않는 경우에는 소제기에 앞서 한국저작권위원회에 조정을 신청할 수 있다.

(2) 제1항에 따라 해결되지 아니할 때에는 대한민국의 민사소송법 등에 따른 관할법원에서의 소송에 의해 해결토록 한다.

제12조 (비밀유지)

양 당사자는 본 계약의 체결 및 이행과정에서 알게 된 상대방에 관한 정보, 본 계약의 내용 및 대상저작물의 내용을, 상대방의 서면에 의한 승낙 없이 제3자에게 공개하여서는 아니 된다.

☐ 상기 내용을 확인하였으며, 동의합니다.

제13조 (기타부속합의)

(1) 권리와 이용자는 본 계약의 내용을 보충하거나, 이 계약에서 정하지 아니한 사항을 규정하기 위하여 부속합의서를 작성할 수 있다.

(2) 제1항에 따른 부속 합의는 본 계약의 내용과 배치되거나 위반하지 않는 범위 내에서 유효하다.

제14조 (계약의 해석 및 보완)

본 계약서에서 명시되어 있지 아니하거나 해석상 이견이 있을 경우에는 저작권법, 민법 등을 준용하고 사회 통념과 조리에 맞게 해결한다.

제15조 (계약 효력 발생일)

본 계약의 효력은 계약 체결일로부터 발생한다.

개인정보 수집 및 이용 동의서

국립국어원 일상 글 수집 ‘한 편 줍쇼’의 원활한 사업 추진을 위해 개인정보 보호법 제15조 제2항 관계 규정에 의거, 아래와 같이 개인정보를 수집 및 이용하고자 하오니, 내용을 숙지하시고 동의 여부를 결정하여 주시기 바랍니다.

수집항목	수집목적	보유기간
성명, 주민등록번호 앞 일곱 자리, 휴대 전화 번호, 전자 우편, 성별, 직업	인적 사항 확인, 공지 사항 통보, 사례금 지급, 통계 분류 등	2년

1. 수집된 정보는 보유 기간 경과 후 자동 폐기됩니다.
2. 개인정보는 공모전 운영 이외의 용도로 이용하거나 제3자에게 제공되지 않습니다.
3. 참가자는 개인정보 보호법 제15조 제2항 제4호에 따라 개인정보 수집 및 이용을 거부할 수 있습니다. 그러나 동의를 거부할 경우 응모자의 인적사항 확인 등이 불가함을 이유로 접수하지 않을 수 있습니다.
4. 상기 내용이 허위일 경우 접수 또는 사례금 지급이 취소됨을 사전에 미리 양지하고 있습니다.

상기 본인은 국립국어원 일상 글 수집 ‘한 편 줍쇼’ 운영 과정에서 본인의 개인정보를 수집·이용하는 것에 동의합니다.

※ 만 14세 미만의 경우 법정대리인 동의

상기 내용에 동의합니다 ☐ 상기 내용에 동의하지 않습니다 ☐

2020년 __월 __일

권리자 :

성명 : _____ (인)

주민등록번호 :

--	--	--	--	--	--

--

 × × × × × ×

* 주민등록번호를 앞에서 일곱째 자리까지 입력해 주세요.

휴대폰 번호 (000-0000-0000)

이용자 :

성명 국립국어원장 (인)

주소 서울특별시 강서구 금남화로 154



(권리자가 미성년자일 경우)

법정대리인 :

성명 : _____ (인)

주민등록번호 :

