

● DH 겨울학교 연구계획서 및 관심사

[DH 겨울학교에 지원하게 된 계기와 기대하는 바]

저는 이번 봄학기 카이스트 디지털인문사회과학부 대학원에 불합격하게 되었습니다. 면접을 준비하며 부족함을 많이 느껴 결과는 겸손하게 받아들이기로 하였으며 일주일이라는 기간동안 불합격한 이유에 대해 곰곰이 생각해보았습니다. 저는 ‘디지털 인문학’이라는 학문에 대해 정확히 알고 있는 것 같지 않습니다. 인문학적인 사고를 할 줄 알고, AI 지식을 받아들일 줄 아는 학생으로써 디지털 인문학 논문이 굉장히 흥미롭고, 미래지향적이고, 역동적인 것으로 느껴졌으나, 제가 ‘디지털 인문학’에 대해 설명하려고 하면 항상 말문이 막히곤 했습니다. 저는 포기하지 않고 ‘디지털 인문학’을 더 적극적으로 체험해보기로 다짐했습니다. 이번 기회를 통해 비슷한 관심사를 가지고 계시는 교수님, 조교님, 참여하신 모든 분들로부터 진정한 ‘디지털 인문학 연구’가 무엇인지 가르침 받고자 합니다.

[DH 겨울학교에서 다루고 싶은 연구 주제]

저는 학부시절 ‘영미문학입문’, ‘영어학개론’, ‘스와힐리어 형태통사론’과 같은 수업을 들으며 언어적 관심을 꾸준히 키워왔습니다. 이런 배경에 따라 저는 디지털인문학 연구 중, 구체적으로 단어의 시대별 의미 변화 연구에 많은 관심을 갖고 있습니다. 예를 들어, 제가 발견한 하나의 작은 발견이 있습니다. 진화론에서 언급되는 ‘자연선택’과 ‘진화’라는 단어의 의미 차이를 알기 위해, 혹은 시대적 배경에 따른 의미 변화를 알기 위해, 구글 N-gram에 ‘natural selection’과 ‘evolution’의 시대별 단어 사용 빈도에 대해 실제 실행해보았습니다. 저는 사회문화적 맥락 안에서 기독교적 이상을 추구했던 19세기 빅토리아 시대에 ‘natural selection’이라는 단어가 거의 사용되지 않았다는 걸 알게 되었습니다. 그 이유는 ‘natural selection’이 신의 개입을 제외하는 단어였기 때문입니다. 실제 이를 직접 시행해 보면서 다윈에 대해 새로운 인문학적 평가를 할 수 있게 되었는데, 이를 문학과 같은 방대한 텍스트 데이터에 디지털 기술을 활용하여 데이터화한다면 인문학은 변화무쌍한 진화의 역동적인 모습으로 재탄생할 수 있다고 믿습니다.

단어의 의미 변화 연구의 범위는 굉장히 넓다고 생각합니다. 위 같은 어떠한 발견에서는 단순히 N-gram과 인문학적 지식만을 이용할 수 있지만, 더 복잡한 의미 변화 연구와 같은 경우에는, 문장 분할, 토큰나이징, 품사 태깅, 불용어 제거와 같은 정규표현식을 이용한 문자데이터 전처리할 수 있는 능력이 필요합니다. 단순히 빈도 수를 이용하여 의미 변화를 알아낸다고 한다면 TF-IDF를 활용할 수도 있지만, 단어의 의미를 고려해야하는 경우에는 토픽 모델링이나 제가 통계학과에서 배운 SVD를 사용한 잠재 의미 분석, LSA에 관한 지식 또한 필요하다고 생각합니다. 이러한 기술들에 대해서는 DH 겨울학교를 통해 더 자세히 고민하고, 교수님들로부터 더 정확한 지식을 습득하고자 합니다.