

‘팩트’의 형식적 구성 요건에 대한 탐색적 연구
: SNU FactCheck 기사 메타데이터 분석을 통한 제언

오세욱*
(한국언론진흥재단 선임연구위원)
황구현**
(SNU팩트체크센터 연구원)

< 초 록 >

뉴스 형식의 허위정보가 확산되면서 ‘팩트체크’가 유력한 대응 방안 중 하나로 주목받고 있다. 그러나, 방대한 양의 정보를 사람이 일일이 확인하는 것은 어렵고 시간이 오래 걸린다. 이에 따라, 자동 사실 확인 기술(automated factchecking technology)을 통해 이러한 문제를 해결하려는 시도들이 이어지고 있다. 하지만, 자동 사실 확인 기술들은 아직까지 사실 여부를 최종적으로 확인하기보다는 사실 여부 확인이 필요한 정보를 빠르게 찾아주는 수준이다. 이러한 문제 해결을 위해 팩트를 구성하는 기술적인 형식 요건을 정립하고 이의 충족 여부에 따라 자동으로 사실을 확인하려는 시도들이 이어지고 있다. 이 연구는 구글 뉴스의 ‘팩트 체크’에 사용되는 ‘클레임리뷰 마크업(ClaimReview Markup)’, BBC의 ‘스토리라인 온톨로지 (Storyline Ontology)’, ‘투명성 선언(Transparency Initiative)’의 ‘hnews’ 마이크로 포맷 등 이러한 형식적 요건과 관련한 세 가지 사례를 검토했다. 이후 한국 ‘팩트체크’ 정보의 데이터 베이스라고 할 수 있는 ‘SNU 팩트체크’ 등록 정보의 메타데이터 분석을 실시했다. 총 740개의 정보를 분석한 결과 대부분의 기사에서 팩트를 뒷받침하는 근거자료, 원문 출처 URL 등 표기를 제대로 하고 있지 않은 것으로 나타났다. 이러한 수치는 SNU 팩트체크 정보들이 절차적 투명성에 대해 상당히 무관심하다는 점을 의미하며, 이 연구는 이러한 문제 해결을 위해 한국 상황에 맞는 팩트의 형식적 요건과 이의 실현을 위한 방안을 제시했다.

키워드 : 팩트, 팩트체크, 객관성, 투명성, 메타데이터, 플랫폼

* kinpah@gmail.com
** guhyunhwang@gmail.com

1. 서론

국립국어원이 펴낸 표준 국어대사전에 따르면 '사실'은 "실제로 있었던 일이나 현재에 있는 일"을 말한다. 그런데 이 사실을 확인하기는 쉽지 않다. 실제로 있었던 현장에서 누군가가 사실을 전달하더라도 그것을 정확하게 전달했는지는 그 사람만 알기 때문이다. 사실은 별도로 존재하는 것이 아니라 구성되기도 한다. 설(Searle, 1995)에 따르면, 사실은 원초적 사실(brute fact)과 사회적 사실(social fact)로 구분할 수 있다. 원초적 사실은 인간의 제도(institution)와 독립적이다. '물 분자는 수소 원자 두 개와 산소 원자 하나로 구성되어 있다' 등 대부분의 자연과학적 사실들을 의미한다. 사회적 사실은 그 존재를 누군가에게 이해시키기 위해 제도를 필요로 하는 것들(Searle, 1995, pp. 27-29)을 말한다. '청와대는 대한민국 대통령의 집무실이자 관저'라는 내용을 이해하기 위해서는 '대한민국', '대통령', '집무실', '관저'라는 제도에 대한 설명이 따라야 하기 때문에 사회적 사실로 분류된다. 즉, 사회적 사실은 사회적으로 구성된 개념에 따라 그 사실 여부를 입증한다.