--	--	--	--	--	--

--

 × × × × × ×

* 귀하께서는 제출 글에 대한 저자명을 공개하시기를 원하십니까? ☐ 예 ☐ 아니오

2. 말뚱치 지식 강연회 접속자³⁾

〈표 27〉 말뚱치 지식 강연회 참석자 명단

번호	이름	번호	이름	번호	이름	번호	이름
1	PHAN * * DUONG	39	김*언	77	김*연	115	김*연
2	P**H	40	김*현	78	김*연	116	김*연
3	강*영	41	김*현	79	김*이	117	김*영
4	강*은	42	김*관	80	김*진	118	김*원
5	강*선	43	김*훈	81	김*화	119	김*윤
6	강*영	44	김*보	82	김*미	120	김*은
7	강*규	45	김*주	83	김*유	121	김*희
8	강*숙	46	김*주	84	김*연	122	김*영
9	강*민	47	김*미	85	김*혜	123	김*혜
10	강*은	48	김*성	86	김*영	124	김*중
11	강*호	49	김*연	87	김*근	125	김*하
12	강*화	50	김*영	88	김*경	126	김*경
13	강*미	51	김*진	89	김*옥	127	김*연
14	강*윤	52	김*현	90	김*의	128	김*웅
15	강*은	53	김*경	91	김*주	129	김*희
16	강*아	54	김*옥	92	김*림	130	김*연
17	고*현	55	김*정	93	김*슬	131	김*수
18	고*연	56	김*혁	94	김*은	132	김*주
19	고*미	57	김*환	95	김*서	133	김*주
20	공*웅	58	김*환	96	김*걸	134	김*지
21	곽*름	59	김*조	97	김*희	135	김*지
22	곽*애	60	김*경	98	김*나	136	김*창
23	곽*석	61	김*현	99	김*리	137	김*순
24	구*영	62	김*현	100	김*미	138	나*정
25	구*정	63	김*희	101	김*영	139	나*민
26	권*	64	김*경	102	김*광	140	나*선
27	권*경	65	김*복	103	김*숙	141	남*정
28	권*지	66	김*철	104	김*수	142	남*민
29	권*두	67	김*하	105	김*아	143	남*희
30	권*희	68	김*태	106	김*애	144	남*호
31	기*선	69	김*현	107	김*옥	145	남*아
32	김*아	70	김*희	108	김*온	146	노*진
33	김*영	71	김*연	109	김*원	147	노*희
34	김*엽	72	김*연	110	김*은	148	노*주
35	김*형	73	김*윤	111	김*은	149	노*근
36	김*영	74	김*희	112	김*호	150	노*정
37	김*영	75	김*희	113	김*화	151	라*정
38	김*균	76	김*빈	114	김*서	152	류*미

3) 최소 로그인 1회 기준

번호	이름
153	류*정
154	류*연
155	류*주
156	류*현
157	명*언
158	문*원
159	문*현
160	문*선
161	민*은
162	민*미
163	민*준
164	민*선
165	민*현
166	박*민
167	박*용
168	박*식
169	박*련
170	박*래
171	박*은
172	박*은
173	박*진
174	박*숙
175	박*애
176	박*현
177	박*형
178	박*좌
179	박*선
180	박*영
181	박*윤
182	박*
183	박*민
184	박*랑
185	박*현
186	박*영
187	박*빈
188	박*연
189	박*희
190	박*주
191	박*희
192	박*미
193	박*영
194	박*혜
195	박*선
196	박*숙
197	박*섭
198	박*민
199	박*배
200	박*아
201	박*훈
202	박*해
203	박*화

번호	이름
204	박*선
205	박*
206	박*준
207	박*주
208	박*연
209	방*원
210	배*빈
211	배*애
212	배*진
213	배*경
214	백*연
215	백*라
216	변*경
217	서*화
218	서*주
219	서*환
220	서*경
221	서*별
222	서*희
223	서*향
224	서*희
225	서*현
226	서*희
227	서*연
228	서*란
229	서*진
230	서*원
231	선*철
232	선*리
233	성*진
234	성*진
235	소*옥
236	손*형
237	손*호
238	손*진
239	송*경
240	송*일
241	송*숙
242	송*승
243	송*미
244	송*경
245	송*리
246	송*재
247	송*정
248	신*혜
249	신*진
250	신*남
251	신*재
252	신*은
253	신*환
254	신*석

