**[방송자막시스템 POC]**

|  |  |  |
| --- | --- | --- |
|  | **제 안 요 청 서** |  |

**2022.09**

**㈜케이티스카이라이프**

**목   차**

1. 사업개요 3

1.1 사업명

1.2 추진목적

1.3 POC과제 제안 요청 사항

2. 제안 일반조건 3

3. 제안서 작성내용 3

**1. 사업개요**

**1.1 사업명**

○ 음성인식 학습용 데이터 가공 제안

○ 미디어 콘텐츠 전용 자막시스템 검증 데이터 제안

**1. 2 추진목적**

○ 당사에서 제공 하는 오디오 파일(1시간 단위)과 자막 파일을 10~15초 간격의 문장으로 정확하게 분리 및 매칭하는 변환

○ 미디어 콘텐츠 전용 자막 시스템 구축 후 인식률 확인용 전사 데이터 구축

**1. 3 POC과제** **제안 요청 사항**

○ 음성인식 학습용 데이터 가공

* 데이터 정제 프로그램(알고리즘) 개발

① 당사 에서 제공 하는 음성과 자막 데이터를 문장 단위로 분리 하고 해당 하는 자막을 정확하게 일치 하도록 매칭 시키는 알고리즘 개발 및 전체 데이터 가공 수행

1. 문장 단위로 음성 데이터를 분리(1~30초)

2. 분리된 문장에 해당 하는 자막의 제공 자막을 일치

3. 정확하게 일치하는 데이터만 정제(최소 50% 이상 변환)

4. 숫자 표기는 한글로 변경 또는 유지

5. 영문 표기는 한글로 변경 또는 유지

※ 제안 가점사항

1. 전체 데이터 중 변환된 데이터 비중이 높은 경우(최소 50%)

2. 화자를 분리하여 별도 구분

3. 숫자/외국어를 한글로 변환 하여 통일 하는 작업

4. 대화형 답변을 분리(예시 : 질문, 답변)

② 산출물 형태

1. 음성 학습용 데이터(WAV+TEXT 매칭) 제공

2. (화자를 구분 하여)1~30초 내외 길이 문장 단위로 저장

3. 전체 데이터에서 활용된 데이터 수량 별도 체크

③ 데이터 포맷

|  |  |
| --- | --- |
| File Type | WAV |
| File Type Extension | WAV |
| MIME Type | audio/x-wav |
| Encoding | Microsoft PCM |
| Num Channels | 2 |
| Sample Rate | 48000 |
| Avg Bytes Per Sec | 192000 |
| Bits Per Sample | 16 |
| Captions | text |

④ 데이터 정제 대상 : 총 210TB 기준(HDD 70개 내외로 제공, 2TB ~ 8TB)

○ 미디어 콘텐츠 전용 자막 시스템 검증 데이터 제안

* 음성 인식률 확인 전사 데이터 작업(데이터는 당사 제공)
* 수량 : 총 40시간(뉴스/드라마/예능/유트브 각 10시간
* 산출물 규격 : ETRI 전사 규격

○ H/W 등 기타 제안

* 본 POC과제에 필요한 H/W부분 까지 입찰사가 포함하여 제안
* IDC/클라우드 시스템 선택하여 이용 가능. 단, 해상서버는 POC전용으로만 별도로 구성
* 반드시 당사에서 제공한 데이터가 유출 되지 않도록 시스템을 구성
* 당사 데이터는 HDD로(70여개) 제공 되며 산출물을 별도 저장 공간에 제출 한다.
* 방식 제안 사항 ex)NAS스토리지)
* 데이터 정제 기술 확인을 위해 총 30분의 샘플 데이터 제공 후 정제 데이터 제공(샘플 제공 후 3시간 이내 제출)

**2. 제안일반조건**

○ 제안 일시 : 입찰공고 참조

○ 추진 기간 : 계약일로부터 9개월 예정

○ 업체 선정 방식 : 입찰공고 참조(제한 종합평가 경쟁입찰)

○ 제안/선정 일정 : 입찰공고 참조

※기타 문의사항은 담당자에 전화 또는 이메일 문의 가능

- 제안서 문의 : CT TF팀 임민우 010-8916-0957, [chansdad@skylife.co.kr](mailto:chansdad@skylife.co.kr)

- 입찰문의 : scm팀 박춘규 010-7914-1288, ppbros@skylife.co.kr

**3. 제안서 작성 내용**

○ 과업 개요

* 과업에 대한 이해(목적, 범위 기대 효과 등)

○ 제안사 소개

* 일반현황, 재무현황, 관련 수행실적 등

○ 수행 계획

* 정제된 데이터/정제 방법 등 기술적인 수행 설명
* 데이터 유출 방지 방안, 하드웨어 시스템 구성
* 프로젝트 수행 계획(기간/인원/수행 방법 등)

○ 기타

* 프로젝트 종료 후 유상/무상 지원 계획
* 기타 특별 제안 등