사실은 주로 언어를 통해 표상된다. 전통적으로 기자들은 사실을 취재하여 사실적 진술을 통해 그 내용이 진실이라고 주장해 왔다. 이러한 내용이 담긴 기사는 일반적으로 사실을 담고 있을 것이라고 믿어져 왔다. 하지만, 완전한 사실을 완전하게 전달하는 것은 어렵다. 완전하게 알기 어려우며 그 완전성을 입증하는 것 또한 어렵기 때문이다. 워터게이트 특종으로 풀리처 상을 수상한 칼 번스타인(Carl Bernstein)은 "기사는 최선을 다해 얻을 수 있는 진실의 한 조각(getting the best obtainable version of the truth)"¹⁾일 뿐이라고 말한다. 일반적으로 사실을 확인하고 전달하는 언론과 언론인들은 확인 가능한 사실이 있을 것으로 가정하고 최선을 다해 모든 조각들로 '종합적 진실(the whole truth)'을 전달할 수밖에 없다(이민웅, 2002). 절대적 사실 혹은 진실이 존재하기보다는 사실은 여러 맥락의 조각이 모여 종합적으로 구성되기 때문이다.

사실의 전달자로서 언론은 전달하는 사실이 사실임을 증명하고, 인정받기 위해 다양한 규범을 준수할 수밖에 없었다. 그 규범 중 첫 번째는 객관성이라고 할 수 있다. 물론, 객관성은 그 개념 자체가 문제적이며 역사적으로 긴 발전 과정을 거쳐 왔다(Ward, 2004/2007). 객관성의 이상은 독립적, 공정한 저널리즘을 실천하려는 시도의 확장이며, 당파성과 편향성을 감소시키려는 노력으로 나타나게 되었다. 언론인들은 전문성의 상징이자, 과도한 황색 저널리즘의 처방으로서 객관성 규범을 찬성했으며, 객관성의 기준을 강조하고 추가하는 과정에서, 객관적 뉴스와 주관적 의견 사이의 선을 긋고, 기자들은 사실과 가치를 구분할 수 있다고 주장했다(Ward, 2004/2007, 249~365쪽). 하지만, 사실은 모순적이고 복합적인 것으로, 정보를 추구하고 그 정보와 관련된 그릇된 정보를 확인하고 의도적인 유포나 선전적인 편견을 제거하여 진실에 다가가기 위해 노력하는 과정으로 객관성을 이해해야 할 필요가 있다(Kovach & Rosenstiel, 2007/2009, 68~88쪽).

합리적 취재의 규율로서 기본적인 원칙들이 존재하고 이러한 방법론이 객관성을 추구할 수 있게 한다. 없었던 것을 추가하지 않고, 독자를 속이지 않고, 방법과 동기에 대해서 투명하게 제시하고, 기자 스스로 취재하고 겸허하게 접근한다는 원칙 같은 것 등이다. 이러한 원칙들은 투명성 원칙으로 이어지는데 투명성 원칙은 사실 확인에 있어서 중요한 요소다. 취재원과 취

재 방법에 대한 공개, 취재원들은 누구이며 그들이 가진 지식의 성격과 편견, 충돌하는 내용이나 모르는 내용이 무엇인지를 투명하게 드러내고 정직하게 대하면 객관성의 추구에서 발생 하는 문제는 대체로 해결될 수 있다(Kovach & Rosenstiel, 2007/2009, 146~155쪽).

이러한 투명성 논의는 디지털 시대가 본격화되면서 언론이 전달하는 내용이 형식적으로 특정 구성 요건을 갖출다면, 그 내용의 사실 근접 여부를 판단할 수 있을 것이라는 가정으로도 이어지고 있다. 즉, 사실(fact)의 여부는 판단할 수 없지만, 사실에 얼마나 근접하였는지를 형식적 구성 요건의 충족 여부로 판단하는 것이다. 예를 들어, PEJ(Project for Excellence in Journalism)의 보도지수는 투명 취재원 4개 이상 포함, 복합적 관점 제시, 이해 당사자가 4개 이상 포함될 경우 최고 수준의 기사로 사실에 근접할 것으로 판단한다. 박재영·이완수(2010)도 토대 정보의 신뢰성이 확인된 기사, 이해 당사자가 2개 이상, 투명 취재원이 2개 이상, 인용구의 주관적 술어가 1개 이하, 무주체 주관적 술어가 1개 이하, 복합적 관점 제시 등을 구성 요건으로 제시한다. 그동안 이러한 요인들에 대한 분석은 주로 사람에 의해 이루어졌는데, 기술 발전에 따라 사람의 내용 분석이 아닌 기계적 분석을 통해서 자동으로 이러한 요인들을 추출해 분석할 수 있게 됐다. 이에 따라, 디지털 환경에서 사실(fact)의 '사실(truth)' 근접성 여부를 특정 형식적 구성 요건의 충족 여부로 판단하려는 기술적 시도들이 이루어지고 있다.

