

인터넷 검색엔진: 사용자의 관심을 흡수하여 전문성을 강화하는 기술[†]

김지연*

본 논문에서는 일반적인 웹검색 기술에 대한 분석을 포함하여, 한글검색엔진에 대한 접근도 전개하고자 한다. 최근 검색엔진의 독점적 지위문제, 검색순위의 공정성 문제가 발생하면서 우려가 높아지고 있다. 검색엔진을 둘러싸고 발생하는 상이한 해석적 주장들에는 기술결정론적 관점에서부터 도구주의적 관점까지 다양하다. 여러 해석적 주장들은 강한 지향성을 가지고 있지만 또한 그럴만한 기원도 가지고 있다. 이처럼 다양한 해석이 병존하는 것은 이 기술에 대한 통합적 재해석이 제기되고 있다는 의미일 것이다. 검색엔진은 사용자의 관심을 흡수하여 자신의 질서를 부여하는 기술양식이다. 이는 검색엔진의 합리성으로부터 기원한다. 특히 한글검색엔진은 사용자들의 관심을 끌어들이므로써 자신의 합리성을 증식시켜왔다. 한글검색은 일반적인 검색엔진이 수행하는 문서들 사이의 관계만이 아니라, 사용자들이 입력하는 단어 사이의 기호학적 관계를 양식화해냈고 그로써 강력한 권위를 구성해냈다. 이제 검색엔진의 판단은 단순한 안내자의 지위를 넘어서 민주주의의 문제가 되었다. 어떻게 검색엔진의 전문성을 승인하면서도 민주주의에 봉사하도록 할 것인가? 검색엔진과 인간행위자를 분절적으로 보는 대신에, 둘 사이의 관계방식에 초점을 둘 때 검색기술에 대한 새로운 전망을 내놓을 수 있을 것이다.

【주제어】 검색엔진, 네이버, 다음, 네이트, 구글, 기술의 합리성, 전문성

[†] 이 논문은 2010년도 정부 재원(교육과학기술부 인문사회연구역량강화사업비)으로 한국연구재단의 지원을 받아 연구되었음 (NRF-2010-330-B00170).

* 가톨릭대학교 인문과학연구소 연구교수

전자우편: spring900@gmail.com

1. 서론: 검색엔진에 대한 편향들과 문제제기

오늘날 인터넷 검색엔진은 정보사회를 대표하는 기술이며 가장 널리 알려진 기술이다. 널리 사용된다는 것은 널리 수용되었다는 의미이며 또한 그런 만큼 논쟁의 대상이다. 이미 국내외에서 특정 검색엔진의 독점적 지위문제, 검색순위의 공정성 또는 오래된 문서의 노출로 인한 분쟁 등 사회정치적이고 윤리적인 문제들이 제기되고 있다.

현재 구글 검색엔진은 영어권만이 아니라 다수의 나라에서 우세한 점유율을 보이고 있고, 한국(네이버, 68.8%), 중국(BAIDU, 72.4%), 러시아(YANDEX, 64.7%)만이 현지의 검색엔진이 우세한 상태이다¹⁾. 왜 그런 현상이 발생하는지에 주목해 볼 만하다. 대부분의 기술들이 그렇듯이 검색엔진도 국제적인 보편적 특질을 지니면서도 지역적 스타일로 분기하는 현상이 존재하지만, 한편으로는 점차 보편적 특질이 지역적 스타일을 압도하는 경향도 보이고 있다. 검색엔진의 기술적 스타일을 관찰함으로써 기술과 사회의 공동 진화의 한 측면을 발견할 수 있을 것이다.

일반적으로 기술에 대한 두 가지 편향이 있다. 하나는 기술결정론적 관점이다. 이것은 기술이 사회와 독립적으로 자신의 고유한 자율성에 의해 발전한다는 주장이다. 더 나아가 기술이 사회의 발전에 결정적 영향을 준다는 주장도 포함한다. 기술결정론적 관점은 종종 기술에 대한 맹목적 신뢰로 나타나거나, 그와 정반대로 기술에 대한 혐오로 드러나곤 한다. 기술에 대한 맹목적 신뢰나 기술혐오는 외관상 달라 보이지만 사실은 같은 인식론적 토대 위에 있다. 기술이 더 결정적인 역할을 한다고 보기 때문에 중대한 문제가 발생했을 때 기술이 그 문제의 핵심적인 원인이라고 지목하는 것이다.

기술에 대한 또 다른 편향은 도구주의적 관점이다. 이것은 기술결정론과는 대조적으로 기술 자체에 아무런 독자적 행위성이 없으므로 기술은 가치중립

1) 아운코리아마케팅 2013년 자료를 참조했다.

적이라고 간주한다. 기술은 단지 인간의 도구에 불과하기 때문에 기술 자체는 투명하다고 보는 것이다. 이 주장에 따른다면 총과 같은 무기는 단순한 도구이므로 누가 사용하느냐에 따라 그 용도가 달라질 뿐, 총 자체에는 어떤 행위성도 없으므로 총 또는 총의 제작자에게 어떤 책임도 물을 수 없다. 이 관점은 결과적으로 책임소재를 인간행위자-특히 사용자-에게 귀속시킨다. 중요한 문제가 일어났다면 그 사용자를 찾아서 그에게 책임을 물어 대가를 치르게 하면 충분하다.

기술결정론이나 도구주의만으로는 현대의 기술들을 해석하고 그 문제를 분석하는데 충분하지 않다는 것이 최근 기술사회학의 공통 입장이다. 기술이나 인간행위자 어느 한쪽에 책임을 모두 귀속시키는 방식은 해석상의 편의성은 있으나 문제의 해법을 도출하는데 그다지 효과적이지 못하다. 대표적으로 행위자-연결망 이론(ANT, Actor Network Theory)은 기술결정론이나 도구주의 대신에 기술과 사회 모두에게 행위성이 있고 그 행위자들 사이의 연결 방식에 초점을 두어 기술현상을 해석하고 있다. ANT에 따르면, 인공물이 안정화되면 그것은 지식체계가 되고 사회적 관계를 변화시킬 수 있는데, 그 지식의 배포 과정은 권력의 배포를 수반한다.

그런 점에서 권력은 소유되는 것이 아니며 행위자들에 의해 만들어지는 일종의 합성물이다. 행위자들은 어떤 지식주장, 사회질서, 인공물의 사용방식 등을 다음 행위자에게 전달한다. 다시 말하자면 어떤 주장, 질서, 인공물, 상품의 시공간적 확산은 행위자들의 손에 달려 있다(Callon, 1980; Law & Callon, 1992; Latour, 1999, 2005). 각각의 행위자들은 다양한 방법으로 행동하고, 자신의 목표에 따라 그것을 변형하기도 한다. 행위자들 사이에서 채택되고 전달된 것은 권력이 된다. 인공물의 안정화 과정은 구체적인 현장에서 발생하는 권력의 작동관계이며 그 위임의 결과이다. 기술은 구체적으로 현장의 결정과 상세 내역을 통해 그 결과가 취약해 질 수도 있고, 반대로 더욱 확고해질 수도 있다(Latour, 1986).

본 논문에서는 일반적인 웹검색 기술에 대한 분석을 포함하여, 동시에 한

글검색엔진에 대한 접근도 전개하고자 한다. 검색엔진을 둘러싸고 발생하는 상이한 해석적 주장들 역시 기술결정론적 관점에서부터 도구주의적 관점까지 다양하다. 우리는 검색엔진에 대한 기술결정론적 관점을 쉽게 만날 수 있다. 예를 들어 한글 검색서비스에 대해서 검색 알고리즘에 의존하지 않고 인간편집자의 개입을 허용하기 때문에 객관성을 훼손할 것이라는 주장이 있다(김평호, 2007; 설진아, 2011; 김인성·김내리, 2012). 이것은 의미 있는 비판이다. 그런데 동시에 우리는 “과연 기술적 알고리즘은 항상 객관적이고 믿을만하며, 인간편집자는 언제나 주관적이고 위험한 것인가?”라는 질문을 해볼 필요가 있다.

검색엔진 자체가 사용자를 바보로 만든다는 비판도 있다. 검색엔진이 사용자에게 편리하고 익숙한 정보서비스를 제공함으로써 사용자가 과도하게 의존하여 마침내 다른 생각을 하지 못하도록 만든다는 것이다(Haigh, 2006; 바이디야나단, 2012). 이 주장에는 검색엔진기술에 대한 혐오의 관점이 포함되어 있다. 이들의 비판은 타당성이 있지만 되짚어 볼 문제도 있다. “과연 검색엔진은 사용자 또는 사회적 맥락에 전혀 개의치 않고 홀로 그런 진화의 길을 걸어간 것인가?”, “검색엔진에게만 문제의 책임을 지울 수 있는가?” 또는 “과연 그렇게 해서 문제가 해소될 수 있을 것인가?”도 생각해볼 일이다.

최근에는 검색엔진이 새로운 권력으로 등장하는 것을 경계하는 우려가 높다. 그들의 주장에 따르면 검색엔진은 콘텐츠를 생산하지도 않고 정보를 소유하지도 않는다. 검색엔진은 단지 정보를 측정하고 그에 관한 도식만 가지는 것으로 지대이익(Rent profit)을 얻고 있으므로 기생적 성격을 가진다는 것이다. 그러므로 사용자들은 자신이 제작한 정보에 대해서 검색엔진회사에게 비용을 요구해야 한다는 주장이다(Pasquinelli, 2009; 백옥인, 2013 재인용). 이 또한 의미 있는 비판이지만 이 해결방식은 저작권과 같은 기존의 질서를 오히려 강화시킬 우려가 있다.

한편 검색엔진 자체는 특정 가치로부터 중립적이라고 보는 관점도 존재한다(황성기, 2007). 이러한 인식론은 검색기술을 단순한 도구로 보는 태도에 기반을 둔다. 도구주의적 관점은 가장 널리 유포된 주장인데 기술에 대한 비판

을 무력화시킨다는 점에서 상당한 편향을 포함하고 있다. 검색엔진은 웹문서와 사용자를 연결하는 역할이므로 그 시스템 자체는 단지 경로만을 제공한다 는 관점이다. 그동안 검색엔진이 어떤 다른 매체보다도 ‘투명해지려고’ 스스로 노력해왔다는 점에서 이 주장 역시 상당한 타당성이 있다. 그럼에도 불구하고 “과연 검색엔진이 아무것도 보낸 것이나 뺀 것이 없이 웹문서와 사용자를 그 자체로 연결하는 것이 가능한가?”라는 질문에 대해서는 회의적이다.

