

뉴스 어뷰징 추방, 검색 알고리즘 개선해야



네이버나 구글에서 실시간 검색어나 조회 수가 상위에 올라있는 글들은 어떻게 해서 그렇게 되었을까? 어떤 기사는 온라인과 모바일에서 어떻게 사람들의 주목을 그토록 많이 낚을 수 있었을까? 밑밥을 뿌리지 않았는데도 저절로 상위 수준으로 올라온 것일까?

결론부터 말하자면 검색어를 추출하는 알고리즘(algorithm)에 따라 검색 순위가 전혀 달라진다는 것이다.

우리나라 온라인과 모바일 광장은 정녕 낚시꾼들이 물려드는 낚시터인가? 어떤 언론사들의 뉴스 어뷰징 행태는 막장 드라마와 다를 바 없다.

어뷰징(abusing)이란 ‘남용’과 ‘오용’을 뜻하는 영어 어부스(abuse)에서 파생된 어휘다. 따라서 기사 어뷰징이란 온라인과 모바일에서 조회 수를 높이기 위해 같은 기사를 제목이나 내용만 조금 바꿔 반복으로 전송하는 행위다.

네이버와 구글의 검색 알고리즘 비교

2006년 12월, 네이버에서 키워드를 입력하면 자동으로 언론사 홈페이지의 기사 검색으로 연결되도록 검색 방식을 바꾼 다음부터, 뉴스 어뷰징이 전염병처럼 창궐하기 시작했다. 그 이전까지는 독자가 기사 제목을 클릭하면 네이버 안에 저장된 기사가 소개되는데 그쳤지만(인 링크), 검색 방식을 바꾼 다음부터는 해당 언론사의 홈페이지로 넘어가 자동으로 기사 화면으로 접속되었다(아웃 링크).

어떤 언론사에서 자사 기사에 대한 조회 수를 높이려고 어뷰징을 하는 이유는 간단하다. 기사의 조회 수가 많으면 저절로 광고 수익과 직결되기 때문이다. 지금 이 순간에도 어떤 언론사들은 비슷비슷한 기사를 다량으로 중복 전송하고 있다. 뉴스 어뷰징에는 성, 연애, 연예인 관련 뉴스들이 특히 많다. 문제는 내용을 약간씩 바꾼 다음 자극적인 제목을 달아 저널리즘 가치를 스스로 훼손하고 있다는 점이다. 이번 세월호 참사와 관련된 기사에서도 어뷰징 기사들이 수백 건에 이르렀다. 이는 포털이 실시간 검색어 장사를 하는 것이나 마찬가지인데, 이 역시 알고리즘 문제로 귀결된다.

알고리즘이란 어떤 문제를 해결하기 위해 정해진 일련의 절차나 방법을 의미하며, 다음과 같은 5가지 조건을 만족시켜야 한다. 즉, 외부에서 제공되는 자료가 존재해야 하는 ‘입력성’, 적어도 1개 이상의 결과를 도출해야 하는 ‘출력성’, 모호하지 않은 명확한 명령어로 기능을 수행해야 하는 ‘명확성’, 정해진 명령어를 수행한 다음 주어진 시간 내에 종료해야 하는 ‘유한성’, 그리고 모든 과정을 명백히 검증할 수 있어야 하는 ‘효율성’이 그것이다.

구글은 이미 뉴스 순위를 정하는 알고리즘을 공개한 바 있다. 알고리즘을 공개한다는 것은 기업의 비밀을 공개하는 것과 마찬가지며, 나아가 경쟁의 규칙을 천명한 것이라고 할 수 있다. 구글은 ‘정확한 정보 검색’과 ‘사용자 편

의 제고'라는 두 가지 원칙을 바탕으로 알고리즘을 발전시켜 왔다. 키워드 검색에서 출발한 구글은 '자연어' 검색은 물론 '대화형' 검색까지 가능하게 만들었다. 예컨대, 널리 알려진 구글의 '페이지 링크' 알고리즘 기술도 구글에서 만든 수천 개의 알고리즘 가운데 하나에 불과하다.