이 연구는 우선 특정 형식적 구성 요건 충족 여부를 통해 언론이 전달하는 사실의 사실 근접 여부를 판단하려는 기술적 시도들의 사례를 제시하고 그 특성들을 분석한다. 구글 뉴스의 '팩트 체크'에 사용되는 '클레임리뷰 마크업(ClaimReview Markup)', BBC의 '스토리라인 온톨로지(Storyline Ontology)', '투명성 선언(Transparency Initiative)'의 'hnews' 마이크로 포맷 등 세 가지 사례들에서 사용하고 있는 형식적 구성요건들이 무엇인지, 그 요건들을 어떻게 확인하고 적용하고 있는지 등을 분석한다. 이후 현재 'SNU 팩트체크(factcheck)'를 통해 제공되고 있는 팩트체크 정보들의 메타데이터를 수집해 분석한다. 'SNU 팩트체크'는 참여 언론사의 팩트체크 정보를 한데 모아 표출하는 국내에서 가장 큰 팩트체크 데이터베이스라고 할 수 있다. 이 팩트체크 정보들의 메타데이터들은 각각 특정 사실 구성을 위한 형식적 요건이라고 할 수 있다. 2017년 3월 29일부터 2018년 6월 30일까지 SNU 팩트체크에 등록된 팩트체크 정보의 메타데이터를 분석해 현재 한국의 언론들이 어떠한 형식적 요건을 지키고 있는지를 보고자 한다. 이를 통해, 이른바 '팩트'를 구성하는 요건이 무엇인지에 대해 탐색적으로 연구하고 요건 준수의 필요성 및 문제점, 향후 활용 가능성 등에 대해 논의하고자 한다.

이른바 '가짜뉴스(fake news)' 등 허위정보의 확산으로 인해 유통되고 있는 정보들에 대한 사실 여부 확인의 중요성이 강조되고 있다. 하지만, 유통되는 정보의 방대함과 신속함 때문에 효율적 대응이 어려운 상황이다. 이러한 상황의 극복을 위해 기술에 의한 자동화된 사실 확인(automated factchecking)이 늘어나고 있다(오세욱, 2017). 그러나 어떻게 사실을 확인할 것인지에 대한 논의가 이루어지기도 전에 특정 기술적 방식의 적용으로 이를 해결하려는 시도들이 이어지고 있다. 사실은 전술했듯이 복합적이다. 이 연구는 '팩트'의 형식적 구성 요건들에 대해 탐색함으로써 향후 기술에 의한 사실 확인에 있어서 적용할 수 있는 가능성, 혹은 효율적 적용 방안의 제시를 목적으로 한다. 이와 함께, SNU 팩트체크의 분석을 통해 한국 언론이 갖고 있는 특성을 감안한 구체적인 적용 방식의 제안을 담고자 한다.

1) 2017년 5월 8일 예루살렘 프레스 클럽 개최 The Freedom of the Press 컨퍼런스 연설 내용("Best Obtainable Version of the Truth". Retrieved from <https://www.youtube.com/watch?v=16WiuU1Lbdw>) 중 일부

2) 이 글에서는 홑따옴표로 '팩트'라는 용어를 사용한다. 팩트는 기본적으로 사실을 의미하지만, 형식적 구성요건을 충족한다고 해서 사실임을 입증하지는 못하기 때문에 그 충족여부로 판단되는 사실이라는 의미에서 '팩트'라고 구분했다.

