

뉴스스트리스트위원회 ‘디지털시대 뉴스 유통과 알고리즘 세미나’

언론계·학계 협력 ‘신뢰도 높은 뉴스 알고리즘’ 개발

구본권 / 뉴스트리스트위원회 부위원장 · 한겨레신문사 사람과디지털연구소장



인터넷과 모바일 환경에서 저널리즘 수준과 역할을 제고하기 위해 신뢰도 높은 뉴스를 찾는 알고리즘이 필요하다는 인식 아래, 뉴스트리스트위원회가 출범했다. (사진 출처=필자 제공)

뉴스트러스트위원회가 주관한 ‘디지털시대 뉴스 유통과 알고리즘 세미나’가 2016년 12월 13일 프레스센터 내셔널프레스클럽에서 한국언론학회, 한국방송학회, 한국언론정보학회, 한국기자협회, 한국신문협회, 한국방송협회 공동 주최로 열렸다. 2016년 5월 출범한 뉴스트러스트위원회가 개최한 이날 세미나에서는 그동안의 작업 결과와 향후 방향에 대한 내용을 공개하고 6개 학회와 협회가 추천한 패널들의 토론이 진행됐다. 필자는 뉴스트러스트위원회 위원으로 활동해 오며 이날 세미나에서 발표를 했다. 뉴스트러스트위원회의 목표와 활동 상황, 향후 과제와 함께 세미나에서 토론한 내용을 전달한다.

뉴스트러스트

출범 배경

행태가 근본적으로 달라져 포털과 검색 서비스 등 플랫폼과 정보기술이 끼치는 영향력이 막강해졌다. 미디어 생태계는 항상 기술의 영향을 받지만 알고리즘으로 대표되는 기술 환경은 뉴스 생산과 유통, 이용 관행에 혁명적 변화를 불러오고 있다. 포털 사이트, 검색엔진 등 기술 플랫폼은 뉴스를 이용자에게 전달하는 유통망을 넘어 뉴스의 생산과 이용 등 언론 환경 전반에 지대한 영향을 끼치고 있다.

인터넷과 모바일을 이용해 뉴스를 이용하는 환경은 비단 미디어의 유통 채널과 이용 수단의 변화에 그치지 않는다. 미디어 플랫폼과 알고리즘은 뉴스의 형식을 새롭게 규정하는 것은 물론이고 뉴스의 내용과 가치에도 중대한 변화를 가져온다. 포털과 검색은 편리한 뉴스 이용 도구이고, 이용자 대부분은 모바일의 포털 홈페이지에서 뉴스를 만난다. 이용자들은 개별 언론사를 방문하지 않고도 다양한 뉴스를 만나 왔고 이에 따라 뉴스의 이용 빈도와 영향력도 함께 높아졌다.

인터넷과 모바일 환경에서

뉴스의 생산과 유통, 이용

뉴스 유통과 이용 환경의 변화는 저널리즘의 내용과 가치에도 새로운 차원의 문제를 제기한다. 전에 없이 많은 뉴스가 이용되고 있지만, 저널리즘은 전보다 개선됐는가? 이용자들은 모바일에서 어떤 기준과 가치에 따라 뉴스를 소비하는가? 뉴스 알고리즘과 플랫폼은 어떻게 만들어지고 누가 구조를 결정하는가? 모바일과 알고리즘 환경에서 저널리즘은 개인과 사회에 정보 제공과 여론 형성 기능을 충실히 수행하고 있는가 같은 질문이다. 모바일과 알고리즘이 언론 환경에 던진 다양한 효과는 긍정적 측면과 부정적 측면이 섞여 있지만, 지금까지는 언론이 어떻게 접근해야 하는지조차 알기 어려웠다. 알고리즘은 기본적으로 작동 원리와 구조가 바깥으로 드러나지 않는 블랙박스 속의 기술로 불린다.