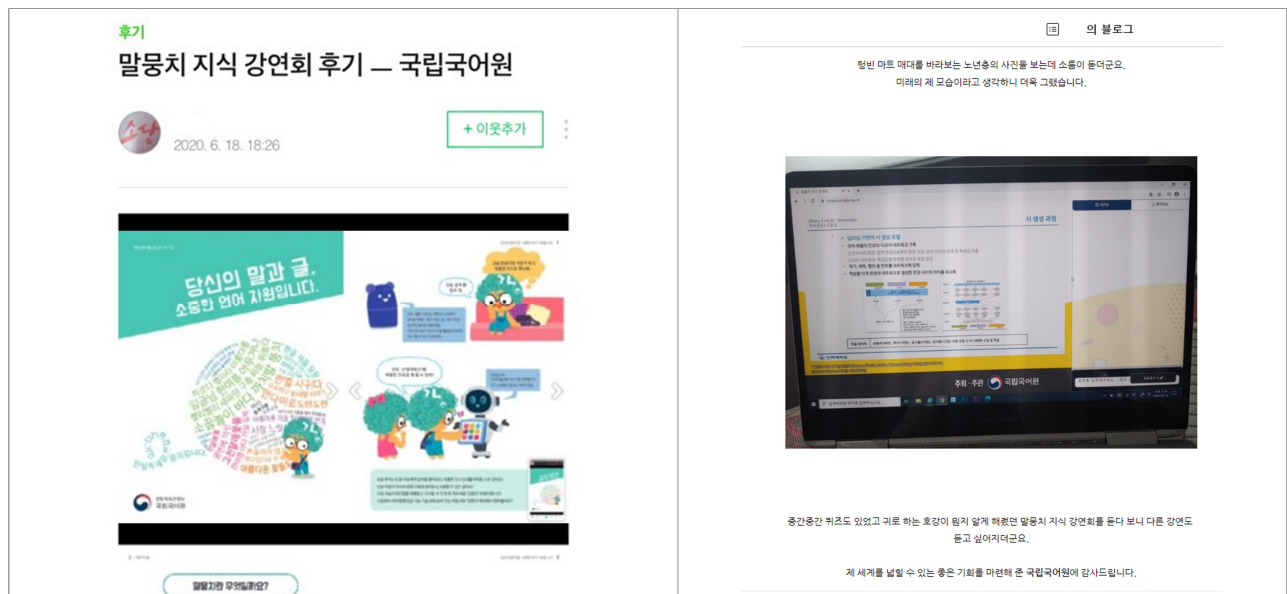
번호	이름
255	신*희
256	심*현
257	심*란
258	심*현
259	심*희
260	안*정
261	안*용
262	안*진
263	안*림
264	안*순
265	안*엽
266	안*숙
267	양*철
268	양*화
269	양*혁
270	양*민
271	양*은
272	양*경
273	양*신
274	양*정
275	양*현
276	양*영
277	양*이
278	여*만
279	연*진
280	오*비
281	오*람
282	오*주
283	오*연
284	오*진
285	오*비
286	오*진
287	오*연
288	오*성
289	오*정
290	오*인
291	옥*희
292	왕*현
293	우*혁
294	우*용
295	유*림
296	유*숙
297	유*경
298	유*미
299	유*식
300	유*희
301	유*석
302	유*선
303	유*순
304	유*은
305	유*훈

번호	이름
306	유*진
307	유*정
308	윤*은
309	윤*롬
310	윤*미
311	윤*영
312	윤*녕
313	윤*은
314	윤*영
315	윤*연
316	윤*은
317	윤*진
318	윤*찬
319	이*희
320	이*래
321	이*원
322	이*직
323	이*
324	이*수
325	이*민
326	이*재
327	이*희
328	이*연
329	이*홍
330	이*겸
331	이*래
332	이*성
333	이*은
334	이*용
335	이*란
336	이*옥
337	이*우
338	이*주
339	이*수
340	이*희
341	이*영
342	이*영
343	이*원
344	이*정
345	이*미
346	이*빈
347	이*정
348	이*정
349	이*화
350	이*희
351	이*
352	이*비
353	이*아
354	이*재
355	이*재
356	이*지

3. 말뚱치 지식 강연회 후기 작성 행사

〈표 28〉 강연회 후기 작성 행사 참여자

번호	이름	주소
1	노*진	https://vo.la/myI4
2	강*화	https://vo.la/pTzV
3	W***yC**i	https://vo.la/R0No
4	강*민	https://vo.la/lmle
5	김*지	(당첨자 발표 후 삭제)
6	김*현	https://vo.la/2Xm3 (당첨자 발표 후 삭제)
7	김*연	https://vo.la/J2tP
8	박*민	https://vo.la/J9yS
9	박*배	https://vo.la/866C
10	이*림	https://vo.la/3WMZ
11	임*숙	https://vo.la/tW5e
12	장*덕	https://vo.la/LP7Z
13	정*태	https://vo.la/a04b
14	조*미	https://vo.la/ziZv
15	최*훈	https://vo.la/9fBs
16	김*아	https://vo.la/7eiw



〈그림 44〉 말뚱치 지식 강연회 후기 1



6월 18일 오후 6:12 · 🌐

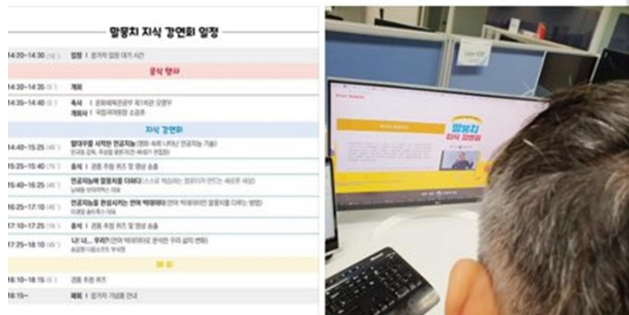
말뭉치 지식 강연회 후기 #말뭉치지식강연회

종합 선물세트같은 하루였다.