2. 이론적 논의 및 선행 연구 검토

1) 객관성에서 투명성으로

그동안 민주주의 공동체 사회의 유지와 발전을 위한 공적 언론으로서 갖추어야 할 핵심적 자질로 꼽혀온 것은 객관성이다. 물론 전술했듯이 객관성은 그 개념 자체가 문제적이며 역사적으로 긴 발전 과정을 거쳐 왔다. 워드(Ward, 2004/2007)는 언론 윤리의 발전을 역사적으로 추적하면서, 객관성의 문제가 19세기 자유주의 신문으로부터 나타남을 밝혔다. 이 시기 영국 신문의 경우 엘리트주의적 입장에서 대중을 위한 공정한 뉴스 제공자의 입장에서 뉴스를 강조했다. 하지만 미국의 평등주의적 대중지 관점에서는 모두에게 접근 가능한 형태로 정보를 제공하는 것이 가장 중요했다. 특히 19세기 들어 언론이 상업화되고, 기술이 발전하면서 정보 수집력이 강화됨에 따라, 어떤 정보가 선택되고 신문에 실리는가와 관련해 편파성에 대한 질문이 발생했다. 즉, 자유주의적이고 상업적인 신문의 맥락에서 대중에게 쉽게 전달되거나 대중이 흥미를 느낄 것으로 생각되는 정보를 선택하는 저널리스트의 책임 문제가 부각되기 시작한 것이다.

이론적으로 객관성의 이상은 독립적, 공정한 저널리즘을 실천하려는 시도의 확장이며, 당파성과 편향성을 감소시키려는 노력으로 나타나게 됐다. 이 시기 객관성은 세 가지 지표를 통해 구성됐다. 첫째는 사실성이다. 객관성에 대한 최초의 담론은 이 사실성 개념을 중심으로 구성됐다. 사진의 발전은 현실을 있는 그대로 드러낸다는 사실성의 신화의 정점이었다. 하지만 수동적인 관찰자의 사실적 진술이 보도이고, 이것이 객관성이라고 간주되기 시작하면서, 이는 이후 객관성에 대한 철학적 혼란을 일으키는 원인이 됐다. 두 번째 지표는 독립성이었다. 정부 권력으로부터의 독립적인 실천은 저널리즘에 있어 가장 중요한 가치로 간주됐다. 사실적 보도와 독립적 논평은 자율성에 근거한 객관성의 의미를 구성했다. 세 번째 지표는 전문직주의로, 전문직은 공공의 이익에 복무하며, 객관적 태도를 견지해야 한다는 주장을 통해 객관성의 지표로 작동했다.

워드 는 이 규범이 곧 객관성의 한계와 그와 관련된 사회적 논란에 부딪히게 됐다고 지적한다. 20세기 중반 이후 객관성이 더 이상 윤리 규범의 중심이 되지 못했다고 보는 것이다. 전통적 객관성의 개념은 산업의 입장에서는 사업상 불편부당한 뉴스가 좀 더 금전적으로 유리하고 명예훼손의 우려가 적었기 때문에 환영받았던 점이 있다. 언론인들은 전문성의 상징이자, 과도한 황색 저널리즘의 처방으로서 객관성 규범을 찬성했다. 객관성의 기준을 강조하고 추가하는 과정에서, 객관적 뉴스와 주관적 의견 사이의 선을 긋고, 기자들은 사실과 가치를 구분할 수 있다고 주장했다(Ward, 2004/2007, 249~365쪽).

객관성 개념에 논란의 소지가 있다고 해서 이 가치가 의미 없다는 것은 아니다. 단, 철학적으로 볼 때, 이 가치가 중요하며 추구되어야 한다는 것이 변함없는 사실임에도 불구하고 인간적 편향을 배제한다는 의미로서의 객관성은 가능하지 않다는 문제가 있다. 객관성을 추구하기 위해 형식적 요건을 강조하게 됨에 따라 오히려 뉴스의 표현 양태나 기사 취재 방식 등이 제한되는 결과가 발생하게 되는 우려도 있다. 객관보도와 관련된 논란을 다루는 연구들이 보여주는 것은 언론보도가 다양한 이해관계로 인해 편향될 수 있다는 점이며, 객관성은 단지 자본주의적, 혹은 상업적인 이해관계에 기반을 둔 이데올로기에 불과하다는 주장도 제기된다. 메릴과 로웬슈타인(Merrill & Lowenstein, 1971/1975)은 객관보도란 애초에 존재하지 않는 것으로서 불가능한 것이라는 입장과, 현실적으로 어렵지만 가능하며 동시에 부단히 추구해야 할

목표로 바라보는 입장으로 객관성에 대한 입장이 나뉜다고 정리한 바 있다.