구글은 PC보다 더욱 개인화된 기기인 모바일에 적합한 검색 서비스를 준비해 '지식 그래프'를 발표했는데, 바로 개인 맞춤형 검색이다. 사용자가 원하는 정보를 일목요연하게 보여주기 위해 카테고리화 DB를 구축한 것이다. 지식그래프는 2012년 5월 미국에서 적용된 이후, 한국어, 프랑스어, 독일어, 이탈리아어, 스페인어, 일본어 등을 지원하고 있다. 구글은 5억 8천만개의 대상에 지식그래프를 적용해 정보를 연결한다. 이 정보를 연결하고 선별하는데 컴퓨터 알고리즘이 결정적인 기능을 발휘한다.

네이버 역시 '지식인' 서비스를 적극 활용해서 데이터베이스(DB)를 구축해왔다. 네이버는 포털(Portal)의 수많은 정보와 뉴스 중에서 중요도와 재미에 따라 편집자가 직접 골라 링크를 연결하는 방식을 적용했고, 모바일 정답형 정보 서비스를 모바일 앱과 PC로 확대해왔다. 정답형 자동완성 기능이 제공되는 검색어를 검색하면 검색 버튼을 누르지 않아도 검색 창에서 바로 원하는 정보를 확인할 수도 있다. PC와 모바일 앱에서도 지원되는 검색어가 확대되었고, 필요한 정보를 연결하고 큐레이션하는 개선 작업도 계속해 왔다. 그렇지만 어뷰징 문제를 해결하지 못하고 있다. 따라서 이 문제를 개선할 수 있는 특단의 알고리즘 개발 전략이 필요하다. 네이버와 구글의 검색 기술의 차이를 비교하면 <표 1>과 같다.

<표 1> 네이버와 구글의 검색기술 비교

	네이버 (NHN)	구글 (Google)
명칭	정답형 검색	지식 그래프
중심 플랫폼	모바일 중심	웹 중심
정보 성격	빠르게 찾는 정답형 정보 중시	정보 간 연관성 중시
정보처리방식	정보 큐레이션	기계적 처리

뉴스 어뷰징 추방 위해 알고리즘 한계 보완해야

그렇다면 뛰어난 알고리즘 기술이 모든 문제를 해결할 수 있는 만능 처방약이 될 수 있을까? 그렇지 못하다. 그 놀라운 장점에도 불구하고 모든 알고리즘은 일정한 한계를 지니기 때문이다. 알고리즘의 특성은 대체로 다음과 같다.

첫째, 알고리즘은 원래 불공정하고 객관적이지 않다. 알고리즘은 설계자의 의도나 여러 가지 사회적 요소들이 개입될 수밖에 없는 논리 언어이기 때문에 완벽한 공정성과 객관성을 구현하기 어렵다. 아무리 공정한 배열을 목표로 뉴스 랭킹 알고리즘을 설계한다고 해도, 그 과정에서 설계자의 가치 기준이 개입될 수밖에 없다. 따라서 누구나 만족할만한 공정한 가치가 반영된 가치중립적이고 객관성을 담보할 알고리즘은 처음부터 불가능하다.

둘째, 언제나 어뷰징의 가능성을 내포하고 있다. 알고리즘 원칙이 세상에 공개되면 어떤 형태로든 어뷰징 행위가 나타나게 마련이다. 사실 뉴스 검색의 순위를 정하는 문제는 알고리즘 설계자의 아이디어 발상과 언론사 기자들의 노출 욕구 사이에 벌어지는 총성 없는 전쟁과 같다. 어떤 알고리즘을 개발하려고 아이디어 발상을 하는 순간부터 선정적인 제목을 어떻게 막느냐와 그럼에도 불구하고 어떻게 진입하느냐의 머리싸움이 벌어진다고 봐야 한다.

그렇다면 해결책은 없는 것일까? 바람직한 대안은 무엇일까?

