

---

# 저널리즘 가치에 기초한 알고리즘을 이용한 뉴스의 시각화

A visualization of news with an algorithm based on journalistic values

박대민(Dae-Min Park)\*, 김기남(Gi-Nam Kim)\*\*, 강남용(Nam-Yong Kang)\*\*\*, 서봉원(Bong-Won Suh)\*\*\*\*, 하효지(Hyo-Ji Ha)\*\*\*\*\*, 온병원(Byung-Won On)\*\*\*\*\*

---

요약 현재 온라인 뉴스 서비스는 선정적인 연성뉴스 중심으로 제공된다. 이에 따라 저널리즘 가치를 구현한 뉴스 서비스의 필요성이 대두되고 있다. 정보원과 공동 인용 여부에 따라 기사를 클러스터링하고 가중치를 부여하여 사실성, 다양성, 심층성, 비판성 등 주요 저널리즘 가치를 구현한 알고리즘은 뉴스정보원연결망분석(news source network analysis)으로 제안된 바 있다. 본 연구는 이를 사용자 친화적으로 시각화한 뉴스그래프를 제안한다. 뉴스그래프는 시간과 정보원에 따라 뉴스를 막대그래프 형식으로 어떤 토픽에 대해 분야별, 소속별로 얼마만큼의 중요도에 따라 논의되는지를 대조적으로 보여준다. 본 연구에서는 뉴스 아카이브인 카인즈 기사를 활용하여 뉴스그래프의 베타버전을 구현하였다.

(<http://147.47.125.161/NSNA/> 에서 베타서비스 중이며, 구글 크롬에 최적화 되어있음)

## Abstract

There was widespread criticism of the online news services due to their bias toward sensational and soft news. Thus, news services based on journalist values are socially requested. News source network analysis(NSNA), an algorithm to cluster and weight news sources, quotes, and articles, are suggested as a method to emphasize on journalist values like facts, variety, and depth in the previous study. This study suggests 'news graphs' as a visualization tool of NSNA. News graphs show news as bar graphs weighted by facts and arguments and arranged by organization and subjects. This study designed a beta version using KINDS, the news archive of Korean Press Foundation.

**핵심어:** 뉴스그래프, 뉴스정보원연결망분석, 저널리즘 가치, 사실성, 다양성, 심층성, 뉴스 UI/UX

---

본 연구는 한국정보화진흥원 빅데이터 활용 스마트서비스 시범사업의 지원을 통하여 연구되었음.

\* 서울대학교 언론정보학과 박사수료; e-mail: [heathe1@snu.ac.kr](mailto:heathe1@snu.ac.kr)

\*\* 아주대학교 미디어학과 석사과정; e-mail: [gnkim@ajou.ac.kr](mailto:gnkim@ajou.ac.kr)

\*\*\* 아주대학교 정보컴퓨터공학과 학사과정; e-mail: [fnantia@ajou.ac.kr](mailto:fnantia@ajou.ac.kr)

\*\*\*\* 서울대학교 융합과학기술대학원 융합과학부 부교수; [bongwon@snu.ac.kr](mailto:bongwon@snu.ac.kr)

\*\*\*\*\* 아주대학교 미디어학과 석사과정; e-mail: [hjha0508@ajou.ac.kr](mailto:hjha0508@ajou.ac.kr)

\*\*\*\*\* 교신저자: 서울대학교 차세대융합기술연구원 연구교수; e-mail: [bwon@snu.ac.kr](mailto:bwon@snu.ac.kr)

## 1. 서론

현재 포털 뉴스 검색 서비스에서는 연예, 스포츠, 성(sex), 범죄, 인물, 미담 등 선정적 기사 또는 연성뉴스(soft news)가 중시되는 황색 저널리즘(yellow journalism)의 성격이 강하게 나타난다.[1] 포털 저널리즘이 인터넷 담론 공중[2]의 토론 기초가 되는 등 긍정적 기능이 있음에도 불구하고, 적지 않은 비판을 받는 것도 이 때문이다.[3] 이는 광고 수익 극대화를 위해 클릭 수를 늘리려는 언론사의 온라인 뉴스 편집 관행 때문이기도 하지만, 포털 역시 실시간 검색어 등 인기도에 기반해 뉴스 가중치를 부여하는 등 저널리즘 가치와는 무관한 알고리즘을 적용하는 탓도 있다.