언론인들은 객관적 보도란 달성할 수 없는 이상적 목표라는 인식을 하고 있었지만, 자신의 주관적 편향을 통제하기 위해 전략적으로 객관성을 일종의 직업적 규범으로 확립했다(Shudson, 2001). 언론인들은 확인이 불가능한 사실을 보도할 필요가 있을 때, 이른바 '사실성의 망(the web of facticity)'을 이용해 그것을 사실처럼 보이게 만들었다(Tuchman, 1978). "누가 어떻게 말했다"는 것을 사실로 삼아 보도, 사실에 근접한 것으로 인정되는 인물이나 기관을 중심으로 보도, 직업 인용문을 이용해 발언 사실을 강조한 것이다. 객관성 규범의 핵심인 사실과 의견의 분리는 현실적이지는 않지만 일종의 규제적 이념으로 작동했다(이준웅, 2010).

저널리즘은 진실을 전달해야 한다. 그런데 진실이 사실을 정확하게 전달하는 것이라고 한다 면, 그 개념은 단순하지 않다. 전술했듯이 절대적 수준에서 진실이라는 것이 존재하지 않을 수도 있다. 더욱 중요한 것은 저널리즘의 진실은 단순한 정확성을 넘어선다는 것이다. 이제까지의 저널리즘은 기능적 진실의 개념을 발전시켰다. 그러나 단순한 사실 확인과 정확성이란 진실에 대한 독자들의 기대에 미치지 못한다. 정확성은 해석과 논쟁, 맥락과 공적 커뮤니케이션의 토대이지만 이것만으로는 불충분하다. 진실은 모순적이고 복잡한 것으로, 정보를 추구하고 그 정보와 관련된 그릇된 정보를 확인하고 의도적인 유포나 선전적인 편견을 제거하여 진실에 다가가기 위해 노력하는 과정으로 이해해야 한다(Kovach & Rosenstiel, 2007/2009, 68~88쪽).

객관성은 저널리즘의 방법론으로, 기자는 객관적일 수 없지만 저널리즘은 객관적일 수 있다. 물론 현재 사실 확인의 저널리즘 방법론은 잘 확립되었다기보다는 관행이나 전승처럼 되어 있으며 체계적이지 않다. 그러나 합리적 취재의 규율로서 기본적인 원칙들이 존재하고 이러한 방법론이 객관성을 추구할 수 있게 해준다. 없었던 것을 추가하지 않고, 수용자를 속이지 않고, 방법과 동기에 대해서 투명하게 제시하고, 기자 스스로 취재하고 겸허하게 접근한다는 원칙 같은 것이 그 사례이다. 특히, 투명성의 원칙은 사실 확인의 규율에서 중요한 요소이다. 취재원과 취재 방법에 대한 공개, 취재원들은 누구이며 그들이 가진 지식의 성격과 편견, 충돌하는 내용이나 모르는 내용이 무엇인지를 투명하게 드러내고 정직하게 대하면 객관성의 추구에서 발생하는 문제는 대체로 해결될 수 있다고 코바치와 로젠스텔은 주장한다(Kovach & Rosenstiel, 2007/2009, 146~155쪽).

디지털 기술의 신속성, 상호작용성, 투명성(transparency), 개방성(openness)은 기존 저널리즘의 원칙과는 다른 새로운 윤리 원칙을 요구하고 있다(Singer, 2010). 디지털 미디어의 신속성과 상호작용성 때문에 진실성이 영향을 받고 있으며, 신속성에 대한 요구 때문에 정보 수집이 빨라지는 한편으로 정보가 부정확해진다는 양가성이 생겼으며, 상호작용성 때문에 뉴스에 대한 수정 가능성이 높아지고 정정 요구가 발생할 수 있어 사실 확인의 부담은 더 커지고 있다. 맥브라이드와 로젠스텔(Mcbride & Rosenstiel, 2014/2015)은 디지털 시대 저널리즘의 윤리 원칙으로 진실, 투명성, 공동체라는 세 가지 기준을 제시한다. 먼저 진실의 원칙은 "진실을 추구하고 가능한 한 완벽하게 보도하라"는 명제로, 정확성의 추구, 발언권 없는 집단의 대변, 권력층 비판 등을 강조하는 것이다. 투명성의 원칙은 디지털 저널리즘 시대에 강조되는 새로운 원칙으로 저널리즘의 생산 방식이 진실성을 보장할 수 있는 현시점에서, 언론인에게 취재 과정에 대해 알리고, 문제가 있으면 신속히 수정해야 함을 의미한다. 공동체에 대한 책임은 저널리즘 차원의 결정이 공동체의 선을 위한 것이며, 다양성을 추구하고 공동선을 추구하는 과정과 관련되어 있다는 점을 강조하는 것이다.