모바일 환경에서 이용자들은 다양한 뉴스를 개인적 관심에 따라 이용하는 편의를 누리지만, 저널리즘은 새로운 도전과 위기를 만났다. 공동체의 주요한 사안을 보도하고 여론을 형성해야 하는 언론이 흥미와 선정성 위주의 기사를 생산하고 트래픽 증대에 집중하고 있다. 일부 선정적·상업적 매체의 일이라고 넘기기에는 언론 환경 전반에 끼치는 영향이 심대하다. 공공적 중요 사안에 대한 뉴스보다 개인적 관심 위주의 가십들이 과도하게 생산, 유통, 소비되는 현상은 언론의 역할과 여론 형성 기능을 위축시켜 민주주의 사회를 위협할 수 있다. 2016년 미국 대선은 소셜네트워크서비스에서 알고리즘에 의해 뉴스가 소비되면서 벌어질 수 있는 폐해를 보여주며 뉴스 알고리즘에 대한 사회적 통제 필요성을 공론장으로 불러왔다. 페이스북과 구글에서 ‘가짜 뉴스(Fake News)’가 난무한 결과다.

포털의 영향력이 큰 국내에서는 클릭을 유도하기 위해 포털의 인기 검색어를 조합한 뉴스, 선정적 제목과 표현으로 눈길을 끄는 기사, 취재 없이 다른 기사를 베끼는 어뷰징 기사 등이 난무하면서

해당 언론사만이 아니라 저널리즘 전체의 품질과 신뢰가 함께 하락하는 현상이 일어나고 있다. 영국의 로이터저널리즘연구소가 발행하는 〈디지털 뉴스리포트〉 2016년 보고서에서 한국은 뉴스 신뢰도가 조사 대상 25개국 중 22위로 최하위 수준을 기록했다. 한국은 뉴스 소비에서 포털과 스마트폰 의존도가 유난히 높았으며, 일본과 함께 경성 뉴스에 대한 관심도가 가장 낮고, 연성 뉴스에 대한 관심도가 가장 높은 나라로 나타났다. 방치할 경우 저널리즘의 위기로 이어질 수 있는 상황이다.

이날 발표에서 최수진 국민대 교수는 기계적이고 자동화한 알고리즘이 객관적이고 중립적일 수 있다는 견해는 성립하지 않는다면 알고리즘 역시 인간의 판단이 개입한 창작물이기 때문에 투명성과 책무성이 요구된다고 말했다. 뉴스 알고리즘을 기계의 영역에 두는 대신 사회적 감시와 비판을 받아야 하는 공공의 영역으로 삼아야 한다는 것이다.

뉴스트러스트위원회의 구성과 역할

역할을 제고하기 위해 신뢰도 높은 뉴스를 찾는 알고리즘이 필요하다는 인식에 이르렀고, 뉴스트러스트 위원회가 출범하게 됐다. 뉴스트러스트 위원회는 윤영철 연세대 교수(언론학)를 위원장으로 한 15인의 협의체다. 위원회는 구본권(한겨레), 김사승(송실대), 김성태(고려대), 김춘식(외대), 안혜리(중앙일보), 우병현(조선일보), 이상구(서울대), 이성규(메디아티), 이종근(데일리안), 정철웅(KBS), 조문기(MBC), 최수진(국민대), 최민재(언론진흥재단), 한운희(연합뉴스) 등 학계와 언론계의 다양한 전문가로 구성됐다. 뉴스트러스트위원회를 운영하고 어떤 결정을 할 때 언론진흥재단은 지원 업무를 담당하고 모든 것은 위원회가 공식 회의를 통해 독립성과 책임성을 갖고 진행한다.