현장감은 좀 덜했지만, 자리에 편하게 앉아서 5인 5색의 강연을 몰입감 있게 볼 수 있었던 건 정말 행복한 경험이었다

모든 세션이 다 좋았지만, 송길영 다음소프트 부사장의 나! 너... 우리? 가 정말 인상 깊었는데, 코로나19 시대를 빅데이터로 명쾌히 분석해 주시고, 우리 삶의 변화를 짚어 주셨다.

많은 사람이 접속해서인지, 초중반까지 자꾸 튕기고 했던 게 옥의 티라면 티일까? 못 본 부분은 녹화영상으로 꼭 챙겨 봐야겠다.



3개

공유 1회



님이 사진첩에 사진 1장과 동영상 1개를 추가했습니다
니다: 말뭉치 지식 강연회.

6월 18일 오후 7:12 · 🌐

말뭉치?

컴퓨터가 말을 학습하는 시대

4차 산업 혁명 시대에 들어서면서 말뭉치는

기존 언어 연구 외에도 인공지능 기술 개발에 꼭 필요한 자원으로 관심이 모아지고 있다

인공 지능 로봇이나 스피커 등에 있는 음성 인식 기능은 이러한 언어 처리 기술을 바탕으로 하는데 말뭉치를 학습하는 방식으로 발전하고 있다.

한국-서울+도쿄=?

돈+돈=?

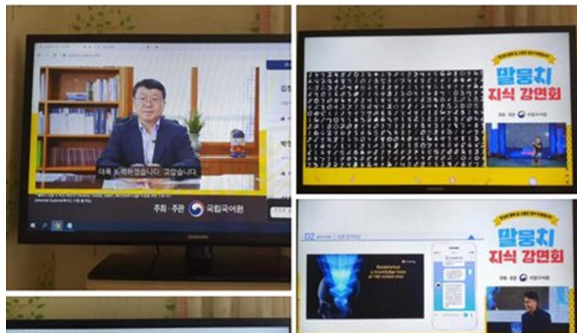
*말뭉치란? 언어학 용어인 코퍼스(corpus)를 우리말로 번역한 말. 말을 한데 모아 놓은 덩어리

#말뭉치지식강연회



6월 18일 오후 9:30 · 🌐

전계영 만화가님의 오디션과 좋아하면 울리는 이라는 작품을 정말 좋아하는데 퇴행성 관절염으로 인해 목소리 음성인식으로 그림을 그리신다는 걸 보고 깜짝 놀랐다. 대단하시다 라는 생각과 동시에 인공지능 기술이 이렇게 발달했다고? 하는 놀라움이 있었다. 태어나 영화관에서 처음 본 영화가 A.I.인데 어릴 때 아무 생각없이 봤던 게 점점 현실화 되어 가는 것 같아 살짝 무서웠다. 남세동 대표님 강연은 태어나 처음 들어보는 단어들 투성이라 솔직히 듣는데 좀... 답답... 워드투벡터... 지도학습 비지도학습... 이해가 될랑말랑 말랑카우요~ 이경일 대표님 강연은 이해하기 조금 더 쉬워서 재미있게 들었다. 쿵 심인데 쿵 나고 팔 심인데 팔 나는 건 사람이나 인공지능이나 마찬가지더라. 인공지능 어린이의 개짜증이 아직까지도 생생한ㅋㅋ 그리고 ADAMs의 엄청난 데이터랑! 정말 놀랍더라. 이따 사람 이겨먹는 인공지능도 진짜 곧 아닌가 싶다. 마지막으로 송길영 부사장의 강연은 빅 데이터 분석을 2020년 가장 큰 이슈인 코로나19를 키워드로 잡아 진행해주셔서 집중이 확 되었다. 너무 빨리 지나간 것 같은 게 아쉬웠다. 초반 서버 접속 불안정으로 인해 시청에 다소 불편함을 느꼈으나 담당자분들의 빠른 대처와 대응으로 끝까지 잘 시청할 수 있었다. 말뭉치에 대해 알게 되고 조금 더 친숙해질 수 있었던 시간이었고 조금 더 알고 싶은 마음이 들었던 시간이었다. 이러한 강연회가 더 자주 있으면 좋겠다. #말뭉치지식강연회

