|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **캡스톤디자인 주간보고서 및 회의록** | | **\_9\_차** | 제출일 : 2022년 05월 27일 | |
| **학기** | **2022년 1학기** | **1분반** | **1학기** | |
| **주제** | **한국어의 특성을 반영한 자연어 처리 성능 향상** | | | |
| **지도교수** | **최 희 열 (서명)** | **참여기업(기관)** | | **AITRICS** |
| **참석자** | **허재무, 김정희, 김주환** | | | |
| **[ 13 주차] 캡스톤디자인 진행 상황 주간 보고 내용** | | | | |
| **회의시간** | 2022년 05월 27일 18시 00분 ~ 19시 00분 | **현재 완성도** | | 90 % / 100 % |
| **주요 의제** | **1차 산업체 미팅 피드백 반영 및 반영 결과에 대한 피드백** | | | |
| **이전**  **피드백** | 명사형 존대 어휘(진지, 말씀 등)또한 변환할 수 있는 기능 제공할 것. | | | |
| **금주**  **수행**  **내용**  **요약** | \* 각각의 세부내용은 다음 페이지부터 기술할 것. (분량 무관)   * 자모 단위 BPE가 Vocab size를 줄일 수 있는 확실한 근거 탐색 * 명사형 존대 어휘(진지, 말씀) 등을 치환할 수 있는 기능 구현. | | | |
| **새로운**  **피드백** | 전처리 과정에 멀티 프로세싱을 반영하여 처리 속도를 증가시키는 방법을 찾아볼 것. | | | |
| **차주**  **수행**  **계획** | * 멀티 프로세싱을 전처리 과정에 적용하고, 명사형 존대 어휘를 적용하여 train을 해 볼것. | | | |

**주간 수행 내용 (세부)**

\* 단, 기업체 인사가 참여한 멘토링 회의 시 반드시 회의 장면(사진, 캡쳐 등)이 첨부되어야 함.

1. **BPE Vocab size의 감소**

**음절**

**dictionary = {**

**'학 교 </w>' : 5,**

**'한 교 </w>' : 2,**

**'함 교 </w>' : 6,**

**'할 교 </w>' : 3,**

**'곡 교 </w>' : 1,**

**'곤 교 </w>' : 3,**

**'골 교 </w>' : 4,**

**'곰 교 </w>' : 9**

**}**

**자모**

**dictionary = {**

**'ㅎ ㅏ ㄱ ㄱ ㅛ \_ </w>' : 5,**

**'ㅎ ㅏ ㄴ ㄱ ㅛ \_ </w>' : 2,**

**'ㅎ ㅏ ㅁ ㄱ ㅛ \_ </w>' : 6,**

**'ㅎ ㅏ ㄹ ㄱ ㅛ \_ </w>' : 3,**

**'ㄱ ㅗ ㄱ ㄱ ㅛ \_ </w>' : 1,**

**'ㄱ ㅗ ㄴ ㄱ ㅛ \_ </w>' : 3,**

**'ㄱ ㅗ ㄹ ㄱ ㅛ \_ </w>' : 4,**

**'ㄱ ㅗ ㅁ ㄱ ㅛ \_ </w>' : 9**

**}**

**('학교')**

**('한', '교')**

**('함교')**

**('할', '교')**

**('곡', '교')**

**('곤', '교')**

**('골교')**

**('곰교')**

**('ㅎㅏ', 'ㄱ', 'ㄱㅛ\_')**

**('ㅎㅏ', 'ㄴ', 'ㄱㅛ\_')**

**('ㅎㅏ', 'ㅁ', 'ㄱㅛ\_')**

**('ㅎㅏ', 'ㄹ', 'ㄱㅛ\_')**

**('ㄱㅗ', 'ㄱ', 'ㄱㅛ\_')**

**('ㄱㅗ', 'ㄴ', 'ㄱㅛ\_')**

**('ㄱㅗ', 'ㄹ', 'ㄱㅛ\_')**

**('ㄱㅗ', 'ㅁ', 'ㄱㅛ\_')**

**'학교'**

**'한교'**

**'함교'’**

**'할교'’**

**'곡교'’**

**'곤교'’**

**'골교'’**

**'곰교'’**

**음절**

**자모**

**학교, 한, 교, 함교, 할, 곡, 곤, 곰교, 골교**

**ㅎㅏ, ㄱ, ㄱㅛ\_, ㄴ, ㅁ, ㄹ, ㄱㅗ**

**9개**

**7개**

**예시에서 보이는 것 처럼 자모단위로 자를 경우 같은 수의 단어에 대하여 더 적은 단위들로 이를 구성할 수 있다.**

1. **존대형 어휘 적용**

****

**예시에서 알 수 있는 것 처럼, 존대 어휘를 적용했음.**

**‘말씀’과 같은 어휘의 경우, 간단하게 이를 ‘말’로 치환하는 것으로 해결이 가능하나, ‘진지’의 경우, ‘포진지’의 ‘진지’, ‘그는 매사에 진지하다’ 의 진지 등 다른 의미로 사용될 수 있다. 때문에 이런 어휘들에 대하여 패턴을 분석을 한다.**

**‘진지’의 경우, 첫째, 단독으로 사용되는자, 둘째, ‘장소’를 나타내는 조사 등이 쓰이는지 등이 그 고려의 조건이 된다.**

1. **산업체 자문 ppt**

**5월 27일 저녁 6시에 산업체 자문 미팅을 진행하였다.**



(사진 자료 및 표 첨부)