저널리즘의 주요 가치로는 사실성, 다양성, 심층성 등이 널리 인정받는다. 우선 사실성 관행으로는 대표적으로 인용, 수치, 사례를 꼽을 수 있다.[4] 이 가운데 인용은 가장 중요하다. 특히 인용 대상인 정보원은 기사 내용의 핵심 제공자로 저널리즘 연구자들이 일찍부터 주목하였다.[5][6][7]

정보원과 공동 인용 여부에 따라 기사를 클러스터링하고 가중치를 부여하여 저널리즘 가치를 구현한 알고리즘은 뉴스정보원연결망분석(news source network analysis, NSNA)으로 제안된 바 있다.[8] 뉴스정보원연결망은 같은 기사에 두 정보원이 직접인용문으로 함께 인용됐을 경우 이 정보원들 간에 서로 의미론적인 관계가 있는 것으로 보고 간접적으로 만드는 양방향(undirected) 준 연결망(quasi network)[9]이다.

본 연구는 NSNA를 사용자 친화적으로 시각화한 뉴스그래프를 제안한다. 뉴스그래프는 뉴스 보기 화면을 정보원들의 토론장처럼 시각화한다. 즉 어떤 토픽에 대한 분야별, 소속별 의견을 사실에 가중치를 두어 시간에 따라 대조할 수 있도록 막대그래프를 활용함으로써 사실성, 다양성, 심층성 등 저널리즘 가치를 시각적으로 구현하고자 했다.

## 2. 관련 사례

인포테인먼트(infortainment)는 정보(Information)와 오락(Entertainment)의 합성어로서 정보 전달의 효율성과 같은 기능적 측면과 더불어 정보 전달과정에서 사용자들에게 부가적 장치를 통해 심리적 부담을 줄이고 흥미유발을 통해 정보에 몰입할 수 있도록 한다.[10] 최근에 나타나는 뉴스 어플리케이션은 기존의 뉴스 서비스에서 단순하게 뉴스기사만을 전달하는 기능에서 벗어나 데이터 분석을 통해 새로운

정보를 제공하거나 사용자의 관심을 끌 수 있는 인포테인먼트적인 성격으로 바뀌고 있다. 가장 대표적인 사례로는 Marcos Weskamp의 Newsmap이 있다.[11] Marcos의 Newsmap은 Google의 뉴스 서비스를 기반데이터로 Squarified Treemap Algorithm[12]을 이용하여 그림 1과 같이 실시간으로 뉴스의 클러스터링과 중요도를 각각 색상과 사각형의 크기를 통해 시각적으로 제공한다.



그림 1. Newsmap 웹 사이트

인포테인먼트적인 요소를 강조하여 사용자에게 더 친숙하게 접근한 사례로는 NBC news에서 서비스하는 Spectra[13]가 있다. Spectra는 사용자가 원하는 채널(분야)을 선택함에 따라 해당 채널의 뉴스 기사들이 화면 상단에서 일정한 궤도를 형성하며 회전한다. Spectra는 분야별 필터링 외에 다른 추가적인 정보를 제공하지는 않지만 인포테인먼트적인 요소로 사용자의 흥미를 유도하고 있다.



그림 2 인포테인먼트 요소를 적용한 뉴스 어플리케이션 Spectra

## 3. 뉴스그래프의 구현

뉴스그래프는 정보원연결망 알고리즘으로 각 토픽에 대해 분야별, 소속별로 중요한 정보원을 도출하고, 정보원의 중요도에 따라 가장 중요한 기사를 막대그래프 형식의 목록으로 보여준다. 뉴스 데이터로는 한국언론진흥재단의 뉴스 아카이브인 카인즈의 1990년 이후 66개 매체 2900여 만 건의 기사를 활용하였다. 뉴스그래프는 다양한 미디어기기를 통해 사용자들의 접근이 용이하도록 웹 기반으로 제작되었다.