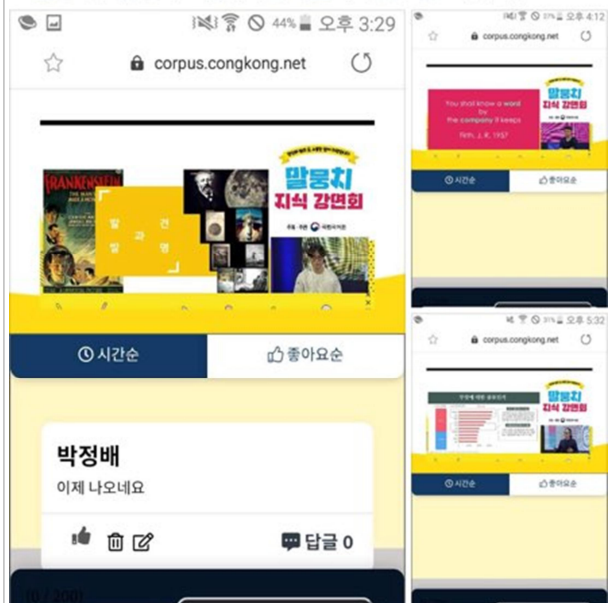


6월 21일 오후 4:15 · 🌐

말뭉치 지식 강연회를 온라인으로 신청해서 보았습니다.

접속이 힘들고 중간 중간 새로고침 현상으로 방송이 끊어져서 아쉬운 점이 있었지만 말뭉치 지식 강연회 내용은 좋았습니다. 말뭉치가 인공지능에 활용되는 것도 신기하고 다양한 언어 빅데이터를 활용한 방법들도 앞으로 더욱 기대되는 모습이었습니다.

#말뭉치지식강연회 #국립국어원 #말뭉치 #인공지능 #빅데이터



<그림 45> 말뭉치 지식 강연회 후기 2



좋아요 16개

#말뚱치지식강연회 #국립국어원 #세미나 #인공지능 #빅데이터 #말뚱치 #코퍼스 #언어정보 #민규동감독 #주성철 편집장 #남세동대표님 #이경일대표 #송길영다음소프트부사장 #코퍼스언어학 #음성코퍼스 #언어데이터 #AI #챗봇 #병렬말뚱치
오래 기다린 말뚱치 지식 강연회가 온라인으로 열렸다.
비행기에서 내려자마자 접속해서 이동하면서 계속 들었다.
네트워크 문제로 새로고침을 한 100번 하면서 스마트폰과 노트북으로 둘러가며 들었는데, 그럼에도 불구하고 유용하고 의미있는 강연들.
발표하신 분들 진짜 존경...!
할 너무 재밌음
#엄마가내전공을모르는데내전공은 #언어정보학 #뭐 공부하는지 비밀
아아입니다
6월 18일

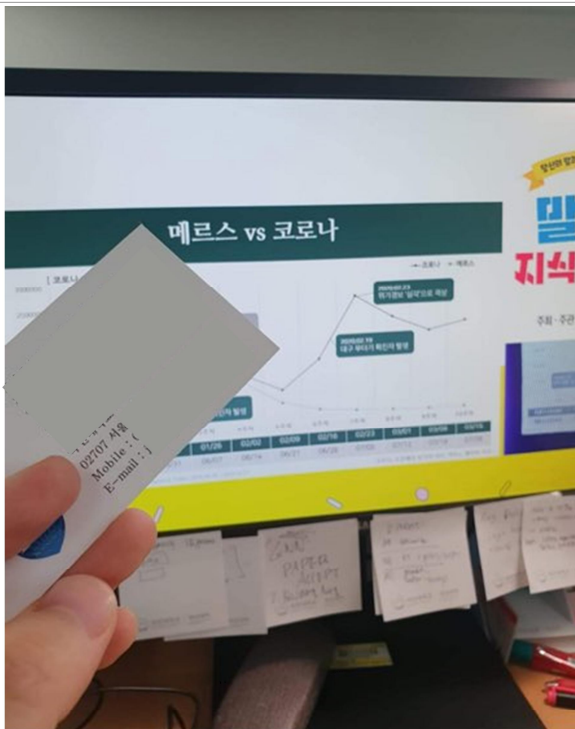


국립국어원지식강연회의 참관을 모두 마쳤다. 좋은 내용들이 많았지만 한가지 아쉬운 점은 시스템의 이상으로 질문을 할 수 없다는 점이였다. 다음에도 이와 같은 좋은 강의를 볼 수 있기를 기원한다.