여러 해석적 주장들은 저마다 주관적 지향성을 가지고 있지만 단순한 편향 이라고 비판할 수만은 없는, 그럴만한 기원도 가지고 있다. 하나의 태도가 주 관적 지향이면서 동시에 기원을 가질 수 있는 것은 이 기술에 대한 통합적 재해석이 제기되고 있다는 의미일 것이다. 검색기술과 인간행위자를 이분법 적으로 보는 대신에, 둘 사이의 관계방식에 초점을 둘 때 이 복잡한 상황이 해명될 수 있을 것이다. 그것이 설명될 수 있다면 우리는 검색기술에 대해 새 로운 전망을 내놓을 수도 있다. 검색엔진을 사용하는 사람은 이 기술이 어떻 게 작동하는지에 대해 관심을 가져야 하고 또 이 기술이 우리 사회에 어떤 의미인지도 알아야 한다.

2. 검색엔진의 합리성

웹은 공평하지 않다. 모든 정보가 웹으로 전환되는 것은 아니기 때문이다. 웹 이 되기에 좀 더 수월한 종류의 정보들이 더 빨리 그리고 더 많이 진입한다. 게다가 웹은 불확실하다. 인터넷에는 특별히 임명된 편집자도 없고 정보의 형 식이나 출처도 극도로 이질적이다. 더구나 전체 웹페이지의 양이 얼마나 되는 지도 모른다. 이같은 웹의 불확실성은 검색엔진에 의해 노출되는 색인과 순위 의 적합성을 흐드는 배경이기도 하다. 그럼에도 불구하고 검색엔진은 도메인 네임시스템만으로는 부족했던 웹접근성을 높이는데 기여해왔다.

1994년 야후(Yahoo)를 비롯하여 일련의 웹디렉토리 서비스들이 쏟아졌다. 이 방식은 검색엔진의 편집자가 수작업으로 웹페이지를 범주별 색인목록에 등록하면, 사용자들이 범주분류를 따라서 자신의 관심사를 점점 좁혀가다가 마침내 특정 웹사이트에 도달하도록 도와준다. 그리고 몇 년 후 “자연어 검색”이 등장했는데²⁾ 이것이 오늘날 익숙한 모습의 검색서비스이다.

자연어 검색에서는 검색엔진이 웹문서들을 자동으로 수집하고, 사용자는 단지 검색창에 몇 개의 일상 단어를 입력하기만 하면 그것에 가장 근접할 것으로 ‘자동-판단된 색인목록을 얻을 수 있다. 디렉토리 서비스에서 자연어 검색방식으로의 전환은 검색엔진의 큰 변화였다. 이를 사용자들 관점에서 본다면 사이트를 클릭하던 방식에서, 찾고자 하는 것을 단어로 바꾸어 질문하기만 하면 구체적인 웹페이지 결과들을 일목요연하게 볼 수 있게 되었다. 자연어 검색으로 인해 사용자들은 더 많은 질문을 하게 되었고, 더불어서 검색엔진을 통해 더 많은 것을 찾을 수 있다고 기대하게 되었다.

1) 검색엔진과 데이터베이스

이제 ‘웹상에 존재한다는 것’은 ‘검색엔진에 등재된다는 것’을 의미한다. 검색엔진은 키워드에 따라 웹페이지를 색인화하고 그 웹페이지의 내용을 키워드나 URL 단위로 연결하는 거대한 데이터베이스를 구축했다. 여기서 중요한 것은 검색엔진이 최소 단위로 키워드 개념을 사용한다는 것이다. 키워드는 색인 과정에서 웹페이지로부터 연역된 것이다. 검색로봇은 웹페이지 소스 코드 안에 포함된 제목 태그에서 적어도 한 개의 단어를 추출한다. 이것은 검색엔진이 수행하는 중요한 선택작업 중 하나이다.

예를 들면 검색엔진은 대중적인 주제를 처리할 때 검색엔진 색인을 웹문서

2) 자연어 처리(Natural Language Processing, NLP)란 1950년대부터 시작된 컴퓨터과학 및 인공지능 영역으로서, 인간이 만들어내는 의미와 자연어 입력을 컴퓨터가 이해하고 반응하는 것을 말한다.

의 제목으로 제한한다. 학술적 문서라도 색인은 몇 개의 키워드만이 추출될 뿐이다. 나머지 문서의 맥락은 배제한 채, 해당 문서에 고유한 이름(docID)을 부여한다. 검색엔진에 등재되길 원하는 웹운영자들은 검색엔진이 어떤 기준으로 색인하는지 그리고 얼마나 자주 재방문해서 정보를 수정하는지 궁금해한다. 그런데 검색엔진회사들은 대체로 구체적인 사항에 대해서는 기업비밀에 붙이고 있다(Halavais, 2005; Kammerer et. al, 2013).

데이터베이스는 사용자 데이터와 메타 데이터, 그리고 데이터베이스 수행을 개선하는 체계들, 응용 메타데이터들로 구성된다. 데이터베이스는 거대하기 때문에 원하는 관계항을 찾으려면 시간이 많이 걸리기 때문에 이를 개선하기 위해 색인을 만든다. 데이터베이스 관리시스템(DBMS)은 데이터베이스 내부구조를 보여주는 환경설정도구를 제공한다. 데이터베이스의 강력한 힘은 방대한 양의 데이터 배열을 질서 짓는 능력이고, 다양한 유형의 사용자 질의에 대해서도 각각 지원가능하다는 점에 있다.

검색시스템은 데이터베이스에 근거하여 작동하는데, 알고리즘이라는 숨겨진 ‘뒷단(back end)’과 사용자에게 경험되는 ‘앞단(front end)’으로 구성된다(Paul, 2007). 이 체계에서 뒷단은 비가시적이다. 또는 비가시적이어야 한다. 데이터베이스 알고리즘은 모든 유형의 정보에 논리를 부여하거나 데이터 수집을 걸러 내거나 데이터를 가상화하는 원리를 담아낸다. 그런 점에서 알고리즘의 형식은 지식과 신념과 사회적 행위의 양태를 드러내는 방식이다. 사용자의 눈앞에 드러나는 앞단은 그런 뒷단에 의해 생산된 것이다.

데이터베이스는 전형적으로 자기-기술적으로 자료들을 수집(self-describing collection)한다. 여기서 자기-기술적이라는 말은 데이터베이스 구조에 대한 기술이 그 데이터베이스 자체 내에 포함되어 있다는 의미이다. 데이터베이스는 항상 자기 자신 내부를 탐색함으로써만 의사결정을 할 수 있다. 때문에 자기 자신이 아닌 외부의 다른 곳을 탐색할 필요가 없다(Kroenke & Auer, 2011). 데이터베이스의 내용은 그 내부에서 보는 것에 의해서만 결정될 수 있다(Hayles, 2012). 그러므로 데이터베이스의 합리성은 전지전능을 요구하지 않는다. 다만

데이터베이스의 합리적인 선택은 지금까지 자신이 인지한 시퀀스에 의존할 뿐이다(Russell & Norvig, 2010).

2) 색인과 순위

검색엔진 알고리즘의 주요 역할은 색인하는 것이다. 알고리즘 개발자들은 인간의 관심과 가치를 기계적 알고리즘으로 전환하고자 한다. 그래야만 사용자의 검색의도를 만족시킬 것으로 예상되기 때문이다. 검색 알고리즘의 색인은 재현의 폭과 깊이에 따라 달라질 수 있다. 검색로봇이 깊이보다 폭에 더 초점을 둔다면 더 많은 웹사이트를 방문하게 될 것이다. 반면에 색인은 부분적으로만 하게 될 것이다. 대체로 검색엔진들은 웹포털과 같은 거대 사이트에 대한 색인을 할 때는 그 사이트의 내용 중 약 10-15%만 색인한다³⁾. 검색엔진이 모든 웹페이지를 색인하는 것은 아니기 때문에 웹운영자는 검색엔진이 더 많은 웹페이지를 색인하게 하려고 노력한다. 자신의 웹페이지가 검색엔진의 색인목표가 되도록 하는 것은 웹운영자들이 넘어야 할 첫 번째 장벽이다.

다음으로 웹운영자가 넘어야 할 두 번째 장벽은 자신의 웹페이지가 잘 보이도록 높은 검색결과 순위에 들어야 한다. 웹페이지들이 기하급수적으로 증가하면서 사용자들은 검색결과와 첫 페이지만 참조하는 경향이 증가했다⁴⁾. 이는 사용자들이 점점 기계적인 정보처리에 더 의존하고 있고, 그만큼 검색엔진의 자동화된 판단작업은 중요해졌다는 의미이다. 잘 알려진 순위결정 방식으로는 “백링크 계산(Backlink Count)” 방식과 “페이지랭크(PageRank)” 방식을 들 수 있다. “백링크”란 다른 웹페이지로부터 들어오는 링크를 말하는 것으로, 전체 웹에 출현하는 웹페이지에 연결된 링크의 수를 계산하는 것이다. 백

3) 반면에 검색로봇이 뛰어나더라도 웹사이트가 색인될 수 없는 경우도 있는데 그 중 하나는 웹운영자가 로봇배제규칙에 따라 “robots.txt”를 작성해 놓았다면 검색로봇은 그 사이트에 원칙적으로 ‘들어갈 수 없다’.

4) 검색결과 첫페이지만 보는 비율이 1997년 29%에서 2002년에는 73%로 증가했다 (Jansen & Spink, 2006).

링크 수가 많다면 그 웹페이지는 더 중요하다고 추론한다. “페이지랭크”는 학술 논문의 우수성 평가 방식을 참조한 것이다. 어떤 문서가 누군가에 의해서 참조되었다면 그 문서는 아마도 우수한 논문일 것이다(Introna & Nissenbaum, 2000).

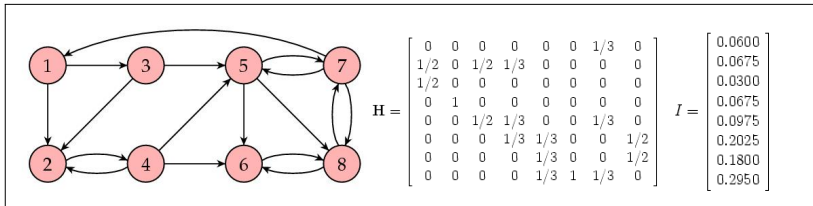
또한 순위 알고리즘은 키워드의 위치와 빈도도 참조한다. 예를 들면 웹문서 앞부분에 해당 키워드 출현빈도가 높으면 관련성이 높은 것으로 판단한다. 이처럼 검색엔진 알고리즘은 합리적이면서 자기발견적이다. 그러나 이런 방식은 스팸이나 광고성 웹에 의해 쉽게 흔들리곤 했다. 일부 웹운영자들은 더 많은 백링크를 보유한 것처럼 보이게 하려고 거짓 웹문서들을 만들어서 자신의 웹문서로 하이퍼링크를 한다. 그래서 그러한 기술적 속임수를 배제하는 것도 검색엔진의 중요한 역량이 되었다. 한편 더 높은 검색순위에 오르려고 하는 웹운영자들의 열망은 검색최적화(SEO, Search Engine Optimization)라는 새로운 지식 영역을 창출했다.

