

디지털 사회

DIGITAL SOCIETY

2017-1호



연세대학교 사회과학데이터혁신연구센터
SOCIAL SCIENCE DATA INNOVATION CENTER

발행인: 조화순

발행일: 2017년 9월 15일

홈페이지: <http://democracy3.0.yonsei.ac.kr>

왜 데이터 혁신인가?

이선형

(연세대 사회과학데이터혁신연구센터 연구교수)

2016년 3월, 알파고가 바둑 고수인 이세돌 기사를 이기면서 인공지능기술은 전 세계적으로 주목받았다. 그리고 2017년 5월에 세계 바둑 최강자인 커제와의 대국에서 알파고는 다시 3-0으로 승리하면서 우리사회에 적지 않은 충격과 놀라움을 주었다. 이 '사건'은 과학기술의 발전에 대한 긍정적 호기심을 촉발하였지만, 한편으로는 인류의 위기가 도래했다며 우려하거나 구글의 상업적 이벤트에 불과하다는 비판과 같은 부정적 주장이 제기되었다. 정보기술의 발달이 인류 역사에서 처음이 아닌데도 알파고의 승리가 충격적인 이유는 컴퓨터가 몇 가지 정보만으로 '스스로' 학습하여 정보처리를 할 수 있는 정도로 발전했기 때문이다. 기존 컴퓨터가 인간이 입력한 내용만을 처리했다면, 알파고에 사용된 인공지능은 거의 '무'에서 '유'를 창조할 수 있게 되었고, 이러한 혁명은 상상했던 것이지만 피하고 싶은 공포의 대상이었을지 모른다.

알파고와 같은 인공지능 프로그램이 일상생활에까지 보급되는 데에는 오랜 시간이 걸리지 않을 것이다. 그리고 이에 적응하기 위해서는 또 다른 알파고들의 등장에 대하여 무작정 걱정만 하기보다는, 변화에 대한 이해가 먼저 요구된다. 이미 우리사회는 20세기 후반부터 급속한 정보기술의 발전을 경험하고 있지만, 정작 정보에 대한 이해와 이로 인한 혜택은 사회적으로 광범위하게 확산되지 못하고 있다. 특히, 데이터의 공급과 소비가 균형을 이루지 못하고 있고, 기술혁신의 장점을 활용하기 위한 서비스 제공이 제한적이다. 연구분야에서도 정보산업을 주도하는 곳과 그렇지 않은 곳의 활용 속도가 차이가 나며, 이는 융합연구를 더욱 어렵게 만들 것이다. 그렇다면 이러한 문제들을 해결할 수 있는 방법은 무엇일까? 바로 데이터 혁신이다.

데이터 혁신이란 데이터 활용을 극대화하기 위한 데이터의 수집-저장-분석-변환의 전 과정을 의미한다. 이를 통해 기초적인 혜택뿐만 아니라 실용적인 혜택을 누릴 수 있는 기회를 창출하는 것이 데이터 혁신의 핵심 목표이다. 이제 데이터는 단순히 빅데이터와 같은 어마어마한 자료의 규모만을 의미하는 것이 아니라, 새로운 결과를 도출하기 위한 일종의 자원으로써 기능하게 되었다. 사실, 알파고의 등장은 한 기업의 기술 자랑도, 인공지능의 위협도 아닌 데이터 혁신의 과정과 일맥상통한다. 정보기술에서 반드시 많은 데이터가 필요하지 않으며 핵심적인 자료만으로 빠른 반복과정을 통해 의미 있는 결과를 생산할 수 있게 된 것이다. 이를 가능하게 한 것이 알고리즘 개발인데, 결국 알고리즘도 데이터를 기초로 만들어지고 데이터가 없으면 기능하지 못한다.

알파고의 사례에서 확인할 수 있듯이, 정보기술의 혁신은 대부분 민간 산업 분야에서 주도하고 있다. 학계에서도 빅데이터, 머신러닝, 딥러닝 등 새로운 분석방법을 활용하기 위해 노력하고 있으며, 정부 기관에서도 빅데이터 분석을 위한 전문가를 양성하는 등 움직임을 보인다. 그러나 산업 분야와 비교하여 상대적으로 미약한 노력이며, 이러한 현상은 우리사회의 고질적인 문제이기도 하다. 물론, 꼭 학계나 정부기관이 기술발전을 주도해야만 하는 것은 아니지만 이들 역시 사회의 변화에 따라 적절하게 대응하고 문제점을 해결해 나가기 위해 혁신에 적극적으로 참여해야 함을 강조할 필요가 있다. 특히, 정부기관은 다양하고 방대한 공공데이터를 가지고 있지만, 통일되거나 통합되지 않아 아직까지 분석할 수 있는 데이터가 많지 않다. 데이터를 분석하기 위해서는 원시 자료를 정제(cleaning)하고 활용 가능한 수준으로 처리하는 작업이 선행되어야 하며, 이 작업은 나름의 기준과 많은 시간이 요구된다. 때문에 아무리 많은 데이터가 있다고 해도 분석 가능한 상태로 처리하지 못한다면 그 데이터로부터 잠재적 가치를 창출할 수 없다. 그리고 정제된 데이터라고 해도 접근할 수 없다면 소용이 없다.

미국데이터혁신센터(The Center for Data Innovation)는 2015년 5월에 데이터 혁신을 촉진하기 위한 입법관련 12개 권고안을 제시하였다. 내용을 간략하게 살펴보면, 정부데이터 개방 법제화, 위성 데이터 오류 최소화, 해안지역 데이터 구축, 교육정보시스템 향상 등 공공데이터의 공유와 개방을 위한 포괄적인 내용을 담고 있다. 이러한 노력은 경제적 효과뿐만 아니라 학술적 활용을 위한 것으로, 사회 전반에 긍정적인 영향을 줄 것으로 평가되었다. EU도 2014년부터 3조원 규모의 빅데이터 혁신기술연구를 시작한 바 있다. 이처럼 세계 곳곳에서 정부를 주축으로 하는 데이터 혁신이 진행되고 있다.

아무리 좋은 장비가 있어도 재료가 없으면 아무것도 만들어내지 못하며, 재료가 많아도 활용하지 못하면 가치가 없다. 모든 데이터가 디지털화될 필요는 없으며, 과학기술을 통해 분석되어야만 가치가 있는 것은 아니다. 다만, 지금 생산되고 소비되는 정보의 많은 부분이 디지털 자료임을 부인할 수 없으며 그 데이터의 생산 속도는 더욱 빠르게 진행되고 있다. 따라서 데이터 분류/저장과 같은 기본적인 처리뿐만 아니라 분석도구의 개발은 미래에 끊임없는 과제가 될 것이다. 무엇보다도 데이터 혁신 수준이 국가 경쟁력과 직결되는 만큼 국가 차원에서 데이터 활용의 효율성 증대와 한계를 극복하기 위한 현실적 제안이 필요하다.