#말뚱치지식강연회

타임라인 사진 · 6월 18일 ·

전체 크기 보기 · 옵션 더 보기



#말뚱치지식강연회

너무나 재밌게 자연어처리에 대해 리마인드 했던 시간 좋았습니다. 특히 송길영 부사장의 인사이드 도출로부터 변화에 대응하는 메커니즘에 대해 알게되며 한층 더 성장하는 모티브가 되었습니다.

휴대폰 업로드 · 6월 18일 ·

전체 크기 보기 · 옵션 더 보기



6월 19일 오후 5:41 ·

#말뚱치지식강연회

평소에 언어학이 관심이 많아 사전신청하고 여유롭게 시청하려 했던만 서버문제인지 기술문제인지 접속코드 넣고 계속 에러... 전 소리만 나오고 댓글창은 열리지도 않네요 페북 가보니 저만 그런건 아닌듯
2시간이 시작해서 접속이 잘 안되어서 사과리도 안되고 크롬도 안되고 포기하려던 찰나 4시 조금 넘어서야 다시 나오네요^^
그래도 퀴즈도 풀고 강연자분들의 심도있고 밀도 넘치는 강연 잘 보고 잘 들었습니다
예전 영화 말모이도 떠오르고 그러네요~
강연전에 배포된 안내책자 보면서 들으니까 이해도 잘 되고 쑥쑥 귀에 잘 들어와서 좋았답니다
인공지능으로 명령어만 넣어서 그림 그리는 기술은 참 신기하더라고요
제일 흥미가 돌아온 건
다음부사장 송길영님 강연이었는데 텅빈 마트를 바라보는 노인 사진은 소름이 돋더라고요!

이래저래 이벤트 개최하시느라 고생하신 국립국어원 관계자 여러분께 심심한 감사 인사 올립니다
자주 자주 해주시면 좋겠네요^^
제 어둡고 모자란 해안을 한껏 깨우치고 일깨워준 오랫동안의 좋은 강의였네요^^