투명성의 원칙이 강조되는 이유는 저널리즘이 민주적 공동체를 구성한다는 믿음 때문이다. 민주적 공동체를 위한 정보의 공개는 중요하며, 투명성은 이러한 정보의 신뢰도를 높이는 방법이다. 하지만 사실 '투명'하다는 것이 누구의 무엇을 '보여주는' 것인지 불분명하다. 이에 대해, 크래프트와 하임(Craft & Heim, 2009, p. 219)은 "저널리즘에서 투명함은 사실과 상황, 사건과 의견이 정렬되고, 이동되고, 의미화되고 제시되는 과정을 열어 놓는다"고 답을 제시한다. 투명성의 가치는 자발적으로, 이제까지 저널리즘에 대해 제기되던 비판에 답하는 방법으로, 정보의 다양성을 추구하고 저널리즘의 성과를 평가하는 방식이 된다는 점에 있다 (Craft & Heim, 2009). 또한, 투명성은 기자가 수용자를 존중한다는 것을 보여준다는 점에서 의미를 지닌다(Kovach & Rosenstiel, 2007/2009, 147쪽). 취재원을 드러내고 취재 과정을 공개하여 취재원의 오류나 의도적인 사기를 막을 수 있으며, 수용자가 이를 공정하게 판단할 수 있는 기회를 제공한다. 기자는 투명성을 추구하면서 공익적인 의무를 다하고 있다는 확인을 할 수 있다. 한편 기자들의 성실성이나 전문성이 의심받는 현대에 있어 수용자가 잘못 이해할 수 있는 결정에 대해서 설명하고, 어떤 것을 넣고 어떤 것을 뺐는지에 대해서 그 동기 와 의사 결정을 질문하는 것은 객관성의 본질적 차원을 다시 생각하게 하는 셈이다. 기자가 이를 어떻게 알았는지, 왜 그렇게 믿는지를 드러냄으로써 수용자들도 이를 똑같이 할 수 있고 그렇기 때문에 기자를 믿을 수 있게 된다. 투명성의 다른 차원으로 중요한 것은 기자가 취재원을 속이거나 오도하지 않게 된다는 점이다. 반대로 취재원 역시 기자를 오도하지 않게 된다 (Kovach & Rosenstiel, 2007/2009, 147~151쪽).

2) 내용 분석에서 기계 분석으로

이렇듯 객관성의 이상을 투명성을 통해 실질적으로 구현할 수 있을 것으로 받아들여지게 되면서 기사 내에 투명성을 구현하기 위한 형식적 요건들이 요구되기 시작했다. 특정 사안에 대해 상반된 입장의 취재원이 등장하는지, 실명 취재원을 인용했는지, 무주체 술어 등을 사용하지 않았는지 등 형식적 요건들을 투명하게 제시함으로써 실질적 객관성을 구현할 수 있다는 것이다. 대표적인 사례가 미국의 '우수 저널리즘을 위한 프로젝트(PEJ, Project for Excellence in Journalism)'의 보도 지수³⁾다. PEJ는 좋은 기사의 요건으로 투명 취재원(transparent source), 복합적 관점의 제시(presence of multiple viewpoints), 기사 내 이해 당사자 수(number of stakeholders)를 제시한다. 투명 취재원은 '기사와의 관련성이 명백히 드러난 취재원'으로서 '그가 누구인지 독자가 알 수 있는 취재원'을 의미한다. 복합적 관점의 제시란, 어떤 형태로든 관점이 제시되어 있었던 기사 중 '하나의 견해가 기사의 3분의 2 이상을 구성하지 못하는 경우'로, '대체로 단일 관점'은 '다른 견해가 일시적으로 언급되지만(a passing reference) 한 경우'로, '완전한 단일 관점'은 '그런 일시적인 언급조차 없는 경우'로 정의된다. 이해 당사자는 이해관계가 서로 다른 집단(different interested groups)을 의미한다. PEJ는 구체적으로 투명 취재원 4개 이상 포함, 복합적 관점이 제시된 기사, 이해 당사자가 4개 이상 포함된 기사일 경우 최고 수준의 기사라고 판단한다. 기사가 특정한 형식적 요건을 준수하는지를 분석함으로써 해당 기사의 품질, 즉 종합적 진실을 전달하고 있다고 평가하는 것이다. 투명성을 보여주는 기사 내 형식적 요건들을 분석해 종합적 진실을 전달하기 위한 객관성의 준수 여부를 평가함으로써 사실에 근접한 내용을 담고 있는지를 판단하는 방식이다.

