

○ 강연내용

2003년까지 인류가 만들어 낸 자료의 크기가 5 엑사바이트 정도인데 요즘은 이틀에 한번꼴로 이 정도 규모의 자료가 생성된다고 합니다. 하지만 이러한 정보의 홍수속에서 실제로 유용한 정보를 찾는 것은 건초더미에서 바늘찾기 만큼 어려운 일입니다. 정보가 21세기의 기름이라면 분석은 연소엔진이라는 비유가 있듯이 쏟아지는 정보의 바다에서 등대를 찾아 헤메는 우리에게 데이터과학은 나침반과 같은 존재입니다.

빅데이터의 시대를 맞이하여 인류가 궁금해 왔던 많은 문제에 대해서 답변이 이루어지고 있습니다. 꿈의 재생과 유전체 정보에 대한 분석은 인류가 오랫동안 갈망해오던 시대의 서막을 알리고 있지만 한편에서는 빅데이터의 그늘을 두려워하는 시선도 존재하는 것이 사실입니다..

이러한 데이터의 시대에서 한발짝 옆에 비껴 서있는 것 같은 인문학에서도 변환의 물결은 감지 되고 있습니다. 디지털 인문학은 역사, 문학등 다양한 인문학 분야의 오래된 질문에 대해 디지털화된 대규모 자료를 이용하여 새로운 답변을 얻어내고자 하는 분야입니다.

예를 들면 셰익스피어는 약 18,000개의 단어를 작품에 사용했는데 그 중 “road”, “hurry”와 같은 단어를 포함한 1700개의 단어를 그가 새로 만들어 냈다고 합니다. 그렇다면 그가 실제로 알고 있었던 총 단어의 숫자는 얼마나 될까요? 또한 그의 작품중에 진위여부에 대한 논란이 끊이지 않는 작품들에 대해서 정량적 분석을 통한 진위여부의 판단은 가능할까요?

이와 같이 문학작품을 정량적으로 분석하는 분야를 양식측정학이라고 합니다. 19세기 말에 시작된 양식측정학은 구글 도서관 프로젝트와 구텐베르크 프로젝트를 통한 대규모 문학 작품의 디지털화와 텍스트 마이닝의 등장으로 비약적인 발전을 이루게 됩니다.

이 강연에서는 디지털 인문학을 통하여 문화 예술 전반에 대한 새로운 시각을 제공하고자 합니다. 조산왕조실록의 분석을 통한 예송논쟁의 재조명에서부터 비틀즈 In my life의 실제 작곡자 발견까지, 전혀 어울리지 않는 커플 “데이터과학과 디지털 인문학”과 함께 정보의 바다로 같이 떠나 보지 않으시겠습니까?