네이버나 다음 같은 우리나라의 주요 포털에서 알고리즘의 한계를 보완하려는 적극적인 노력이 그 무엇보다 시급하다. 다음과 같은 세 가지를 생각해볼 수 있겠다.

첫째, 알고리즘 편집에 있어 불공정성을 최소화시키려고 노력해야 한다. 가치중립적인 공정성을 갖춘 뉴스 랭킹 알고리즘의 설계가 불가능한 상황에서, 어떻게 하면 불공정성을 최소화시킬 수 있는지 우리나라 현실에 맞는 아이디어를 찾아내야 한다. 인기 댓글을 유도하거나 낚시성 기사를 우선순위로 올려놓는 문제가 지속된다면 장기적

인 측면에서 포털에 대한 실망감으로 접속자의 대량 이탈 사태가 발생할 수도 있다. 구글의 검색 알고리즘은 키워드 방식이 아니다. 페이지랭크(PageRank)를 기반으로 개발된 구글의 검색엔진은 이제 빅데이터를 바탕으로 분석하는 ‘구글링’이라고 할 정도로 위상을 키워가고 있다. 네이버 역시 어뷰징, 낚시성 기사, 오보를 막을 수 있도록 단순 키워드 방식을 대체할 알고리즘을 제시해야 하는데, 공정성을 최대한 보장하는 방향으로 개발이 이루어져야 한다.

둘째, 포털에서는 알고리즘의 구성 원칙이나 어뷰징 방지 가이드라인을 발표해야 한다. 공개할 수 있다면 알고리즘 체계를 공개하는 것이 가장 좋다. 만약 기업 비밀이라 공개하기 어렵다면 알고리즘의 구성 원칙이라도 공개해야 한다. 즉, 알고리즘의 설계 목적을 제시하고, 그 목적을 위해 어떤 요소에 가중치를 주었다는 큰 틀에서의 원칙만은 공개해야 한다. 또한, 기자들 스스로 어뷰징 경쟁에서 벗어나 저널리즘 원칙을 회복하는 문제도 시급하지만, 포털에서는 제휴 언론사에게 어느 정도 자극적인 내용이면 기사로 올릴 수 없다는 기준을 제시해야 한다. 예를 들어, 자극적인 내용의 정도를 4등급(게재 불가, 대폭 수정 후 게재, 부분 수정 후 게재, 무수정 게재)으로 나눠 어뷰징 행위를 축소시킬 가이드라인을 제시해야 한다.

셋째, 포털이 뉴스 중개자의 책무를 적극적으로 수행해야 한다. 이제, 포털은 단순한 뉴스 중개자가 아니라 사회 구성원들에게 실질적으로 엄청난 영향을 미치는 언론사가 되었다. 아무리 피해가려 해도 엄연한 현실이다. 그렇다면 자율적인 편집권을 바탕으로 저널리즘의 가치를 존중해야 한다. 부동산 중개업자만 하더라도 매물을 사는 사람과 파는 사람 사이에서 어떤 피해를 보지 않도록 중개업자로서의 책무를 다 한다. 그런데 현재의 포털은 뉴스 중개자로서의 중개 수수료(광고비 등 막대한 혜택)는 다 챙기면서도, 뉴스를 생산하는 언론사와 뉴스를 소비하는 수용자 사이에서 부동산 중개업자보다 책임감이 부족한 것 같다. 앞으로 포털은 언론으로서의 사회적 책임감을 가지고 보다 적극적으로 책무를 다해야 한다.

이상에서 제시한 세 가지 대안 외에도 여러 가지 새로운 방안들이 있을 수 있다. 네이버나 다음 같은 우리나라의 주요 포털들은 여기에서 제시한 세 가지만이라도 먼저 실행하기를 바란다. 알고리즘을 개발하는 일은 언제 끝날지 모르는 힘들고 고단한 작업이리라. 그렇지만 그렇게 피흘리는 창의적 정신이 뉴스 어뷰징을 추방하는 기폭제로 작용할 것이다. **K A A**