신문 편집인 746명을 대상으로 한 설문조사 결과를 바탕으로 뉴스 품질 지수(News

3) <http://www.stateofthemediamedia.org/2005/newspapers-intro/content-analysis/>

Quality Indicators) 23개를 선정한 보가트(Bogart, 1989)는 신문의 한 호 전체의 형식적 요건 충족 여부를 분석해 전체적인 신문의 품질을 평가했다. 박재영과 이완수(2010)는 토대 정보의 신뢰성이 확인된 기사, 이해 당사자가 2개 이상, 투명 취재원이 2개 이상, 인용구의 주관적 술어가 1개 이하, 무주체 주관적 술어가 1개 이하, 복합적 관점 제시(갈등 사안인 경우) 등을 충족할 경우 고급 기사로 판단하는 뉴스 평가 지수를 만든 바 있다. 이러한 사례들 모두 투명성과 관련한 형식적 요건을 충족할 경우보다 객관적으로 종합적 진실을 전달해 사실과 근접할 것이라 판단하는 것이다. 이 사례들의 경우 형식적 요건 충족 여부 판단을 사람의 내용 분석으로 수행했다. 연구자로 이루어진 평가자들이 각 기사의 내용을 읽고 이를 바탕으로 그 형식적 요건을 충족했는지를 분석한 것이다. 국내의 기사 892건을 분석해 고품질 기사 비율이 뉴욕타임스는 55.6%인 반면 국내 일간지 1면에 게재된 기사 중 좋은 기사의 비중은 7.5%로 나타났다는 '좋은 저널리즘 연구회'의 연구 결과(김경모·박재영·배정근·이나연·이재경, 2018)가 대표적인 내용분석의 결과물이다.

그런데, 이러한 형식적 요건들을 사람의 내용 분석이 아니라 기계가 자동으로 수행하기도 한다. 자연어 처리 기술(natural language processing)의 발전 덕분이다. 자연어 처리는 일종의 이야기로 기록돼 있는 뉴스 기사들을 문자 또는 문자열(string), 형태소(morpheme), 단어(word), 구(phrase), 절(clause), 단문과 복문 등의 문장(sentence), 문단(paragraph), 문서(document), 복사 문서 등 다양한 수준으로 쪼개어 분석하는 기술을 의미한다(Ingersoll, Morton, & Farris, 2013/2015, 41~42쪽). 자연어처리를 통해 뉴스 기사를 계량화한 후 뉴스 기사들의 형태소분석, 구문분석(syntax analysis), 의미분석, 담론분석 등을 실시해 뉴스 기사들의 내용과 특정 형식적 요건들을 자동으로 추출하는 것이다(박대민, 2016)⁴⁾. 기계에 의한 형식적 요건의 자동 추출 사례로 대표적인 것이 구글 뉴스의 랭킹 알고리즘과 뉴스트러스트위원회 뉴스트러스트 알고리즘이다.

구글은 "뉴스 기사들의 순위 책정을 향상시키기 위한 시스템과 방법(Systems and methods for improving the ranking of news articles)"이라는 특허(Curtiss, Bharat, & Schmitt, 2012)를 통해 뉴스 기사의 순위를 책정하기 위한 13가지 요인을 밝히고 있다. 언론사의 뉴스 생산량, 뉴스의 길이, 보도의 중요성, 속보성(breaking news score), 이용 패턴(usage patterns), 언론사에 대한 사용자 조사(survey), 사용자 수 및 트래픽, 뉴스룸의 크기, 지국(bureaus)의 수, 실명 인용 보도의 수, 언론사의 보도 범위, 보도 뉴스의 글로벌 도달률, 글쓰기 스타일 등 언론사 외형에 대한 평가와 기사 내 형식적 요건에 대한 평가를 자동으로 수행해 뉴스 기사의 순위를 산출하고 있다. 단, 구글은 각 요인들의 구체적 산출 기준 및 가중치 등은 공개하지 않고 있다.