그림 3 뉴스그래프 메인화면

그림 3는 뉴스그래프 어플리케이션 화면이다. 첫 화면에서는 정치, 경제, 사회, 문화, 국제의 분야별로 가장 중요한 뉴스기사의 제목을 순서대로 보여준다. 기사의 분류는 카인즈에서 분류된 것을 참고하여 각 분야에 대한 트레이닝 테이블을 작성한 후 베이지안 텍스트 분류(Baysian Text classification)기법을 통해 이루어졌다. 사용자가 기사 제목을 선택하면 그림 4처럼 해당 기사의 원문을 볼 수 있다.



그림 4 뉴스제목 선택화면

뉴스그래프에서는 기사원문뿐 만 아니라 사용자가 선택한 기사주제를 중심으로 다른 뉴스 기사에 등장하는 정보원들이 언급한 인용문을 함께 제공한다. 그림 5는 추가 정보원들의 인용문이 제공된 화면을 확대한 것이다. 각 정보원들의 이름 옆에 해당 정보원이 언급한 총 인용문의 수가 표기되며, 정보원의 이름을 클릭하여 열람할 수 있다. 이 기능으로 사용자는 같은 사건을 다루는 다른 매체 기사를 따로 검색할 필요 없이 다양한 사람들의 의견을 한 화면에서 살펴볼 수 있으며 이는 기존의 인터넷 뉴스 서비스 및 뉴스 어플리

케이션과의 차별점이라 할 수 있다.



그림 5 추가정보원이 제공된 화면

뉴스그래프의 또 다른 기능으로 주제어 검색 기능이 있다. 이 기능은 설정된 기간 동안 주제어가 포함된 모든 뉴스 기사를 대상으로 분야 및 소속별 정보원들을 중요한 순서대로 제공한다. 이 기능을 통해 사용자는 관심있는 주제에 관하여 분류 및 소속별로 다양한 정보원의 인용문을 확인 할 수 있다. 각각의 정보원들은 5개의 대분류(정치, 경제, 종합/사회, 문화, 국제)와 각 대분류에서 표 1과 같이 46개의 중-소분류로 분류된다. 분류는 언론의 지면 분류 참고와 함께 출력처 관행을 참조하여 연구진이 부여하였다.

표 1 정보원 분류표

대분류	중-소분류	계
정치	법원/검찰/변호사/경찰, 소방/정당/국회/청와대/행정부/지자체/정부출연/정치인/북한/군대/국제기구	14
경제	농림수산업/광업/에너지/IT전자/차, 선박, 항공/건설업/중소기업/석유화학/제철/대기업/배선, 식품, 마트/금융/서비스업/자영업/관광, 외식/운송/협회/노조/직장인	19
종합/사회	의료계/학계/종교계/예술계/대중문화계/스포츠계	6
문화	시민봉사단체/대중교통/복지시설/시민/기타	5
국제	북미유럽/동아시아/중동, 아프리카/남미	4

정보원들을 단순히 중요도 순서에 따라 나열 할 경우 특정 소속의 입장만 집중되는 문제가 발생할 수 있다. 이를 방지하기 위하여 전체 정보원은 소속별로 추가 가중치를 부여 받는다. 예를 들어 검색된 정보원들이 5명의 여당소속과 2명의 야당의 소속으로 분류되고, 중요도 계산 결과 여당의 인원들이 모두 중요도 1~5위를 기록하고 야당의 인원이 6~7위를 기록한다면 사용자의 입장에서는 목록의 하단까지 내려야 야당의 의견을 확인할 수 있다. 이러한 문제점을 해소하기 위하여 뉴스그래프에서는 정보원 목록에서 5명 단위로 서로 다른 소속의 정보원이 최소한 1명이상 포함될 수 있도록 함으로써 저널리즘의 가치 중 다양성을 반영하였다.



그림 6 주제어 '창조경제'의 검색결과

그림 6는 기간을 2013년 9월 1일부터 2013년 9월 5일로 설정하고 '창조경제'라는 주제어를 검색한 결과이다. '창조경제'라는 주제어에 관하여 정치, 경제, 종합/사회, 문화 분야의 인물들이 각각 검색이 되었으며 국제 분야는 검색된 인물이 없으므로 페이지 공간 활용을 위해 영역이 줄어들어 있다. 각각의 정보원들은 대분류에 따라서 동일한 계통의 색을 가지지만 중분류에 따라서는 서로 다른 색상을 가진다. 정부소속의 '윤상직'과 변호사인 '배정환'은 같은 계통의 색을 가지나 소속이 다르기 때문에 다른 색상을 나타내는 것을 확인할 수 있다.

