

| 메타데이터<br>정보<br>(다중기입가능) | 분야                   | 데이터 유형 <sup>1)</sup>  | 구축 데이터량  | 원천데이터 형식 <sup>2)</sup> | 라벨링 형식 <sup>3)</sup> | 라벨링<br>유형 <sup>4)</sup> |
|-------------------------|----------------------|---|----------|------------------------|----------------------|-------------------------|
|                         | 한국어                  | 오디오   | 4,000시간  | wav                    | json, csv            | 전사(음성)                  |
|                         | 데이터 출처 <sup>5)</sup> | 데이터 구축년도  | 구축기관(총괄) | 가공기관                   | 검수기관                 |                         |
|                         | 자체 수집                | 2021년   | 씨에스리     | 드림비트, 디그랩              | 씨에스리                 |                         |
|                         | 데이터<br>문의처           | 기관명   | 문의 담당자명  | 전화번호<br>(유선전화번호기입)     | 메일주소                 |                         |
|                         |                      | 씨에스리  | 윤희우      | 070-4756-4580          | hwyoon@cslee.co.kr   |                         |
|                         | 데이터 소개               | 외국어가 모국어인 화자의 한국어 음성 데이터로, 주어진 대본 내용을 읽은 음성과 주어진 질문에 자유롭게 답변한 음성을 내용으로 포함 |          |                        |                      |                         |
|                         | 주요키워드                | 한국어 음성 인식, 외국인, 이주민, 한국어, 한국어 학습, 음성 전사                                   |          |                        |                      |                         |
| 카테고리 정의서                |                      | 카테고리 정의서 엑셀 파일에 데이터 카테고리 작성하여 제출  |          |                        |                      |                         |

1) 텍스트, 오디오, 이미지, 비디오,

2) txt, jpg,.....

3) json, csv,.....

4) 내용요약(텍스트), 번역(자연어), 질의응답(자연어), 바운딩박스(이미지/동영상), 키포인트(이미지/동영상), 세그멘테이션(이미지/동영상), 전자(음성) .....

5) 4대 언론기사, 자체 수집,.....