또 다른 순위 결정 방식은 순위를 구매하는 것이다. 특정 키워드에 대한 검색결과 순위에 대하여 적절한 가격을 지불하면, 경쟁자를 물리치고 일정기간 동안 높은 순위를 차지할 수 있다. 이것은 문서의 관련성과 공정성을 왜곡시킬 수 있다. 그러나 경우에 따라서는 오히려 검색품질을 높여주기도 한다. “꽃배달”과 같은 잠재적으로 상업적인 키워드는 비용을 지불한 웹운영자가 더 신뢰할만한 정보를 제공한다고 평가받고 있다. 이런 키워드들은 그 자체로 사용자의 검색 의도가 분명하게 포함되어 있기 때문에, 상업적 웹사이트를 우선순위에 둘 때 사용자가 만족할 가능성이 높다.

검색결과 첫 페이지에 노출되고 순위가 높다는 것은 더 많이 배포될 기회를 가지게 된다는 것이다. 다른 문서에 비해서 또는 경쟁하는 문서에 비해서 더 많이 배포된다면 더 많은 지배력을 가질 가능성이 있다. 최근 연구들에 따르면, 검색엔진 알고리즘이 문서의 적합성이나 관련성을 반영하지 않더라도, 사용자들이 그 검색결과 순위에 영향을 받는 것으로 알려지면서 검색엔진은 더욱 논쟁적인 주제가 되고 있다(Hargittai, 2007).

3) 기호적 질서를 부여하는 기술

검색엔진의 색인과 순위 결정과정에서 보듯이, 검색엔진은 웹문서를 발견하고 키워드를 추출하고 중요성이나 적합성을 판단하는데 있어서 가장 합리적인 방식을 따르고자 분투한다. 검색엔진은 이질적이고 다양한 텍스트들 사이의 관계를 측정하고 위계를 구성한다. 예를 들어 <그림 1>에서 보듯이, 8개의 문서가 있고 이들 사이의 연결관계를 추출했다면, 이것은 숫자로 전환되어 행렬(H)과 중요도(I)라는 수학적 도식 안으로 흡수된다.



<그림 1> 문서들의 하이퍼링크, 문서링크 행렬(H)과 문서의 중요도(I)

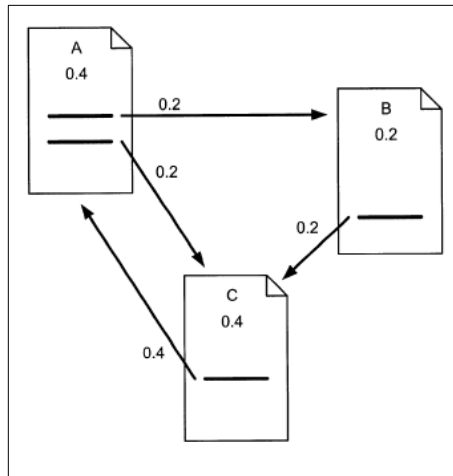
출처: Austin(2011).

여기서 중요도(I)는 고유벡터인데, 이 벡터를 구하는 방식은 다양할 수 있다. 5번문서와 6번문서는 링크 값의 합(7/6)은 같지만, 중요도(I)는 각각 0.0975와 0.2025로 다르다. 6번문서는 중요도가 높은 8번문서의 백링크를 받고 있기 때문이다. 중요도(I)의 수식은 다음과 같다. $I^{k+1} = H * I^k$ 첫 번째 중요도(I)와 행렬(H)의 곱이 다음 번 중요도(I)를 구성한다. 그래서 최초에 “1”을 어떻게 결정하느냐에 따라서 이 값은 계속 검색순위에 영향을 준다(Austin, 2011). 그러므로 다른 변수가 없다면 8번문서의 중요도(I)는 검색엔진이 상태를 갱신할 때마다 다시 참조되면서 강화될 가능성이 높다.

오늘날 검색엔진이 보유하고 있는 행렬(H)의 행과 열의 수는 각각 수백 만 개 이상이다. 따라서 검색결과 첫 페이지에 놓이기 위해서는 거대한 양의 링크를 보유하고 있어야 한다. 이처럼 검색엔진의 문서 관련성 평가기준은 고도

의 알고리즘적 합리성을 지향하고 있다. 검색엔진은 문서들 사이에 위계질서를 부여하는 기술양식이며, 그 질서양식에 의해 문서의 관련성이 검색엔진에 등재되는 순간, 그 문서들 사이의 관련성은 스스로 증폭되고 강화될 수 있다 (Hinman, 2005; Halavais, 2009).

그러므로 이렇게 검색엔진이 부여한 위계는 ‘객관적’인 것이 아니다. 검색엔진이 정의한 문서들의 중요도 또는 그에 따라 제시되는 순위는 본래적으로 존재하는 것이 아니다(Page, 1998). 그것은 단지 수집되어 들어온 문서들에게 검색엔진이 부여한 임의적이고도 자기충족적인 값일 뿐이다. 본래 문서들 사이에는 순위나 중요도가 없다. 즉, 8번문서가 6번문서 보다 더 ‘중요하다’고 볼 수는 없다. 단지 검색엔진이 부여한 질서 내에서, 그리고 검색엔진이 불러들인 문서들 사이에서만 그러하다.



<그림 2> 검색순위 결정 예시

출처: Page(1998).

<그림 2>에서 보듯이, 검색엔진 알고리즘은 수집된 문서들이 서로 순환적으로 보증되도록 값을 부여한다. 문서 A의 순위값은 문서 C의 순위값(0.4)과 같

다. 문서 A가 문서 C로 부터 유일한 백링크를 받았기 때문이다. 문서 A는 두 개의 백링크를 걸었기 때문에 자신의 순위값을 둘로 나눈 값(0.2)을 문서 C와 문서 B에 부여한다. 그래서 문서 A로부터만 백링크를 받은 문서 B(0.2)의 순위 값은 문서 A(0.4)의 절반 값과 같다. 문서 C의 순위값은 문서 B의 값(0.2)과 문서 A로부터 나오는 두 개의 백링크 중 하나의 값(0.2)을 합한 것이다.

$$r(A)=r(C)=0.4$$

$$r(B)=r(A)/2=0.2$$

$$r(C)=r(B)+r(A)/2=0.4$$

이처럼 문서들의 순위 값은 그 알고리즘 내에서만 타당하며, 수집된 문서들 사이에서만 설정된 관계이다. 그 외 어떤 다른 보증도 존재하지 않는다. 이런 상황은 검색엔진의 관심에 의해 더 강화된다. 검색엔진은 문서의 적합성을 더 잘 반영하기 위해 계속 노력하지만, 실제 더 중요한 관심사는 사용자의 만족도이다. 검색엔진은 사용자를 만족시키기 위해 다양한 유형의 질서양식을 개발해 낸다. 그 과정에서 효율성을 위해 종종 ‘길고 두꺼운 꼬리’ 부분을 잘라내곤 한다. 그래서 다수의 관심사는 선택되는 반면에 소수의 관심사는 선택되기 어려워진다.

결과적으로 검색 알고리즘은 사용자와의 상호작용 과정이 되풀이 될수록, 경향적으로 가장 인기 있는 웹문서를 집어내어서 높은 순위를 매기게 된다. 검색엔진은 사용자의 행위에 대하여 자기발견적으로 학습하며 그들의 만족을 추론하기 때문에, 대중적 관심이 되는 것을 더 민감하게 자신의 내적 도식으로 끌어들이고 있다. 그렇게 되면 검색엔진의 관점에서 볼 때 단지 현 상태를 반영하였을 뿐이지만, 일단 반영된 것은 빠른 속도로 재생산되면서 불균형이 강화된다. 이것은 일반적으로 데이터베이스들이 재귀적으로 상승하는 특질을 가지기 때문이다(Abiteboul, Hull & Vianu, 1994). 검색엔진이 흡수한 것은 전체가 아니라 특정한 기준에 따라 선택된 부분이다. 대체로 사용자들에게 매혹적

인 문서는 일상적이거나 가볍고 오락적인 것들이다. 이는 검색엔진회사가 광고에 의존하고 있다는 사실과도 관련이 있다.

3. 검색엔진의 자기증식

1999년 (주)지식발전소는 “엠퍼스”라는 서비스명으로 한글 자연어 검색을 선언했다. 비슷한 시기 네이버 역시 자연어 검색과 통합검색 방식을 개발했다. 오삼균·박희진(2000)에 따르면, 당시 한글 사용자들은 검색결과가 빈약하다는 불만이 있었는데, 특히 사회 정치 분야 보다는 연예, 오락, 음악, 생활 부분의 자료가 부실하다고 지적했다. 엔터테인먼트 분야에 대한 요구는 인터넷 사용자들이 주로 젊은 층이었기 때문일 것으로 보인다. 이러한 사용자층의 불균형은 이후 한글검색엔진의 발전에 누적적으로 영향을 준다.

한글검색엔진은 발전 과정에서 영어권 검색엔진과는 전혀 다른 역돌출부(reverse salience)에 직면했다. 당시 한글검색엔진은 아무런 검색결과도 제시하지 못하는 사례가 많았다. 검색엔진 자체의 역량 부족도 있었지만, 한글 웹문서가 매우 부족했기 때문에 검색엔진은 자신의 역량과 관계없이 무능해 보이는 경우도 많았다. 게다가 한글검색엔진은 한글 형태소 분석이라는 고유한 난관을 만난다. 이 때문에 한글검색결과는 사용자의 검색의도와는 전혀 다른 것을 보여주기 일쑤였다. 이는 검색기술만의 문제가 아니라 한글검색엔진이 극복해야 할 시대적이고 지역적인 맥락이었다.