저널리즘에서의 중요한 가치 중 하나는 정보를 얼마나 심층적으로 제공할 수 있는가이다. 뉴스그래프에서는 각 주제어에 따른 정보원을 보여줄 뿐만 아니라 자아연결망에 따라 직접적으로 관련 있는 정보원을 필터링 해줌으로써 심층성을 반영하였다. 그림 7은 그림 6과 같은 검색조건으로 검색하였을 때, 정보원 연결망 기능을 활성화하여 '최문기'를 선택한 결과다.



그림 7 주제어 '창조경제'의 '최문기'의 자아연결망

선택된 '최문기'의 자아연결망에 포함되지 않는 정보원의 경우 사라지게 되며 화면에는 자아연결망에 포함되는 정보원만 남아있게 된다. '최문기'의 자아연결망에 포함되는 정보원으로 정치 분야의 '박근혜'와 종합/사회 분야의 '배영훈' 등이, 문화 분야에서는 '이경재' 등이 남아있는 것을 확인할 수 있다. 이들은 "창조경제"라는 주제와 관련하여 '최문기'와 함께 기사에 인용된 정보원들이다.

## 5. 결론

본 논문은 뉴스정보원연결망분석(news source network analysis, NSNA)을 기반으로 일반 사용자들이 쉽게 이해할 수 있는 뉴스그래프를 제안하고 실제로 구현한 어플리케이션을 소개하였다. 뉴스그래프는 저널리즘의 가치에 있어서 사실성, 다양성, 심층성을 반영하고 있다. 사용자는 뉴스 그래프를 통하여 중복된 정보를 필터링하고 최대한 다양한 분야의 의견을 한 화면에서 볼 수 있다. 또한 단순한 인터랙션을 통하여 뉴스기사에 등장하는 정보원의 자아 연결망을 볼 수 있다는 점도 다른 뉴스 어플리케이션과의 차별점이라 할 수 있겠다.

- [1] 한국언론연구원. 매스컴대사전. 한국언론연구원, 1993.
- [2] 이준웅, "인터넷 공론장의 매개된 상호가시성과 담론 공중의 형성". 언론정보연구. 46권 2호. 5-32, 2009.
- [3] 김위근, 김성해, 김동윤, "뉴스의 대중화 혹은 저널리즘의 계도화: 저널리즘 관점에서 본 네이버 '뉴스캐스트' 사례 분석". 사이버커뮤니케이션학보. 30권 2호. 33-72, 2013.
- [4] van Dijk, T. A., "News as discourse". NJ: Lawrence Erlbaum, 1988.
- [5] Sigal, L. V., "Reporters and officials: the organization and politics of newsmaking". Lexington, Mass: DC Heath, 1973.
- [6] Schudson, M., "Discovering the news: a social history of American newspapers". NY: Basic Books, 1978.
- [7] Gans, H., "Deciding what's news". Austin: University of Texas Press, 1979.
- [8] 박대민. "뉴스 기사의 빅데이터 분석 방법으로서 뉴스정보원연결망분석". 한국언론학보. 57권 6호. 233-261, 2013.
- [9] 박대민, "정보원 활용을 통한 신문사의 타당성 요구 응답전략: 동아일보와 한겨레신문의 4대강 관련 정보원 연결망 분석". 제 15회 전국 언론학 대학원생 컨퍼런스. 한국언론학회, 2011.
- [10] 주영희, "인포테인먼트를 통한 웹진의 시각 커뮤니케이션 디자인에 관한 연구 : 국내 웹진의 콘텐츠 내에서의 재미요소를 중심으로", 학위논문(석사), 상명대학교 대학원, 2011
- [11] Weskamp, Marcos. "Newsmap." Webdesigning Magazine, June, 2004.
- [12] Bederson, B.B., Shneiderman, B., and Wattenberg, M., "Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies", ACM Transactions on Graphics (TOG), 21, (4), 833-854, 2002.
- [13] <http://media1.s-nbcnews.com/i/msnbc/components/spec tra/>
- [14] <http://14.63.215.206/NSNA/>