“
뉴스트러스트위원회가
목적하는 뉴스 알고리즘은 뉴스의 생산에
알고리즘을 활용하는 게 아니라,
많은 언론사가 다양한 주제로 생산한 기사를
인터넷에 노출할 때 어떠한 기준과 가치에 따라
선별하고 배열할 것인지에 주목한다.”
”

위원회는 지난 5월 출범 이후 매달 2차례씩 12회의 전체 회의를 개최하고, 7인으로 실무 소위원회를 구성해 구체적 논의를 담당하는 방식으로 업무를 진행해 왔다. 2017년 3월 베타 버전의 뉴스 알고리즘을 공개할 예정이다. 뉴스트러스트위원회의 모든 회의와 활동 내용은 공개되고 논의의 결과물인 뉴스 알고리즘은 일차적으로 언론진흥재단이 운영하는 통합 뉴스 포털 빅카인즈에 적용된다. 뉴스 알고리즘은 각 언론사가 가져다 사용할 수 있도록 공개되고 다양한 활용이 가능하도록 각종 자료와 도구가 API 형태로 제공될 예정이다.

위원회가 목적하는 뉴스 알고리즘은 뉴스의 생산에 알고리즘을 활용하는 게 아니라, 많은 언론사가 다양한 주제로 생산한 기사를 인터넷에 노출할 때 어떠한 기준과 가치에 따라 선별하고 배열할 것인지에 주목한다. 이를 위해 위원회는 초반에 뉴스 알고리즘이 지향해야 할 저널리즘의 가치와 규범이 무엇인지에 대한 논의를 진행해, 그 결과로 ‘신뢰도 높은 뉴스 알고리즘’을 만드는 것이라는 합의를 이뤄냈다.

위원회는 인터넷에서 다양한 뉴스를 고유의 기준을



2016년 5월 출범한 뉴스트러스트위원회의 모든 회의와 활동 내용은 공개되고 논의의 결과물인 뉴스 알고리즘은 언론진흥재단의 빅데이터에 적용된다. (사진 출처-필자 제공)

만들어 평가하는 알고리즘을 위해 기존 연구를 참고했다. 보거트(Bogart, 1989)의 신문사를 평가하기 위한 신문품질지수, 메이어&김(Meyer & Kim, 2003)의 신문 품질 평가 요인, 박재영 · 이완수(2010)의 뉴스평가지수, 헤이건(Hagen 1995)의 개별 진술 단위의 보도 품질 분석 사례 연구들을 통해 뉴스를 평가하기 위한 요인들과 계량화 방법을 참고했다. 또한 커티스(Curtis, 2013) 등이 구글을 분석해 구글 뉴스 편집에서 노출 순위를 결정하는

요소들에 대한 계량화 방법도 검토했다.

위원회는 뉴스 알고리즘이 지향하는 저널리즘의 신뢰성이라는 개념을 11개의 하위 개념으로 구체화한 뒤, 각각의 개념이 지향하는 가치를 측정 가능한 조작적 개념으로 만들었다. ‘신뢰성’이라는 가치를 진실성, 유익성, 윤리성 등 세 카테고리로 구분하고 그 아래에 11개의 개념(사실성, 투명성, 다양성, 균형성, 독창성, 중요성, 심층성, 독이성, 유용성, 선정성, 반복성)으로 구체화했다. 이를 계량화하기

위해 58개 요인을 항목별로 적용하고 그중에서 유의미하다고 판단되는 185개 변인을 선택했다. 향후 뉴스 알고리즘은 이를 기초로 해서 시뮬레이션을 거듭하며 유용한 결과를 내놓을 수 있도록 최적화와 보정 작업을 지속할 예정이다.

뉴스 알고리즘이 작동하는 방식을 설명하기 위해 기사의 사실성을 평가하려는 계량적 측정 기준을 예시하면 다음과 같다. 위원회가 현재 제시한 ‘사실성’의 측정 기준은 △기사에 등장한 개체명이 많을수록 △등장한 정보원의 수가 많을수록 △고유 정보가 기사 내에 많이 포함될수록 점수를 많이 주는 방식이다. 한편 △익명의 정보원 수가 많을수록 △무주체 술어(~로 알려졌다)가 많을수록 △주관적 술어 사용이 많을수록 감점을 받는 구조다. 또한 선정성의 경우는 △비속어 사용이 많을수록 △선정적 표현이 많을수록 △‘앗 이럴 수가’ ‘충격’ 등 감정 노출 단어를 제목에 많이 포함할수록 감점을 받는다.