1) 한글과 검색엔진의 적응

초기 검색엔진이 직면한 한글 검색의 난제는 한글로 작성된 웹문서의 부족이었다. 당시로서 이 문제는 ‘보다 뛰어난’ 검색 알고리즘을 개발해야 하는 문제보다 더 시급했다. 1999년 당시 한글 웹 사이트 개수는 약 5만 여개 정도였는

데, 검색할 만한 유용한 정보가 있는 웹사이트는 그보다 훨씬 더 적었다. 이는 검색엔진 기술 자체의 문제는 아니었지만, 검색엔진 기술이 성장하는데 있어서 극복해야 할 핵심적인 문제였다. 곧 검색회사들은 검색기술을 개발하는 것만이 아니라, 그들이 수집해야 할 대상(데이터)도 스스로 만들어내야 한다는 것을 알게 되었다.

결국 검색회사들은 스스로 한글 데이터들을 디지털로 전환하기 시작했다. 가장 최근의 정보처리 방식에서부터 가장 오래된 정보처리 방식들에 이르기 까지 데이터가 될 수 있는 것을 모두 자기 내부로 끌어들이었다. 신문서비스나 사전(辭典)서비스가 대표적인 사례이다. 한글검색엔진들이 검색결과에 신문기사와 사전폴이를 기본으로 제공하게 되면서, 소위 “빈 페이지 문제”로부터 벗어나기 시작했다. 아직 관련 웹문서가 없는 질의어가 입력되더라도 그에 해당하는 사전적인 낱말폴이를 보여줄 수 있었다(김지연·김성희·정혜경, 2011).

2000년 시작된 통합검색 서비스 역시 부족한 한글자료를 보완하기 위해 시작된 것이었다. 이것은 웹페이지를 나열하는 검색방식이 아니라 카페, 뉴스, 디렉토리, 쇼핑, 이미지, 동영상, MP3, 사전 등 섹션별로 노출해주는 방식이다. 예를 들어 “핑클”이라는 질의어가 입력된다면 사용자의 의도는 다양할 것이다. 핑클 멤버의 프로필을 알고 싶을 수도 있고, 사진을 보고 싶거나 팬클럽을 찾고 싶을 수도 있다. 또는 핑클의 노래를 알고 싶을 수도 있다. 프로필을 알고 싶다면 블로그 범주로, 음악을 듣고 싶다면 동영상이나 MP3 섹션을 보는 것이 원하는 정보를 찾을 가능성을 높다. 각 범주별로 콘텐츠의 성격이 달랐기 때문에 웹문서를 일렬로 나열하기 보다는 범주별로 분류하여 묶어서 보여준다면 사용자들이 원하는 유형의 범주를 먼저 선택한 다음 구체적인 세부 콘텐츠를 찾으면 되기 때문에 시간을 절약할 수도 있다(윤정근·임일, 2009; 박정아, 2012). 통합검색 방식은 부족한 한글 데이터를 ‘알뜰하게’ 사용하기 위한 노력이었고 상당히 성공적이었다(김상호 네이트 검색팀장, 2013-03-02).

통합검색 알고리즘은 섹션별 클릭횟수에 의존하여 섹션을 배치하기 때문에 각 키워드마다 섹션 배열이 다를 수 있다. 해당 키워드 사용자가 평균적으로

많이 클릭한 문서의 섹션은 상위에 노출되고 그렇지 않은 경우는 하위에 노출되거나 아예 통합검색 결과에 노출되지 않는다. 이 알고리즘은 클릭횟수와 문서 적합도가 비례한다고 전제하는 것이다. 그러나 사용자들의 의도는 다양할 수 있는데, 검색알고리즘 상으로 적합해 보이는 문서지만 실제로는 질의와 무관한 문서들도 존재했다(박소연·이준호, 2007)⁵⁾.

한글검색엔진이 직면했던 또 다른 주요 문제는 형태소 분석의 어려움이었다. 검색엔진들은 주로 형태소를 분석하는 방식으로 색인한다. 대체로 명사가 문장의 주요 의미를 나타낸다고 보고, 문서 내 키워드들의 빈도를 조사하고 가중치를 부여하여 찾기 쉬운 형태로 조직하고 저장했다. 그런데 한글문서의 경우 단순히 명사들의 배열 관계로만 문서의 내용을 적절하게 대변하는데 한계가 있었다. 영어권 자연어 문장들은 단어들에 주로 공백으로 분리되어 있어서 이를 중심으로 형태소 분석에 대입할 수 있었던 반면에 한글에서는 그렇지 못했다. 한글은 단어의 경계가 명확하지 않고, 문법적 기능을 하는 각종 조사 및 접사들에 의해서도 의미가 완전히 달라지곤 했기 때문이다. 그래서 단지 형태소 분석만으로는 색인으로서 제 역할을 하지 못하는 경우가 많았다.

이러한 한글의 특성으로 인해, 한글검색엔진 개발자들은 상당한 정도의 추가적 노력을 들이더라도, 정작 사용자는 검색의도와 다른 문서자료를 만날 가능성이 높았다. 이런 장벽을 넘기 위해 검색엔진 개발자들은 가능한 여러 가지 유형의 형태소 분절을 중첩적으로 고려해야 했다. 예를 들어 “먹이”라는 단어는 (1) 동물의 식량으로서 “먹이(명사)”는 물론이고 (2) 서예 도구인 “먹(명사)”과 조사 “이”의 결합어로서, 그리고 (3) “먹다”라는 동사의 변형으로서 “먹”과 사역형 “이”와 현재진행형 어미 “는”의 결합어로서도 예측하여 문서내

5) 그럼에도 불구하고 사용자들의 행위양식은 통합검색 방식에 적응하여 빠르게 변했다. 이준호·박소연·권혁성(2003)의 연구는 최초 질의에서 통합검색을 선택하는 비중이 99.94%에 이른다고 보고하고 있다. 사용자들은 다른 검색방식을 시도하는 일은 거의 하지 않고 있다. 또한 사용자들은 검색을 하는데 그다지 많은 시간과 노력을 들이려고 하지도 않았다. 사용자들의 전체 질의 쿼리 중 최초 쿼리 비중이 55.3%에 이르렀고, 세션 당 한 개의 질의만 하는 비율이 67%로 나타났다(Park, Lee & Bac, 2005; 박소연, 2011).

용을 분석했어야 했다(윤성희, 2003). 그런데 이런 한글 형태소 분석의 난관은 한글 키워드를 다루는 검색엔진 개발자들의 역량을 강화했다. 영어권 검색엔진과 달리 한글검색엔진에서만 나타났던 여러 가지 새로운 질서형식을 불러일으키는 기초로 작동했다.

2) 질의-응답 시스템: 현장지식을 흡수하는 기술형식

2000년 초 검색회사들은 한글 데이터를 빠른 시간에 생산해낼 새로운 방법을 더 찾아야만 했다. 한글검색엔진 개발자들은 사용자들이 던지는 ‘기이한’ 검색요청에 응답할 데이터가 ‘기술적으로’ 존재하기 어렵다는 문제에 직면했던 것이다. 사용자들이 원하는 상당수의 검색요청들은 현장맥락적인 것이었는데, 검색엔진은 그런 현장맥락을 처리하기 어려웠다.

“질의 결과가 “0”건이 나오는 것들을 가져다 놓고 고민했는데... 사용자들은 “우리 동네 ○○슈퍼 언제 문 다시 열어요?” 같은 질문을 해요. 처음에는 “이거 미쳤나?”라고 생각했죠. “이게 검색엔진에게 물어볼 질문인가?” 하지만 답답하니까 물어본 거잖아요. 사람들은 그냥 질문한 것예요. 그런데 이런 거는 기술로 할 수 없어요. 질의한 걸 보면 사용자가 원하는 것이 뭔지 알겠는데, 그리고 누군가는 알고 있는 사람도 있을 텐데.....” (이해진, NHN CSO, 2012-11-22, 인터넷역사프로젝트 인터뷰 중에서)

네이버의 지식iN(2002), 다음의 아고라(2004)와 같은 서비스들은 그에 대한 대응으로 부상했다. 네이버의 지식iN 서비스는 검색데이터의 최신성은 물론이고 사용자의 검색의도에 가장 근접한 결과를 제시할 가능성을 극적으로 높였다. 사용자들 사이의 질의-응답 형식은 단지 더 많은 검색결과와 재현을 가능하게 했을 뿐만이 아니라 현장에 근접한 정보요구에 부응할 수도 있었다⁶⁾.

6) 사용자의 질문이 적절한 잠재적 답변자에게 도달하기만 한다면, 기계적으로 수집·분류된 목록결과보다 더 적합도가 높은 답을 구할 가능성이 있다. 인간-답변자는

유비적으로 보면, 이 형식에서 사용자의 질문은 검색요청 행위에 해당하고, 사용자의 답변은 검색결과 화면에 해당한다. 이것은 마치 놀이와 같이 진행되었다. 질의를 한 사용자는 누군가의 답변을 기다리고, 답변들 중에서 가장 만족하는 것을 '답'으로 채택한다. 이를 구경한 다른 사용자들도 답변의 적절성을 평가하여 추천할 수 있다. 그리고 검색엔진은 그들의 질의-응답 문서를 색인했다. 그렇게 되자 검색엔진은 일상적으로 궁금한 모든 것에 답을 주는 것이 가능해졌다.

다음의 아고라 서비스는 약간 다른 형태로 이런 문제에 부응했다. 이 서비스는 누군가가 주장을 제기하면, 그에 대해 다른 누군가가 찬성과 반대 의사를 포함한 의견을 덧붙일 수 있는 전용 게시판 형식이다. 그 글을 읽는 다른 사용자들은 '찬성'이나 '반대' 투표에 참여할 수 있다. 해당 글의 조회수/찬반 투표수에 따라 베스트 게시글로 등재되었다. 아고라 시스템은 다수의 '논객'과 열광적 '독자'들을 만들었다. 누군가의 질의에 답변하는 지식iN 방식과는 다르지만, 아고라 논객들은 동시대 사람들이 궁금해 할 주제들에 대해 논했다는 점에서 아고라 서비스 역시 일종의 '질의-응답 시스템'이다.

아고라서비스를 유명하게 만든 계기는 미네르바 사건이다. 아고라 아이디명 "미네르바"는 2008년 약 6개월 여간 총 278개의 글을 아고라 경제방 게시판에 등재했다. 그의 글은 매우 열광적으로 구독되었는데, 그 이유는 그의 글이 동시대 사람들이 잠재적으로 묻고 싶었던 문제를 제기하고 있었기 때문이다(김지연, 2012). 아고라 논객들과 사용자들 사이의 의견교류는 즉각 웹문서가 되었고, 그것은 다시 검색엔진에 의해 분류·색인되어, 검색을 위한 데이터 베이스 속으로 들어갔다.