| 데이터셋명  | 국문영문   | 인공지능 학습을 위한 외국인 한국어 발화 음성 데이터  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
|--|--|--|----------|-----------|----------|-----------|---------|-----------|--------|---------|-------|--------|---------|-----|----------|--------------|-------|--------|---------|--------|----------|-------|-------|----------|--------|-------|--------|-----|-------|--------|---------|--------|---------|-----|--------|----------|---------|--------|---------|-----|-------|--------|---------|--------|---------|----|--------|----------|---------|--------|---------|----|--|--------|----------|-----------|----------|-----------|
|  | 문  | Establishing Foreign Speakers' Voice Data of Korean for Artificial Intelligence Learning |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 구축목적   | 외국어가 모국어인 화자의 한국어 음성 발화 내용을 인식할 수 있도록 인공지능을 훈련하기 위한 데이터셋   |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 활용서비스  | 외국인 화자의 한국어 음성을 정확하게 인식할 수 있는 인공지능 스피커 (예시)  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 소개   | 외국어를 모국어로 하는 화자의 음성을 토대로 전사 작업을 진행한 AI 데이터셋이다. 외국인 화자는 모국어의 영향으로 인해 한국어 발화 시 한국어 원어민과는 다른 발화 패턴을 보인다. 이는 한국에서 장기간 거주한 외국인에게서도 보이는 현상이다. 이에 해당 데이터셋은 대본 및 질문 내용 구성 시 외국인 발화 오류 패턴을 고려하여 외국인의 한국어 발화 오류를 의도적으로 유도한 음성 데이터를 수집하였으며, 이에 대한 전사 데이터를 구축하였다.  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 데이터셋 통계<br>(구축 규모 및 분포)  | 1. 데이터 구축 규모   |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
|  | <table><tr><th rowspan="2">데이터 종류</th><th rowspan="2">데이터 형태</th><th colspan="3">원천데이터 규모</th><th colspan="2">메타 데이터 규모</th></tr><tr><th>사이즈(GB)</th><th>시간(h)</th><th>건수</th><th>사이즈(MB)</th><th>건수</th></tr><tr><td>베트남어</td><td rowspan="6">녹음음성 및 전사 결과</td><td>82.07</td><td>764.93</td><td>194,082</td><td>343.25</td><td>388,164</td></tr><tr><td>영어</td><td>13.30</td><td>123.90</td><td>31,554</td><td>55.51</td><td>63,108</td></tr><tr><td>일본어</td><td>81.55</td><td>760.08</td><td>216,902</td><td>379.69</td><td>433,804</td></tr><tr><td>중국어</td><td>114.23</td><td>1,064.66</td><td>292,593</td><td>516.53</td><td>585,186</td></tr><tr><td>태국어</td><td>57.70</td><td>537.82</td><td>139,692</td><td>244.89</td><td>279,384</td></tr><tr><td>기타</td><td>112.79</td><td>1,051.00</td><td>284,842</td><td>501.98</td><td>569,684</td></tr><tr><td colspan="2">총계</td><td>461.64</td><td>4,302.39</td><td>1,159,665</td><td>2,041.85</td><td>2,319,330</td></tr></table> |  | 데이터 종류   | 데이터 형태    | 원천데이터 규모 |           |         | 메타 데이터 규모 |        | 사이즈(GB) | 시간(h) | 건수     | 사이즈(MB) | 건수  | 베트남어     | 녹음음성 및 전사 결과 | 82.07 | 764.93 | 194,082 | 343.25 | 388,164  | 영어    | 13.30 | 123.90   | 31,554 | 55.51 | 63,108 | 일본어 | 81.55 | 760.08 | 216,902 | 379.69 | 433,804 | 중국어 | 114.23 | 1,064.66 | 292,593 | 516.53 | 585,186 | 태국어 | 57.70 | 537.82 | 139,692 | 244.89 | 279,384 | 기타 | 112.79 | 1,051.00 | 284,842 | 501.98 | 569,684 | 총계 |  | 461.64 | 4,302.39 | 1,159,665 | 2,041.85 | 2,319,330 |
|  | 데이터 종류   | 데이터 형태   |          |           | 원천데이터 규모 |           |         | 메타 데이터 규모 |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
|  |  |  | 사이즈(GB)  | 시간(h)     | 건수       | 사이즈(MB)   | 건수      |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
|  | 베트남어   | 녹음음성 및 전사 결과   | 82.07    | 764.93    | 194,082  | 343.25    | 388,164 |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
|  | 영어   |  | 13.30    | 123.90    | 31,554   | 55.51     | 63,108  |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
|  | 일본어  |  | 81.55    | 760.08    | 216,902  | 379.69    | 433,804 |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
|  | 중국어  |  | 114.23   | 1,064.66  | 292,593  | 516.53    | 585,186 |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
|  | 태국어  |  | 57.70    | 537.82    | 139,692  | 244.89    | 279,384 |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
|  | 기타   |  | 112.79   | 1,051.00  | 284,842  | 501.98    | 569,684 |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 총계   |  | 461.64   | 4,302.39 | 1,159,665 | 2,041.85 | 2,319,330 |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 2. 데이터 분포  |  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| - 언어분류별 분포 : 베트남어, 영어, 일본어, 중국어, 태국어, 기타 음성 녹음 시간  |  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| - 세트별 분포 : 한국일반, 한국생활I, 한국생활II, 한국문화I, 한국문화II 음성 녹음 시간   |  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| - 유형별 분포 : 대본읽기, 질문에 답변하기 음성 녹음 시간   |  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 2-1. 언어분류별 분포  |  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| <table><tr><th>언어분류</th><th>녹음 시간</th><th>비율(%)</th></tr><tr><td>베트남어</td><td>764.93</td><td>17.78</td></tr><tr><td>영어</td><td>123.90</td><td>2.88</td></tr><tr><td>일본어</td><td>760.08</td><td>17.67</td></tr><tr><td>중국어</td><td>1,064.66</td><td>24.75</td></tr><tr><td>태국어</td><td>537.82</td><td>12.50</td></tr><tr><td>기타</td><td>1,051.00</td><td>24.43</td></tr><tr><td>총계</td><td>4,302.39</td><td>100</td></tr></table> |  | 언어분류   | 녹음 시간    | 비율(%)     | 베트남어     | 764.93    | 17.78   | 영어        | 123.90 | 2.88    | 일본어   | 760.08 | 17.67   | 중국어 | 1,064.66 | 24.75        | 태국어   | 537.82 | 12.50   | 기타     | 1,051.00 | 24.43 | 총계    | 4,302.39 | 100    |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 언어분류   | 녹음 시간  | 비율(%)  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 베트남어   | 764.93   | 17.78  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 영어   | 123.90   | 2.88   |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 일본어  | 760.08   | 17.67  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 중국어  | 1,064.66   | 24.75  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 태국어  | 537.82   | 12.50  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 기타   | 1,051.00   | 24.43  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 총계   | 4,302.39   | 100  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| <div>언어분류별 녹음 시간</div> <table><tr><th>언어분류</th><th>녹음 시간</th></tr><tr><td>베트남어</td><td>765</td></tr><tr><td>영어</td><td>124</td></tr><tr><td>일본어</td><td>760</td></tr><tr><td>중국어</td><td>1,065</td></tr><tr><td>태국어</td><td>538</td></tr><tr><td>기타</td><td>1,051</td></tr></table>  |  | 언어분류   | 녹음 시간    | 베트남어      | 765      | 영어        | 124     | 일본어       | 760    | 중국어     | 1,065 | 태국어    | 538     | 기타  | 1,051    |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 언어분류   | 녹음 시간  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 베트남어   | 765  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 영어   | 124  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 일본어  | 760  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 중국어  | 1,065  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 태국어  | 538  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |
| 기타   | 1,051  |  |          |           |          |           |         |           |        |         |       |        |         |     |          |              |       |        |         |        |          |       |       |          |        |       |        |     |       |        |         |        |         |     |        |          |         |        |         |     |       |        |         |        |         |    |        |          |         |        |         |    |  |        |          |           |          |           |