앞으로의 과제

뉴스 알고리즘의 목표와 윤곽이 제시되자, 학계와 언론계 패널들은 원칙적으로 취지에 공감한다면서도 알고리즘의 적용 범위와 실행 결과 등 현실에서 애초의 동기가 구현될 수 있을지에 대한 우려를 제기했다. 일부 패널들은 ‘신뢰도 높은 뉴스’를 알고리즘을 통해 평가하겠다는 위원회의 목표가 지나치게 방대하고 기술적으로 이를 구현하기 위한 방법적 구체성과 적용 대상이 명료하지 않다면 실현 가능성 있는 목표를 정하고 적용 대상을 축소해야 한다고 지적했다. 또한 위원회가 측정 수단과 기준으로 예시한 요소들이 애초 의도한 바와 다른 결과를 가져올 수 있으며 측정하려는 목적을 이루기보다 오히려 새로운 어뷰징 기사를 만들어 내는 기준이 될 수 있다는 점도 지적됐다. 인기 검색어가 서비스되면서 이를 단어를 포함시킨 어뷰징 기사가 난무한 것처럼, 뉴스 알고리즘의

기준이 공개되면 이를 염두에 둔 새로운 기사 작성법이 생겨나게 되는 부작용이 나올 수 있다는 얘기다. 페이스북에서 가짜 뉴스가 기승을 부린 것은 가짜 뉴스들이 오히려 페이스북의 뉴스 알고리즘이 요구하는 형식적 조건을 충족시킨 결과일 수 있다는 지적도 제기됐다.

패널들의 지적에 대한 뉴스트리스트위원회의 답변은 현재 단계는 완벽하게 작동하는 알고리즘의 구현을 논의하고 평가하는 시점이라기보다 알고리즘이 영향을 끼치는 뉴스 생태계를 좀 더 잘 이해하고 그 환경에서 언론의 공공성을 고민하기 위한 수단으로 접근해야 하는 시점이라는 것이었다. 최수진 교수는 “뉴스 알고리즘 자체가 투명해야 한다는 취지로 공론장을 만드는 것이다. 트래픽을 추종하는 알고리즘은 이미 많은 만큼 공익을 추구하는 알고리즘도 하나쯤은 있어서 뉴스 환경을 다변화하는 게 이용자에게도 좋은 영향을 미칠 것이라 본다”고 말했다. 구본권 부위원장은 “뉴스트리스트위원회가 시도하는 뉴스 알고리즘은 가보지 않은 길”이라며 “디지털 세상에서의 뉴스 생태계를 포괄적으로 이해하기 위한 노력의 일환”으로 장기적 접근과 지속적인 업그레이드의 필요성을 강조했다.

2017년 3월 등장할 뉴스 알고리즘 베타 버전은 우선적으로 언론진흥재단이 운영하는 빅카인즈 포털에 적용된 이후 지속적인 조정과 고도화 작업을 거칠 예정이다. 또한 뉴스 알고리즘은 각 언론사와 이용자들이 평가 요인별 가중치 조정을 통해 다양한 기준의 뉴스를 선별할 수 있는 도구로도 활용될 여지가 있다. 뉴스 알고리즘 개발 과정에서 만들어져 공개될 각종 용어 사전이나 금칙어 사전을 비롯한 데이터베이스와 형태소 분석기 같은 도구는 개별 언론사들이 API 형태로 가져다가 자사의 서비스에 적용할 수 있어 새로운 효용을 낼 수 것으로 기대된다. ♣