이렇게 검색엔진은 인터넷 사용자 전체를 '답변자'로 동원하는데 성공했다.

기계-답변자 보다 인간-질의자의 의도와 맥락을 더 잘 이해할 수 있기 때문이다. 실제로 주요 검색서비스의 검색적합도 비교평가(박소연 외, 2008)는 네이버 검색의 적합도가 가장 높았고, 그 중에서 강세를 보이는 분야로 지식iN 서비스를 꼽고 있다.

달리 말하자면, 그들은 사용자들을 동원하는 새로운 형식을 발명해냈다. 이 형식을 통해 사용자들은 매일 엄청난 양의 웹문서를 기꺼이 생산해냈다. 결과적으로 한글검색엔진은 인간-질의자의 물음을 충족시키기 위해서, 인간-답변자를 동원하여 더 많은 자료를 생산하는데 성공했을 뿐만 아니라 더 현장에 가까운 지식과 더 전문적인 정보를 흡수하는데 성공했다. 무엇보다도 중요한 것은 그들이 질의-응답시스템을 통하여 생산된 자료에 검색엔진의 질서를 부여했다는 점이다. 사용자들이 생산한 문서는 곧장 검색엔진에 의해 평가받고 위계지어 졌다. 이제 사용자들은 자신이 생산한 자료이더라도 검색엔진의 질서를 통해서 호출되는 것을 지켜보아야 했다. 검색엔진과 사용자들은 그 ‘암묵적 계약’에 동의했다.

3) 연관 검색: 사용자의 관심에 질서를 부여하는 기술

현재 한글검색엔진에 입력되는 일일 질의어 수는 수 억 개에 넘을 것으로 추정된다. 대부분의 검색엔진은 질의어의 빈도와 양을 관찰하고 빈도가 높은 질의어 위주로 내용이나 특징을 분석함으로써 정규성을 규정한다. 질의어의 정규성은 검색엔진 관리자가 서비스 방향을 결정하는 중요한 근거가 된다.

“키워드마다 가중치가 있어요. 많이 들어오는 키워드.. 사업적으로 그것에 집중할 수밖에 없어요. 좀 더 디스플레이에 신경 쓰고 좀 더 많은 정보를 한꺼번에 보여주려고 노력합니다.... 영화타이틀이라면, 비즈니스랑 연결되기도 하고 회사마다 특징을 살리려고 하는 것도 있기 때문에... 처음에는 감독이나 배우, 포스터, 런닝타임 정도를 보여주다가 시간이 가면 극장정보에다가 원작자도 나오고... 나중에는 케이블방송 정보도...” (김상호 네이트 검색팀장, 2013-03-22)

질의어 통계정보 중 일부는 “연관검색어” 또는 “인기검색어”와 같은 대중적 서비스로 구현되었다. 2004년, 다음은 검색서비스의 만족도를 높이기 위한

방법의 하나로 연관 검색어를 배치했다. 사용자가 “지하철”이라는 질의어를 입력하면 “지하철노선”, “지하철시간표”, “서울지하철”, “버스” 등과 같이 관련이 있을 것으로 추정되는 다른 질의어들을 함께 제시해주는 것이다(대한민국 공개특허 10-2005-0079845, 박정아 다음 검색팀장, 2013-04-05). 만약 검색결과가 만족스럽지 않다면, 사용자들은 검색엔진이 제시해주는 연관 검색어를 참조하여 추가적인 검색을 이어갈 수 있다.

연관검색어는 질의어들 사이의 연관도를 측정된 결과를 보유하고 있어야만 가능하다. 연관검색어 추출방식은 입력되는 질의어들을 시간적으로 관측하여 일정한 질서를 부여하는 것이다. 단어 연관도 측정 시스템은 발생빈도 또는 교차빈도에 기초하여 빈도 정보를 생성한다. 발생빈도와 교차빈도를 통하여 단어-문서분류 발생빈도, 문서빈도, 분석분류빈도 그리고 단어-단어분류 교차빈도 및 단어-단어 결합빈도도 생성할 수 있다. 이렇게 단어통계정보를 생성하고 벡터, 랜덤변수, 결합 확률분포 등을 이용하여, 단어 사이의 관계를 수치화함으로써 단어 연관도를 결정한다.

〈표 1〉 단어 연관도 행렬 예

	단어1	단어2	단어3	단어4
단어1	3	3	3	1
단어2	3	4	4	1
단어3	3	4	7	3
단어4	1	1	3	3

출처: 엔에이치엔(2006).

검색엔진은 사용자들로부터 입력되는 매 순간의 질의어의 동향을 주시함으로써 다음 순간의 질의어의 동향을 관측할 수 있었다. 어떤 사용자가 하나의 질의어를 사용하고 이어서 다른 질의어를 사용했다면 연속하여 입력한 질의어들 사이의 연관성이 높다고 추정하는 것이다. 다른 사용자가 동일한 질의어

를 입력하는 순간, 검색엔진은 이전 사용자들의 질의 패턴을 참조하여 연관 단어를 제시해 준다. 그러면 사용자들은 적절한 질의어를 떠올리는데 도움을 받을 수 있고 최종적으로 검색의 성과가 있었다고 판단할 가능성으로 유도될 것이다.

2005년 시작된 네이버의 "실시간 인기검색어" 역시 사용자들이 입력한 질의어들에 대해 일정한 질서를 부여하는 기술이다. 이 통계정보는 이용자들이 던지는 질의어를 "거의 실시간"으로 집계하여 10위까지 표시해준다⁷⁾. 좀 더 엄밀하게 말하면 실시간 인기검색어는 많이 입력된 단어의 순위가 아니라, 그 순간 입력량의 증가폭이 큰 순위를 보여준다⁸⁾. 예를 들어 "내일 날씨", "지하철 노선도" 등과 같은 검색어는 일상적으로 많이 입력되는 단어지만, 평상시에는 입력 횟수 추이에 큰 변화가 없기 때문에 실시간 인기검색어에 나타날 가능성이 적다. 그러나 태풍이나 첫 눈 소식이 있는 날에는 날씨 관련 질의어가 평상시 보다 많이 검색될 것이므로 인기검색어 순위에 등장할 수 있다.

검색엔진은 평소 특정 검색어에 대한 평균적 입력횟수를 지속적으로 산정하여 저장하고 있다가 매 순간 평균적 범위를 넘어서게 되는 순서대로 해당 단어에 순위를 부여한다. 즉, 실시간 인기검색어는 그 순간 '비정상적으로 많이' 검색된 단어이다⁹⁾. 이는 본래 검색품질을 관리하기 위한 것이었다. 실시간

-
- 7) "거의 실시간"이란 상당히 짧은 간격을 의미하며, 네이버의 경우 15초를 말한다. 시간간격은 변할 수 있다.
 - 8) 실시간 인기 검색어에 대한 높은 관심 때문에, 네이버 검색시스템은 혼동을 줄이기 위해 2007년 이름을 "실시간 급상승 검색어"로 서비스 명칭을 변경했다.
 - 9) 대한민국 공개특허(00522029)에 따르면, 실시간 급상승 검색어 추출방식은 다음과 같다: ① 사용자 단말기로부터 입력횟수가 급상승하는 검색어를 검출하기 위해서, 소정의 시간 간격마다 하나 이상의 검색서버로부터 로그 데이터를 수집, ② 수집된 로그 데이터를 이용하여 검색어별 입력횟수를 산정하고, 선정된 검색어별 입력횟수를 각 타임데이터와 연관하여 데이터베이스에 저장, ③ 데이터베이스를 참조하여 소정의 슬라이딩 윈도우에 속하는 하나 이상의 타임데이터를 식별하고, 식별된 타임데이터와 연관된 질의어별 입력횟수를 각 검색어별로 합산, ④ 합산된 입력횟수에 기초하여 상기 각 검색어별로 추정검색횟수를 연산, ⑤ 합산된 입력횟수에 기초하여 상기 각 검색어별 입력순위를 결정하고, 상기 결정된 입력순위를 이용하여 검색어별 판정 수치를 연산, ⑥ 판정 수치에 기초하여 실시간 급상승 검색

인기검색어는 사용자들의 검색 유형을 패턴화하여 미래 사용자들의 관심과 흥미를 포착함으로써 사용자 만족도를 높이려는 것이다¹⁰⁾. 검색품질을 유지하기 위해서 모든 질의어를 동등하게 관리하기 보다는 상대적으로 관심이 증폭되는 질의어를 중심으로 검색품질을 개선하는 것이 검색엔진의 관점에서 효율적이었을 것이다. 이제 검색엔진은 시시각각 변이하는 사용자들의 관심에 대해서도 질서를 부여하는 방법을 발견한 것이다.

4) 실시간 검색: 사용자와 검색엔진의 상호적 강화

검색엔진은 문서들 사이에 자신의 위계질서를 부여할 뿐만 아니라 단어들 사이에도 자신만의 위계질서를 부여했다. 그를 통해서 사용자들의 검색의도를 추론할 수 있었다. 이것은 검색엔진의 관심에도 부응했다. 사용자들이 주목하는 인기가 많은 단어는 비즈니스의 대상이 될 수 있기 때문이다. 본래 인기검색어는 흥미를 불러일으키기 위한 가벼운 장치였지만 이 목록에 오르는 단어는 사람들에게 높은 주목을 받았다. 어떤 단어가 인기검색어라고 선언되는 순간, 그 단어는 더 많은 관심을 받게 되고 그 자체로 일정한 힘을 얻었다¹¹⁾. 검색엔진은 사람들 사이에 인기를 얻은 단어의 등장을 단지 알려주기만 했을 뿐인데도 그것은 추가적인 변형을 불러올 위력을 가지게 된다.

실시간 검색어의 위력을 단적으로 보여주는 사례는 <표 2>와 <표 3>에서 찾을 수 있다. 각각 네이버와 구글에서 “박근혜”, “문재인”, “안철수”, “이정희” 키워드의 2012년 트렌드를 비교한 것이다. 한글검색에서 네이버의 점유율(76%)이 구글의 점유율(4.5%) 보다 훨씬 높기 때문에 한국 사용자들의 정규성을 더 잘 반영하고 있을 것으로 추정된다¹²⁾. 네이버 트렌드에서 “박근혜”와

어를 추출한다.

10) 네이버 트렌드 연감(2008; 2009)을 참조하라.

11) 주목받는 단어는 더 많은 추가적인 검색요청의 대상이 되고, 검색광고 시장에서 더 높은 가격에 거래될 수 있다.