#국립국어원 #말뚱치지식강연회 #참여행사 #댓글행사 #인증행사 #인공지능 #빅데이터 #말뚱치

<그림 46> 말뚱치 지식 강연회 후기 3

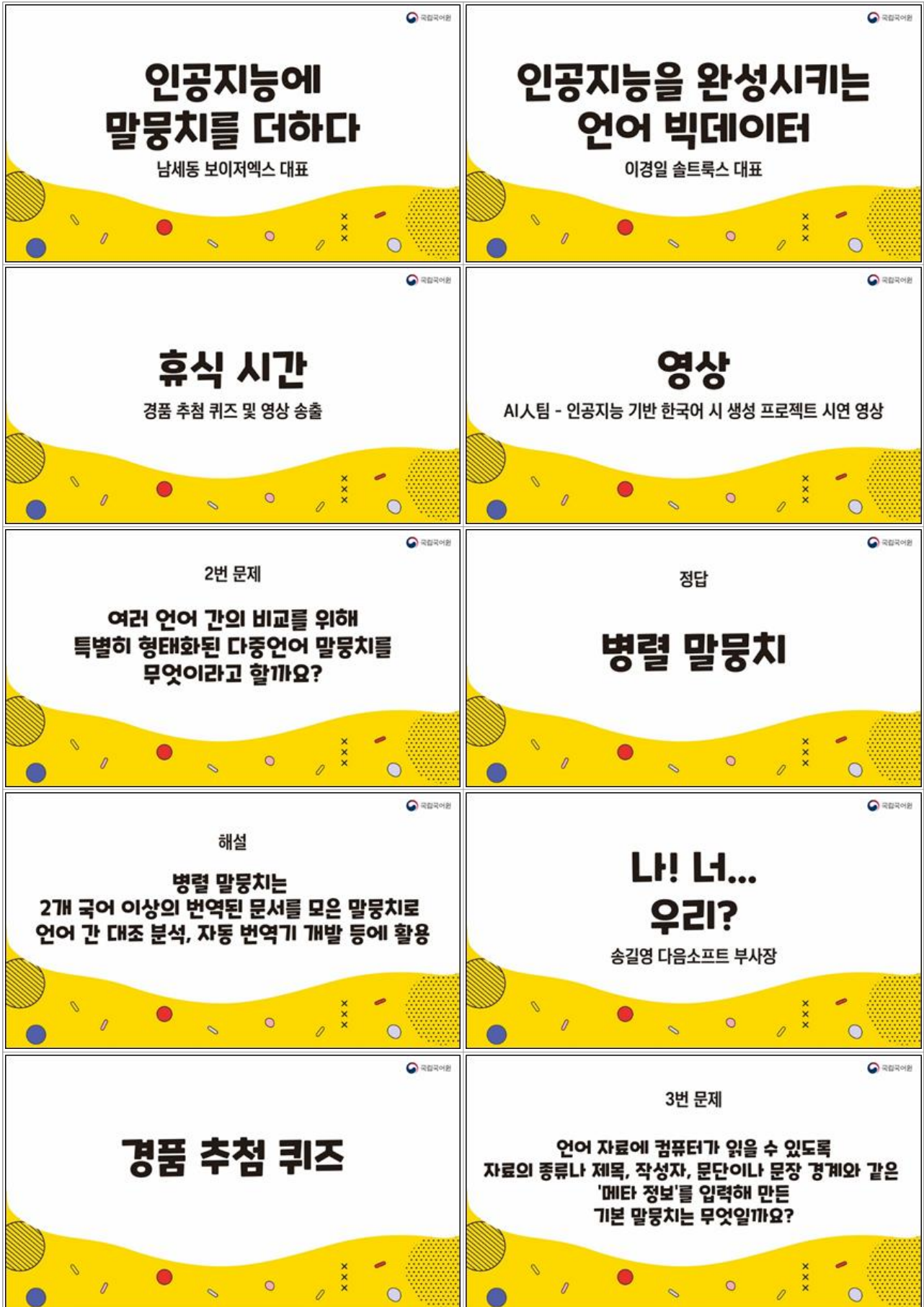
4. 말뚱치 지식 강연회 제작물



<그림 47> 말뚱치 지식 강연회 제작물 1



〈그림 48〉 말뭉치 지식 강연회 제작물 2



〈그림 49〉 말뭉치 지식 강연회 제작물 3



<그림 50> 말뭉치 지식 강연회 제작물 4

<Abstract>

Collecting Personal Writing Data and Holding a Public Lecture Meeting on a Corpus

This project, comprised mainly of the two divisible tasks as below, was initiated to extend the public awareness on large-scale and high-quality Korean language resources for the advancement of the artificial intelligence (hereinafter 'AI') industry. In a task for 'collecting personal writing data,' an opportunity for public participation was provided to collect writing data and to construct a non-publication corpus from writings that were never proofread professionally. In the other task regarding 'Public Lecture on Corpus,' a lecture event on language-based AI was held for the general public. This kind of public participation-style project for the construction and application of the corpus aimed to stimulate public interests in the computerization of the Korean language and to contribute to the establishment of research and industrial environments that would utilize Korean language resources.

The task of collecting personal writing data focused on 'collecting writing data from various generations' and 'securing personal writing data that were never professionally proofread.' The project proceeded in the following order: a composition of writing topics, a detailed planning of data collection methods, an organization and operation of the promotion committee, public relations efforts, a collection of writing data, securing copyright, and making a list of collected works and information.

To collect a good amount of writing data time-effectively, an event titled 'A Piece of Writing, Please' was held, and a web site was mostly used for data collection. Through electronic contracts, the use of copyright was dealt with,

and participants' writings were received. The participants were asked to classify their writings into the following categories: poetry including children's poetry, diary, letter, novel including fairy tale, review, and 'etc.' as an extra category. The submission was limited to previously unposted and unpublished writings, neither online nor offline. The promotion of the 'A Piece of Writing, Please' event was done around internet communities, libraries, community center bulletin boards, and bus information/management systems, both online and offline. In consideration of the computer-illiterate, participants were given options to submit offline as well.

A total of 6,044 people from elementary schoolers to the elderly of age 60 or older participated in the project to collect 10,919 articles and 9,250,097 characters in total. One participant was allowed to submit a maximum of 10 articles and a minimum of 1,000 characters. 800 people aged 7 to 12, 497 people aged 13 to 15, 1,058 people aged 16 to 18, 2,720 people aged 19 to 29, 648 people aged 30 to 59, and 321 people aged 60 or over participated in the process to ensure generational diversity in the writing data. The collected writings were revised to comply with the corpus format of the National Institute of Korean Language (JavaScript Object Notation or 'JSON') to set up a non-publication raw corpus.

For the Public Lecture on Corpus task, a lecture event was planned with the topic, "The Value of Language Resources in the Era of AI and the Status Quo of Natural Language Processing" to present prospects on the usage of Korean language resources. Although it had originally been designed to be offered offline in February 2020, due to coronavirus circumstances, it was changed into an online event offered on Thursday, June 18, 2020. In order to host the event successfully, various online and offline promotions were carried out, such as public relations video transmission, card news distribution, etc.