## 2-2. 세트별 분포

| 세트     | 녹음 시간    | 비율(%) |
|--------|----------|-------|
| 한국일반   | 947.73   | 22.03 |
| 한국생활I  | 836.76   | 19.45 |
| 한국생활II | 895.77   | 20.82 |
| 한국문화I  | 827.47   | 19.23 |
| 한국문화II | 794.67   | 18.47 |
| 총계     | 4,302.39 | 100   |



## 2-3. 유형별 분포

| 세트       | 녹음 시간    | 비율(%) |
|----------|----------|-------|
| 대본 읽기    | 3,101.09 | 72.08 |
| 질문에 답변하기 | 1,201.30 | 27.92 |
| 총계       | 4,302.39 | 100   |



## 1. 대표도면

### 데이터셋 구성

| 구분           | 내용   |
|--------------|--|
| 음성           | ▶ 0:00 / 0:19                                    |
| 파일명          | VN11RC001_VN0001_20211124.wav                    |
| 문장ID         | VN11RC001  |
| 작업자ID        | VN0001   |
| 녹음날짜         | 2021-11-24                                       |
| 언어분류         | 베트남어   |
| 유형           | 대본 읽기  |
| 대본 내용        | 잠시만, 내 지갑이랑 열쇠가 어디 갔지? 유람아, 미안한데 나 물건이 없어진 거 같아. |
| 전사 결과        | 잠시만 내 지갑이랑 열쇠가 어디 갔지 유람아 미안한데 나 물건이 없어진 거 같아     |
| 한국어<br>구사 능력 | 상  |