12) 비즈스프링 2012년 당시 인터넷트렌드 자료를 참조했다.

“문재인” 키워드 비율이 대통령선거에서 두 후보 사이의 득표율과 상당정도 부합했다는 점에서도 알 수 있다.

<표 2> 네이버 트렌드 키워드 비교(2012)

	박근혜	문재인	안철수	이정희
11.19~11.25	11	20	26	1
11.26~12.2	12	12	10	10
12.3~12.9	15	13	22	100
12.10~12.16	14	13	30	14
12.17~12.23	57	34	8	20
12.24~12.30	7	4	2	5

<표 3> 구글 트렌드 키워드 비교(2012)

	박근혜	문재인	안철수	이정희
11.18~11.24	18	17	43	1
11.25~12.1	35	20	16	3
12.2~12.8	46	22	19	38
12.9~12.15	51	31	11	20
12.16~12.22	100	43	11	19
12.23~12.29	24	7	4	5

두 회사 모두 키워드 트렌드 분포를 재현할 때, 해당 선택 기간 내에서 키워드들 중 최대 검색량을 보유한 키워드의 해당 시점을 “100”으로 지정하고, 나머지 키워드 검색량을 상대값으로 계산하여 표시한다. 2012년 대통령 선거 기간 중, 이정희 후보는 토론방송에서 박근혜 후보를 맹렬히 공격하여 주목받았다. 그것을 반영하듯이 “이정희” 키워드는 트렌드 세 번째 주간에서 각각 100과 38로 급격한 상승을 나타내고 있다.

“이정희” 키워드는 전 기간에 걸쳐서 낮은 값을 가지고 있고, 네이버와 구글 모두에서 최소값(1)을 보유하고 있다. 그가 매우 낮은 지지율을 가진 군소 정당 대표였으므로 낮은 키워드 값을 가지는 것은 현실의 상황을 모사했다고 할 수 있다. 그러나 네이버 트렌드 상에서 특정 기간(12월 3~9일) 동안 “이정희” 키워드가 최대값(100)을 차지하면서 ‘당당히’ 다른 모든 키워드의 기준값으로 나타나는 것은 이례적인 것이다. 박근혜와 이정희 두 인물에 대한 지지도를 감안한다면, 네이버 트렌드 상에서 “이정희” 키워드의 최대값(100)은 검색엔진이 제시한 실시간 인기검색어에 대한 사용자의 반응을 포함하고 있는 것이다.

“실시간 검색어에 오르면 그래요... 모르고 있다가도... 보이니까 검색하잖아 요... 영향을 안 미칠 수는 없어요. 실시간 검색어가 원래 그런 거니까...그러나 검색시스템의 영향이라기보다는 포털의 영향이라고 하는 게... 실시간 검색어는 검색서비스라기 보다는 포털 서비스이니까.....” (박정아 다음 검색 팀장, 2013-04-05)

실시간 검색어 노출은 특정 키워드를 향한 사용자의 관심을 추가적으로 촉발함으로써 검색유입량을 자극할 수 있다. 이로써 단어관계 행렬 상에서 해당 키워드의 위계는 달라질 것이다. 어떤 키워드가 실시간 검색어에 등장할 수 있다면, 극히 소수의 사용자만이 관심을 보이던 주변적 키워드이더라도 전체 사용자의 주목을 받을 수 있는 것이다.

이런 주변적 키워드의 부상은 누군가에게는 의미 있는 일일 것이고 다른 누군가에게는 문제로 해석될 수도 있다. 그럼에도 불구하고 실시간 검색어 기술은 다양성을 포섭할 수 있는 가능성도 제시하고 있다. 일반적으로 검색엔진이 부여한 단어들 사이의 위계질서는 각 질의어의 유입총량만을 고려하기 때문에 사용자의 관심의 정도가 큰 키워드에만 집중된다는 취약점이 있다. 그런데 실시간 검색어는 평균적 유입총량은 적더라도 순간적으로 유입총량의 증가가 두드러지는 단어의 존재를 알려준다는 점에서 검색엔진의 새로운 시도

이다. 결과적으로 실시간 인기검색어는 일시적이긴 하더라도 소수의 관심을 다수의 관심으로 전이할 수 있다.

4. 토론 및 결론

1) 검색엔진은 사용자의 명령을 따르는 것이 아니라 시스템 자신의 해석을 따른다.

모든 검색엔진은 사용자의 명령을 잘 이해하고 더 만족할만한 결과를 수행하기 위해 지속적으로 노력하지만, 그럼에도 불구하고 웹문서와 사용자 사이를 ‘투명하게’ 연결하는 것은 아니다. 본래적으로 문서들 사이에는 어떤 순위나 중요도도 존재하지 않으므로, 검색엔진이 부여한 위계는 ‘객관적’이라기보다는, 검색엔진 알고리즘 내에서만 충족적인 관계이다. 검색엔진 알고리즘에 따라 최초에 설정한 중요도(I)는 다음 번 중요도(I)를 갱신할 때 호출되어 다시 참조되며, 문서와 문서 사이, 단어와 단어 사이의 순환적 연관성에 의해 상대적인 순위값이 부여 된다. 그리고 이런 그물망이 반복적으로 불려나오면서 어떤 것들 사이는 덩어리처럼 단단해진다. 이처럼 검색엔진 알고리즘이 부여한 질서양식은 자기완결적이고 재귀적으로 그리고 상승적으로 자기 자신을 강화한다. 검색엔진이 관측한 정규성은 사용자의 관심 그 자체라기보다는 검색엔진이 사용자의 관심이라고 해석한 것이다.

이런 현상은 키워드의 정규성을 측정할 때 발생한다. 수 백 만개의 일간 질의어 중에서 서로 다른 질의어 수는 20-25만개 정도인데, 이중 최대빈도 일간 질의어 비율은 보통 2-4%에 불과하다. 나머지 많은 수의 질의어는 한두 번 정도만 입력된다. 가장 많이 입력되는 상위 1만개의 키워드가 차지하는 비중은 전체 입력 키워드의 극히 일부일 뿐이다(박상규 외, 2007). 검색엔진은 최대빈도

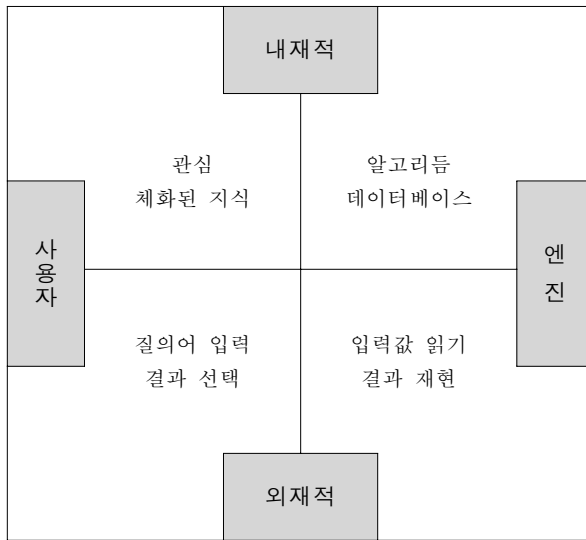
를 나타내는 질의어들 쪽으로 무게 중심을 이동하고 다수의 ‘빈도가 낮은 질의어’는 추론의 효율성을 위해 폐기한다. 검색엔진에게 있어서 사용자의 질의어의 유입량은 ‘사용자의 관심의 양을 의미한다. 검색엔진 데이터베이스 상의 ‘단어의 위계는 ‘관심의 위계’로 해석된다. 선택된 키워드는 더 많이 참조되고 강화될 수 있고 선택되지 않은 키워드는 그럴 기회를 잃는다.

실시간 검색어 역시 검색엔진 자신의 내적 강화 과정이다. 검색엔진 내부로 들어온 키워드는 매번 이전의 평균적 입력값과 비교하여 평가되고 배치되기를 반복한다. 검색엔진은 특정 순간만이 아니라 그 검색어에 대한 과거의 입력행위 전체를 응축적으로 재현하여 표시해준다. 그리고 그 다음 순간 다시 그 자신의 값을 참조하면서 자신의 질서양식을 반복한다. 실시간 인기검색어는 ‘객관적 사실’이라기보다는 검색엔진의 질서양식에 의해 ‘생산된 사실’이다. 자동화된 규칙을 통해 판단함으로써 적은 비용으로 사용자의 요구에 반응할 수 있었다. 이것이 이질적이고 무질서한 웹문서와 사용자들의 관심을 제어하는 검색엔진의 기술이다.

그러므로 ‘사용자의 의도’ 자체와 검색엔진이 ‘그의 의도라고 해석하는 것’ 사이에는 차이가 존재한다. 인간의 의도 또는 관심은 수학적으로 계산되지 않거나 잘 분절되지 않는다. 그런데 검색 알고리즘은 계산가능하고 분절가능한 단위로 그것을 전환하고 구조화하여야만 작동가능하다. 사용자들이 어떤 질의어를 입력할 때, 또 어떤 문서를 클릭할 때 검색엔진은 자신이 분절할 수 있는 단위양식으로 그 행위를 포착하고 저장하며, 다시 호출하거나 다른 데이터와 결합하여 새로운 해석을 도출한다. 검색엔진은 그 모든 과정에서 자신의 규칙에 따라서 그리고 자신이 구성한 질서양식 내에서 스스로 판단하고 선택한다. 검색엔진은 단순한 도구가 아니다. 오히려 자신의 내적 질서를 통하여 ‘거짓(Fiction)’을 생산할 역량을 가지고 있다.

2) 검색엔진은 사용자의 관심을 흡수함으로써 자신의 전문성을 강화한다.

검색엔진은 자신이 부여한 질서양식 안에서만 입력된 명령을 해석할 수 있다. 검색엔진은 수집된 문서에 위계를 부여하는 기술, 입력된 단어들 사이에 위계를 부여하는 기술, 그리고 그 위계를 다시 참조할 수 있는 기술양식이다. 그럼에도 불구하고 검색엔진의 내적 질서는 홀로 형성된 것은 아니다. 검색엔진은 사용자와의 상호작용을 통하여 매 순간의 사용자의 관심을 추출하고, 자신의 내적 질서 속으로 흡수하여 검색결과라는 ‘사실’을 생산한다.



<그림 3> 검색엔진과 사용자의 작동

그런데 사용자의 관심은 사용자의 내재적 작동이므로 검색엔진은 그것을 직접 알 수 없다. <그림 3>에서 보듯이, 다만 사용자가 특정 단어를 검색엔진으로 투입하거나 반대로 검색결과목록 중에서 어떤 것을 선택할 때만 검색엔진에게 포착될 수 있다. 검색엔진은 사용자가 입력하는 여러 가지 값들-입력되

는 단어나 검색결과를 선택하는 클릭을 받아들여 사용자의 관심을 추론한다.