A total of 652 people originally signed up for the online lecture event, and 524 of them eventually participated live online. The five speakers who delivered four lectures were Director MIN Kyu-dong and Commentator JU SUNGCHUL as a team, Representative Sedong Nam, Representative Tony Lee(Kyungil Lee), and Vice President Gilyoung Song. Around 120 to 130 participants attended per lecture. To pursue a colorful event, during breaks, an exhibition video was shown and a giveaway quiz event was offered. An evaluation was done by participants at the end of the lectures to understand their responses. Many responded positively mentioning that the event was well-structured and that what could be difficult content was well-delivered.

A while after the lecture event, another satisfaction survey was conducted to design a better public lecture event in the future. According to its result, more than half of the respondents (191 out of 253), showed high satisfaction levels regarding the lecture event. About 44% of the survey respondents answered that they had the highest expectations on lectures from 'Director MIN Kyu-dong and Commentator JU SUNGCHUL,' and 'Vice President Gilyoung Song,' while about 48% of the survey respondents answered the most helpful lecture for their future business or studies turned out to be that of 'Representative Tony Lee(Kyungil Lee).' Such results may reflect that, from the lectures, the participators rather prefer professional content to what is easy-to-understand from their perspectives. Therefore, it is possible to conclude that the public lecture in the future may be planned to deliver content of a more professional sort.

Keywords: personal writing data, non-publication corpus, Public Lecture on
Corpus, artificial intelligence (or AI)
Project Director: Na Yun-jung(Metrix Corporation)

사업 책임자	나윤정((주)메트릭스코퍼레이션 리서치부문 부사장)
사업 참여자	박두진((주)메트릭스코퍼레이션 리서치부문 이사)
	민지현((주)메트릭스코퍼레이션 리서치부문 차장)
	김홍건((주)메트릭스코퍼레이션 리서치부문 부장)
	김기현((주)메트릭스코퍼레이션 리서치부문 이사)
	조일상((주)메트릭스코퍼레이션 리서치부문 대표)
	조신애((주)메트릭스코퍼레이션 리서치부문 부장)
	남궁건((주)메트릭스코퍼레이션 리서치부문 부장)
	고유미((주)메트릭스코퍼레이션 리서치부문 차장)
	박준현((주)메트릭스코퍼레이션 리서치부문 대리)
	김규일((주)메트릭스코퍼레이션 리서치부문 과장)
	이경화((주)메트릭스코퍼레이션 리서치부문 대리)
	윤지은((주)메트릭스코퍼레이션 리서치부문 과장)
	하지영((주)메트릭스코퍼레이션 리서치부문 팀장)
	채윤철((주)메트릭스코퍼레이션 리서치부문 팀장)
	유혜영((주)메트릭스코퍼레이션 리서치부문 과장)
	임상호((주)메트릭스코퍼레이션 리서치부문 과장)
	김수연((주)메트릭스코퍼레이션 리서치부문 사원)
	김종서((주)메트릭스코퍼레이션 리서치부문 대리)
	김은영((주)메트릭스코퍼레이션 리서치부문 주임)
	이태연((주)메트릭스코퍼레이션 리서치부문 사원)
	한이슬((주)메트릭스코퍼레이션 리서치부문 대리)
	박지유((주)메트릭스코퍼레이션 리서치부문 대리)
	권혁윤((주)메트릭스코퍼레이션 리서치부문 대리)
	김다정((주)메트릭스코퍼레이션 리서치부문 대리)
	김소현((주)메트릭스코퍼레이션 리서치부문 대리)
	곽성호((주)메트릭스코퍼레이션 리서치부문 대리)
	유영석((주)레인보우커뮤니케이션 대표)
	이진철((주)레인보우커뮤니케이션 실장)
	나진수((주)레인보우커뮤니케이션 부실장)

이종진((주)레인보우커뮤니케이션 팀장)
 이화복((주)레인보우커뮤니케이션 팀장)
 황성하((주)레인보우커뮤니케이션 대리)
 정선호((주)레인보우커뮤니케이션 대리)
 최세희((주)레인보우커뮤니케이션 대리)
 김소윤((주)레인보우커뮤니케이션 사원)
 이다연((주)레인보우커뮤니케이션 사원)

담당 연구원 이승재(국립국어원 언어정보과장)
 서셋별(국립국어원 언어정보과 학예연구사)

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9727

인쇄일: 2020년 6월 30일

발행일: 2020년 6월 30일

인 쇄: 씨오엠

※ 이 책은 국립국어원의 용역비로 수행한 ‘개인적 글쓰기 자료 수집 및 말뭉치 지식 강연회 개최’ 사업의 결과물을 발간한 것입니다.