## 2. 라벨링데이터 구성

| NO  | 항목               |            | 타입     | 길이 |      | 단위 | 필수여부 | 유효값  |
|-----|------------------|------------|--------|----|------|----|------|--|
|     | 항목명              | 항목 설명      |        | 최소 | 최대   |    |      |  |
| 1   | fileName         | 파일명        | string | 29 | 29   | -  | Y    | -  |
| 2   | file_info        | 녹음파일 정보    | Object | -  | -    | -  | Y    | -  |
| 2-1 | speakerID        | 녹음자ID      | string | 6  | 6    | -  | Y    | -  |
| 2-2 | sentenceID       | 문장ID       | string | 9  | 9    | -  | Y    | -  |
| 2-3 | recordUnit       | 녹음단말기 유형   | string | 3  | 7    | -  | Y    | android, ios                                   |
| 2-4 | recordQuality    | 녹음 품질      | string | 16 | 16   | -  | Y    | 16bit 16KHz MONO                               |
| 2-5 | recordDate       | 녹음날짜       | string | 19 | 19   | -  | Y    | -  |
| 2-6 | recordTime       | 녹음 시간      | string | 1  | 7    | 초  | Y    | -  |
| 3   | transcription    | 대본/질문 정보   | Object | -  | -    | -  | Y    | -  |
| 3-1 | Reading          | 대본 내용      | string | 0  | 100  | -  | N    | -  |
| 3-2 | ReadingLabelText | 대본읽기 전사 결과 | string | 0  | 200  | -  | N    | -  |
| 3-3 | Question         | 질문 내용      | string | 0  | 150  | -  | N    | -  |
| 3-4 | AnswerLabelText  | 질문답변 전사 결과 | string | 0  | 1300 | -  | N    | -  |
| 3-5 | SentenceSpeechLV | 문장 구사 실력   | string | 1  | 1    | -  | Y    | 상, 중, 하  |
| 4   | SpeakerID        | 녹음자ID      | string | 6  | 6    | -  | Y    | -  |
| 5   | basic_info       | 녹음자 기본 정보  | Object | -  | -    | -  | Y    | -  |
| 5-1 | gender           | 성별         | string | 1  | 1    | -  | Y    | M, F   |
| 5-2 | birthYear        | 출생연도       | string | 4  | 4    | 년  | Y    | -  |
| 5-3 | eduBackground    | 최종학력       | string | 2  | 4    | -  | Y    | 초졸, 중졸, 고졸, 대졸, 석사이상                           |
| 6   | residence_info   | 녹음자 거주 정보  | Object | -  | -    | -  | Y    | -  |
| 6-1 | country          | 국적         | string | 2  | 2    | -  | Y    | -  |
| 6-2 | residencePeriod  | 국내 체류 기간   | string | 5  | 11   | -  | Y    | 1년 미만, 1년 이상 3년 미만, 3년 이상 5년 미만, 5년 이상         |
| 6-3 | residenceCity    | 거주 지역      | string | 4  | 6    | -  | Y    | -  |
| 7   | skill_info       | 녹음자 실력 정보  | Object | -  | -    | -  | Y    | -  |
| 7-1 | languageClass    | 언어 분류      | string | 2  | 4    | -  | Y    | 베트남어, 영어, 일본어, 중국어, 태국어, 기타                    |
| 7-2 | motherTongue     | 모국어        | string | 2  | 9    | -  | Y    | -  |
| 7-3 | selfAssessment   | 자가평가실력     | string | 1  | 1    | -  | Y    | 상, 중, 하  |
| 7-4 | topikGrade       | TOPIK 등급   | string | 1  | 5    | -  | Y    | 1, 2, 3, 4, 5, 6, 해당없음                         |
| 7-5 | LearningPeriod   | 한국어 학습 기간  | string | 1  | 3    | 개월 | Y    | -  |
| 7-6 | learningSource   | 한국어 학습 방법  | string | 2  | 13   | -  | Y    | 학교수업(대학 포함), 학원, 인터넷 강의, 서적, 한국인 친구, 영화/음악, 기타 |

### 3. 라벨링데이터 실제예시

```
{
  "fileName": "VN11RC001_VN0001_20211124.wav",
  "file_info": {
    "speakerID": "VN0001",
    "sentenceID": "VN11RC001",
    "recordUnit": "ios",
    "recordQuality": "16bit 16kHz MONO",
    "recordDate": "2021-11-24 12:01:01",
    "recordTime": "9.316"
  },
  "transcription": {
    "Reading": "잠시만, 내 지갑이랑 열쇠가 어디 갔지? 유람아, 미안한데 나 물건이 없어진 거 같아.",
    "ReadingLabelText": "잠시만 내 지갑이랑 열쇠가 어디 갔지 유람아 미안한데 나 물건이 없어진 거 같아",
    "Question": "",
    "AnswerLabelText": "",
    "SentenceSpeechLV": "상"
  },
  "SpeakerID": "VN0001",
  "basic_info": {
    "gender": "F",
    "birthYear": "1995",
    "eduBackground": "고졸"
  },
  "residence_info": {
    "country": "VN",
    "residencePeriod": "5년 이상",
    "residenceCity": "KR-41"
  },
  "skill_info": {
    "languageClass": "베트남어",
    "motherTongue": "베트남어",
    "selfAssessment": "중",
    "topikGrade": "4",
    "LearningPeriod": "36",
    "learningSource": "기타"
  }
}
```

| 데이터셋 구축<br>수행기관<br>담당자 | 주관기관 | 기관명     | 책임자명             | 전화번호<br>(유선전화번호기입) | 메일주소               | 담당업무              |
|------------------------|------|---------|------------------|--------------------|--------------------|-------------------|
|                        |      | 씨에스리    | 윤희우              | 010-5645-9912      | hwoyon@cslee.co.kr | 사업관리 및<br>전체 공정관리 |
|                        | 참여기관 | 기관명     | 담당업무             | 기관명                | 담당업무               |                   |
|                        |      | 세종대학교   | 데이터 설계<br>수집 관리  | 디그랩                | 데이터 가공             |                   |
|                        |      | 이화여자대학교 | 데이터 설계<br>수집 관리  | 액션파워               | 학습모델 구현            |                   |
|                        |      | 드림비트    | 데이터 정제<br>데이터 가공 |                    |                    |                   |