물론 그 추론 과정은 검색엔진 자신의 내적 알고리즘에 의존한다. 한편으로 검색엔진은 사용자의 외적 행위를 내재화하기 위해 자신의 알고리즘과 데이터베이스를 꾸준히 개선한다. 검색엔진과 사용자 사이의 상호작용이 거듭 될수록 검색 알고리즘 내에는 사용자의 명령을 통하여 추론된 ‘사용자의 관심’이 재귀적으로 집적될 것이다. 검색 알고리즘의 재귀성은 실시간 인기검색어에서도 볼 수 있다. 검색엔진은 실시간 검색어를 생산하고 배포함으로써 다시 자신의 검색총량에 변동을 준다. 즉, 검색엔진이 외재적 요소(사용자 관심)를 단지 그대로 재현하는 것이 아니라 검색엔진이 재현하는 것 안에는 검색엔진 자신의 작동이 포함되는 것이다. 그것은 일회적 재현으로 종결되지 않고, 다음 번 순환에 영향을 주며 강화된다. 이것은 검색엔진 알고리즘의 일반적인 특질이며, 검색엔진의 합리성을 구성한다.

그러한 합리성은 몇 가지 문제를 불러일으킬 수 있다. 가장 먼저 문제될 수 있는 것은 검색엔진의 대중주의이다. 검색엔진은 무엇보다 사용자의 대중적 관심에 민감하다. 한국정보사회진흥원 보고서(2008)에 따르면, 검색서비스의 영향력이 가장 큰 영역을 묻는 항목에 대해서 사용자의 67.3%가 “엔터테인먼트/문화 분야”라고 답했다.¹³⁾ 이는 불과 몇 년 전 사용자들이 가장 불만스러워했던 문제였다. 검색엔진은 지난 수년 동안 사용자들이 특히 많은 관심을 보이는 것을 가장 열심히 강화하려 노력한 것이다. 검색엔진은 대다수를 차지하는 극소수의 질의어들에 대해 노력을 기울이기 보다는 상대적으로 관심이 몰려있는 가벼운 오락 관련 질의에 대한 검색결과를 풍부하게 재현하기 위해 더 집중적으로 노력했다. 이는 검색엔진이 본성적으로 체화하고 있는 효율성이기도 하다.

다음으로 문제될 수 있는 점은 검색엔진들이 경향적으로 논쟁이 없는 내용을 선호하게 된다는 것이다. 그래서 논쟁적인 사건은 검색결과로 뿌러지지 못

13) 2006년 엔터테인먼트에 관한 질의 비중은 23%로 전체 질의어 중 가장 높은 비중을 차지하고 있다(박소연, 2011).

할 가능성이 점점 높아지고 있다. 법률에서 명시하고 있는 “온라인서비스사업자의 의무”는 온라인서비스사업자에게 직접적인 책임을 묻는 대신 문제 콘텐츠의 삭제고지 의무를 부여함으로써 면책특권을 주고 있다. 명예훼손이나 저작권에 관한 문제가 발생했을 때 검색회사는 링크를 삭제하거나 해당 문서의 주인에게 이를 고지하는 일련의 프로세스를 진행하는 것으로 최소한의 의무를 수행한 것으로 면책된다(황성기, 2007). 그런데 대기업이나 정부기관들과 같은 권력주체들은 자신들에게 불리한 내용이 어디에 있는지 모니터링을 할 여유가 있기 때문에 이들의 삭제요청이 급증했다. 따라서 권력주체들에 대한 비난과 비판은 검색결과에서 사라지기 시작했다.

상대적으로 아무도 문제제기 하지 않는 내용들로 검색결과가 구성된다면 사람들은 사회적 문제가 없다고 생각하게 될 것이다. 검색엔진이 접근 경로를 안내해주지 않는다면 그 주장(관심)은 존재하지 않는 것이나 다름없기 때문이다. 이는 상당수의 사람들이 논쟁적 내용보다는 ‘깔끔하게 정돈된 내용’을 보길 원하기 때문에 더욱 강화된다. 그렇게 다듬어진 검색결과는 다수의 사용자가 보기 때문에 더 강화되고, 다시 다수가 채택하는 내용을 보고자 하는 사용자들에게 의해 재선택될 것이다. 그리고 다시 검색엔진은 그런 내용의 중요도를 더 높이 평가할 것이다. 따라서 검색엔진의 합리성과 효율성을 따라 ‘매끈한 것’은 선택되어 지속적으로 재강화되는 반면에 일단 보류된 것은 다시 현실의 표면 위로 부상하기 쉽지 않다.

검색엔진은 사용자의 관심을 흡수하여 자신의 질서로 내재화하고 사용자의 ‘암묵적 동의’를 끌어내는데 성공함으로써 더욱 확고해지고 있다. 검색엔진의 합리적인 판단은 오랜 시간동안의 사용자와의 상호적 작용이 체화되어 있다는 점에서 위력적이다. 그러나 동시에 그러한 검색엔진의 판단은 본질적으로 객관적일 수 없으며 단지 자기 강화의 결과이기 때문에 언제든지 윤리적인 문제를 불러일으킬 수 있다.

3) 검색엔진의 민주주의 문제를 묻다

한글검색엔진은 한글 데이터 부족과 형태소 분석의 난점과 같은 여러 가지 역돌출부를 해결하는 과정에서 통합검색, 질의-응답 시스템, 연관 검색, 실시간 검색과 같은 새로운 기술양식을 발명하며 적응과 자기증식을 실현했다. 한글검색엔진은 사용자의 관심을 더 잘 이해하기 위해서 다양한 기법을 개발했는데, 그 중 하나가 인간편집자를 배치하는 것이었다. 인간편집자는 기계알고리즘 보다 더 적절하게 사용자의 명령을 추론할 수 있다. 특히 한글의 형태소 분석의 난점이나 한글 데이터의 부족과 같은 특수성은 인간 편집자의 역할을 더 높였을 것으로 보인다. 이 때문에 한글 검색엔진의 경우, 상대적으로 검색 결과를 조작한다는 비판을 받고 있다. 그러나 앞선 분석에서 보듯이 검색 알고리즘이 인간편집자 보다 더 ‘순수’하다고 보는 것은 적절치 않다.

2013년 3월 “이순신” 검색어에 대한 논란이 있었다. 드라마 “최고다 이순신”의 영향으로 “이순신”이라는 키워드 검색질의가 상승했다. 검색엔진은 자신의 알고리즘에 의거하여 “이순신” 키워드가 해당 드라마 관련 키워드와 연관되어 등장하기 때문에, 이 키워드를 입력한 사용자가 드라마에 관한 정보를 원한다고 판단했을 것이다. 그래서 “이순신”이라는 질의어에 대해서 “이순신 장군”보다는 해당 드라마에 대한 정보를 상위 검색결과로 노출했는데, 이것은 검색엔진의 입장에서 매우 정당한 것이었지만 사람들은 당혹해 했다. 한 시민 단체는 해당 드라마 제작사에 대하여 제목사용금지 가처분 소송을 제기했고, 일부 사용자들은 검색결과를 문제 삼았다. 이순신장군은 한국인이 가장 존경하는 인물이지만, 검색엔진은 미처 그런 기준을 가지고 있지 않았다. 결국 검색회사들은 여론에 밀려서 검색순위를 즉각 조정했다. 이처럼 검색엔진의 영향력이 커질수록 ‘순수한’ 검색엔진의 해석이더라도 사회적 문제라고 인식될 수 있는 것이다.

검색엔진의 판단은 단순한 안내자의 지위를 넘어서 사회적 민주주의의 문제로 확장되고 있다. 게다가 검색엔진은 사용자의 관심을 증폭하고 강화하여

사용자의 관심을 전이시킬 수도 있다. 그렇다면 일부에서 주장하듯이 검색 알고리즘을 공개하는 것이 해답이 될 것인가? 아마도 내재적 작동방식이 외부로 드러나게 된다면, 검색엔진은 스스로 작동하기를 멈추거나 자신만의 매력을 잃어버릴 수도 있다. 그렇다면 어떻게 검색엔진의 전문성을 승인하면서도, 검색엔진이 민주주의를 지지하는데 기여하게 만들 수 있을 것인가.

먼저, 사용자의 명령과 기계의 해석 사이의 ‘차이’를 인정할 필요가 있다. 도구주의적 관점은 대체로 그 차이(간격)를 무시함으로써 출현한다. 마치 기계가 사용자의 명령을 ‘투명하게’ 수행할 것으로 간주하기 때문에 기계의 행위성은 고려의 대상이 아니며, 단지 인간행위자의 의지가 핵심적이다. 그러나 이미 살펴보았듯이 검색엔진은 인간의 내적 질서를 직접 알 수 있는 방법이 없다. 다만 인간의 외재적 작동을 자신의 고유한 내적 질서에 의해 수용할 뿐이다. 우리는 검색엔진의 행위성과 사용자의 행위성을 동시에 승인할 필요가 있다. 검색엔진은 단순한 도구가 아니며 독자적인 해석의 주체이다.

이와 반대로 기술혐오의 관점은 기계와 사용자 사이를 이분하여 보기 때문에 발생한다. 이 주장은 검색엔진이 드러내고 있는 문제에 대해서 기술 자체의 문제라고만 간주하고 기계 또는 그 기계의 개발자에게 책임을 묻는 것으로 문제를 해결하려고 한다. 그러나 검색엔진은 가장 효율적으로 사용자의 관심에 부응하려고 했고 그 점에서 사용자를 만족시켰기 때문에 수용되었다. 기계의 알고리즘은 사용자와의 순환적 반응과정에서 강화된 것이다. 비록 기계 알고리즘 자신의 내적 방식으로 작동하더라도, 검색엔진이 사용자 맥락의 일부를 체화하고 있다는 점을 부정할 수 없다. 그러므로 검색엔진에 대한 일방적 혐오로부터도 벗어날 필요가 있다. 검색엔진은 고유한 행위성을 가지지만 그렇다고 해서 절대적인 위력을 가진 ‘괴물’은 아니다. 행위자로서 검색엔진은 다른 행위자와 분리되어 홀로 변이할 수 없고, 다른 행위자와의 관계 안으로 제한될 수밖에 없기 때문이다.

우리는 검색엔진의 기계적 알고리즘만을 선호할 수도 없고 그렇다고 해서 인간편집자의 개입을 막연히 지지할 수도 없다. 게다가 검색엔진은 특정 사회

적 정치적 지향을 가진 사용자들과 상호작용하며 서로를 변형시킬 수도 있다. 그렇다면 의제의 초점을 이동할 필요가 있다. 단순히 ‘검색엔진의 중립성’을 요구하는 것으로부터 벗어날 필요가 있다. 이는 검색엔진이 중립적일 수 있다는 도구주의적 관점에 기반하고 있다. 검색엔진이 매끈하거나 오락적인 주제로만 나아가는 경향을 제고하고, 사회정치적 논쟁을 더 빈번히 제시하도록 하려면 오히려 검색엔진은 본래적으로 중립적이지 않다는 점을 공유할 필요가 있다. 검색엔진에게 중립성을 요구하는 것은 검색엔진이 감당할 수 없는 일이다. 사실 검색엔진이 구성하는 세계가 매력적인 이유는 ‘객관적인 정보’를 제공했기 때문이라기보다, 검색엔진과 사용자들이 기호적인 행위교환에 함께 참여할 수 있었다는 점일 것이다. 검색엔진의 문제는 중립성의 부재라기보다는 오히려 검색엔진이 자신의 합리성과 효율성을 과신하여 ‘지식의 다수성(plurality of knowledge)’에 조응하려 노력하지 않을 때이다. 검색엔진의 중립성을 강제하려 하기보다 지식의 다수성을 포섭하도록 권장하는 방향이 검색엔진 자신의 성장에도 부합할 것이므로 협력을 얻을 가능성이 있다.

□ 참고 문헌 □

- 김인성·김내리 (2012), 『두 얼굴의 네이버: 네이버는 어떻게 우리를 지배해 왔는가』, mgt.
- 김지연·김성희·정혜경 (2011), 「검색엔진과 동시대성; 기술시스템이론의 확장」, 한국과학사학회·한국과학철학회·한국과학기술학회 2011년 추계 연합학술대회 발표.
- 김지연 (2012), 「정보통신과 전문성의 재구성: 미네르바 사례를 중심으로」, 동향과 전망 제86호, pp. 164-200.
- 김평호 (2007), 「지식과 기술의 상업적 왜곡의 문제: 검색산업-검색 사이트의 정치경제학」, 한국방송학보 21-6: 132-155.
- 박상규·이찬규·윤경현·김성희·이준호 (2007), 「검색엔진에서 일간 질의어 분포의 정상성에 관한 연구」, 『정보관리학회지』, 제24권 제4호, pp. 255-265.
- 박소연 (2011), 「웹 검색 행태의 추이 및 변화 분석」, 『한국문헌정보학회지』, 45(1): 377-393.
- 박소연·이준호 (2007), 「웹 검색 분야에서의 로그 분석 방법론의 활용도」, 『한국문헌정보학회지』, 제41권 제1호. pp. 231-242.
- 박소연·이준호 (2008), 「주요 검색 포털들의 통합검색 서비스 비교평가」, 『한국도서관·정보학회지』, 제39권 제1호 pp. 265-278.
- 박정아 (2012), 「통합 검색 환경에서 이용자 적합성 판단 기준에 관한 탐색적 연구」, 『정보관리학회지』 29(2): 113-133.
- 백육인 (2013), 「빅데이터의 형성과 전유체제 비판」, 『동향과 전망』 제87호, pp. 304-331.
- 설진아 (2011), 「대학생 집단의 인터넷 포털 사이트에 대한 수용자 평가 연구: 네이버와 구글을 중심으로」, 『언론과학연구』 제11권 1호(2011.3). pp. 157-186.

- 시바 바이디야나단 (2012), 『당신이 꼭 알아줘야할 구글의 배신』 .
황희창 옮김, 브레인스토아.
- 엔에이치엔 (2006), 문서 집합을 기반으로 하는 단어간의 연관도를
측정하는 방법 및 상기 방법을 수행하는 시스템, 대한민국
특허청 출원번호 10-2006-0126272.
- 엔에이치엔 (2009), 『네이버 트렌드 연감 2008』 , ISBN 978-89-527-5502-5 03320.
- 엔에이치엔 (2010), 『네이버 트렌드 연감 2009』 , ISBN 978-89-93976-11-303320.
- 오삼균·박희진 (2000), 「국내 인터넷 탐색엔진에 대한 이용자 중심
의 평가에 관한 연구-한글알타비스타와 네이버를 중심으
로」 , 『한국문헌정보학회지』 제34권 제2호. pp. 117-133.
- 윤성희 (2003), 「웹기반 정보검색을 위한 자연어 키워드 색인에 관
한 연구」 , 『한국컴퓨터산업교육학회 논문지』 , Vol. 4,
No. 12, pp. 1103-1111.
- 윤정근·임일 (2009), 「한국의 대표 인터넷 기업 NHN (2008)」 ,
『연세경영연구』 제46권 제1호(통권 제88호). pP. 93-118.
- 이준호·박소연·권혁성 (2003), 「질의 로그 분석을 통한 네이버 이용
자의 검색 행태 연구」 , 『정보관리학회지』 제 20권 제2
호. p. 27-41.
- 한국정보사회진흥원 (2008), 「인터넷 검색서비스 공정성 제고를 위
한 법제도 개선방안」 , NIA I-RER-08011.
- 황성기 (2007), 「정보매개서비스제공자의 법적 책임의 적정 범위에
관한 연구: 서울중앙지방법원 2007.5.18. 2005가합64571, 손
해배상(기)등의 평석을 중심으로」 , 『사이버커뮤니케이션
학보』 , 통권 제24호. pp. 149-186.
- Abiteboul, S., Hull, R. & Vianu, V. (1994), *Foundations of Databases : The
Logical Level*. Addison Wesley Longman.

- Austin, D. (2011), "How Google Finds Your Needle in the Web's Haystack", Feature Column. <http://www.ams.org/samplings/feature-column/fcarc-pagerank>
- Callon, M. (1980), "Struggles and Negotiations to define what is Problematic and what is not: the Sociology of Translation ", in Knorr, W., Krohn, R., & Whitley, R. (eds.), *The Social Process of Scientific Investigation - Sociology of the Science Yearbook*, Dordrecht and Boston, Mass., Reidel.
- Haigh, G. (2006), "Information Idol: How Google is Making Us Stupid", Monthly, No. 9, pp. 25-33.
- Hargittai, E. (2007), "The Social, Political, Economic, and Cultural Dimensions of Search Engines: An Introduction", *Journal of Computer-Mediated Communication*, Vol. 12, pp. 769-777.
- Halavais, A. (2005), "Social Informatics: Beyond Emergence", *Bulletin of the American Society for Information Science and Technology*, June/July 2005. pp. 13-16.
- Halavais, A. (2009), *Search Engine Society*. Polity Press.
- Hayles, N. (2012), *How We Think: Digital Media and Contemporary Technogenesis*, The University of Chicago Press: Chicago and London.
- Hinman, L. (2005), "Esse est Indicato in Google: Ethical and Political Issues in Search Engines", *International Review of Information Ethics*, Vol. 3(06/2005) pp 19-25.
- Introna, L. & Nissenbaum, H. (2000), "Shaping the Web: Why the Politics of Search Engines Matters", *The Information Society*, Vol.

16, pp. 169-185.

Jansen, B. & Spink, A. (2006), "How are We Searching the World Wide Web? A Comparison of Nine Search Engine Transaction Logs", *Information Processing and Management*, Vol. 42, pp. 248-263.

Kammerer, Y. et. al, (2013), "The Role of Internet-specific Epistemic Beliefs in Layperson's Source Evaluations and Decision during Web Search on a Medical Issue", *Computers in Human Behavior*, Vol. 29, pp. 1193-1203.

Kroenke, D. & Auer, D. (2011), *Database Concepts Fifth Edition*. Pearson.

Latour, B. (1986), "The powers of association " , in J. Law (ed.) *Power, Action and Belief: a New Sociology of Knowledge?*, The Sociological Review, Routledge and Kegan Paul, pp. 264-280.

_____ (1999), *Pandora's Hope : Essays on the Reality of Science Studies*, Harvard University Press.

_____ (2005), *Reassembling the Social : An Introduction to Actor-Network-Theory*, Oxford University Press.

Law, J. & Callon, M. (1992), "The Life and Death of an Aircraft: A Network Analysis of Technical Change " , in Bijker, W. & Law, J. (eds.), *Shaping Technology/Building Society: Studies in Sociotechnical Change*, The MIT Press. pp. 143-156.

Page, L. (1998), 'Method for Node Ranking in a Linked Database', United States Patent, NO: US 6,285,999 B1.

Park, S., Lee, J., & Bae, H. (2005), "End User Searching: A Web Log Analysis of Naver, a Korean Web Search Engine", *Library and*

Information Science Research, Vol. 27, pp. 203-221.

- Pasquinelli, M. (2009), "Google's PageRank Algorithm: A Diagram of the Cognitive Capitalism and the Rentier of the Common Intellect", http://matteopasquinelli.com/docs/Pasquinelli_PageRank.pdf
- Paul, C. (2007), "The Database as System and Cultural Form: Anatomies of Cultural Narratives", in *Database Aesthetics: In the Age of Information Overflow*, edited by Vesna, V., University of Minnesota Press: Minneapolis·London, pp. 95-109.
- Russell, S. & Norvig P. (2010), *Artificial Intelligence: A Modern Approach Third Edition*, Pearson.

논문 투고일	2013년 5월 10일
논문 수정일	2013년 6월 16일
논문 게재 확정일	2013년 6월 19일

Internet Search Engine: Technological Mode that Draws User's Attention to Make Its Expertise Reinforce

Kim, Ji Yeon

This paper tries to analyze technologies of search engine generally, and reveal the additional modes of Korean search engine at the same time. Recently it said that search engine becomes a self-moving and is getting more strong power than the former one existed. There are many difference interpretative views from technological determination to instrumentalism surrounding this system. Search engine invents the technological mode that draws user's attention to make its own expertise reinforce. It is stemmed from the rationality of its own. Especially Korean search engine exposed unique mutation as self-proliferation of it during past a decade, as for example "related keyword" or "real-time popular keyword" service. Its automatic decision aroused democracy matter, now it is not only web guide. How we do make it to serve in democracy, accepting the independent expertise of it simultaneously? We might find new prospect when focusing on interactional modality between engine and human actor, instead counting both as a separate one.

Key terms

Search Engine, Naver, Daum, Nate, Google, Rationality of Technology, Expertise