뉴스트러스트위원회의 뉴스트러스트 알고리즘은 총 11가지 요인을 공개하고 구체적 산출 기준 및 가중치 등도 공개했다(뉴스트러스트위원회, 2018)⁵⁾. 뉴스트러스트위원회는 기사명, 기사의 길이, 인용문 수, 제목의 길이, 제목의 물음표, 느낌표 수, 수치 인용 수, 이미지의 수, 평균 문장의 길이, 제목의 부사 수, 문장당 평균 부사 수, 기사 본문 중 인용문 비중 등 11개 요인을 뉴스트러스트 알고리즘에 적용했다. 뉴스트러스트위원회는 알고리즘 중간 공개를 통해

4) 형태소분석이란 문서, 문장 등 입력된 문자열을 분석하여 형태소를 기본 단위로 분류하는 것이며, 구문분석은 형태소들이 결합하여 구문이나 문장을 만드는 규칙인 통사규칙에 따라 문장 내에서 형태소의 역할이나 상호관계를 분석하는 것이다. 의미분석은 구문분석 결과를 해석하여 문장 내 단어의 의미를 보다 명확하게 구분하고 문장성분 간 의미관계를 파악하는 작업이다. 담론분석은 문맥 속에서 단어나 문장 등에 어떤 의미가 있는지 분석하는 것이다.

5) <http://www.kpf.or.kr/site/kpf/03/10309000000002016110402.jsp>

11개 요인의 기준을 구체적으로 공개했는데, 예를 들어 기사의 길이의 경우 길이에 따라 0-1 점 사이 가점을 부여했다. 좀 더 구체적으로는 기사의 분류, 매체 유형(신문, 방송 등) 등에 따라 기사의 길이 평균을 추출한 후 평균 이상인 기사에 대해 표준편차에 따라 단계별 가점을 부여했다. 이에 따라 기사의 길이는 0, 0.165, 0.33, 0.495, 0.66, 0.835, 1 등 7단계의 값을 갖게 된다. 각 11가지 요인에 따라 값을 부여하고 그 값이 가장 높은 기사의 신뢰도를 가장 높게 평가하는 방식이다.

물론, 구글의 뉴스랭킹 및 뉴스트러스트 알고리즘은 PEI 등 사람의 내용분석처럼 복잡한 형식적 요건을 파악하지는 못한다. 인용문의 수, 기사의 길이, 실명 기자명 등 기초적인 수준의 형식적 요건만을 자동으로 추출해 분석하고 있다. 하지만, 인공지능 기술의 발전에 따라 자동으로 추출하는 형식적 요건의 수와 깊이가 늘어나고 있다. 아직까지는 내용의 맥락까지 이해하여 추출하는 것은 아니지만, 사람의 기존 내용분석을 기계가 학습하는 방식으로 그 분석의 깊이를 더해가고 있다. 또한, 우리나라에서만 하루 수만 건의 기사가 생산되는 상황에서 사람이 그 내용을 일일이 분석하기는 힘들다. 뉴스 기사 등 정보의 양은 앞으로 급증할 것이기에 기계에 의한 자동 분석 시도는 점점 늘어날 것으로 예측된다.

3) 자동화된 사실 확인으로

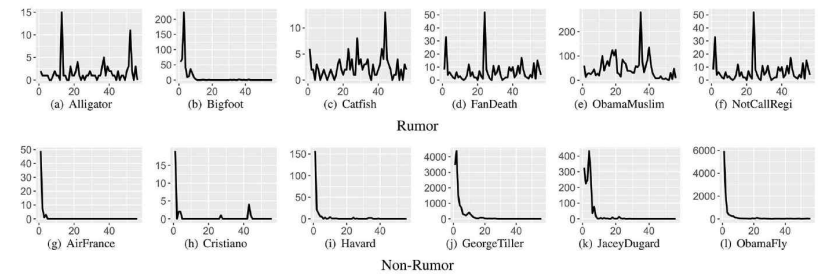
형식적 요건들을 기계에 의해 자동으로 확인하려는 시도들은 자연스럽게 사실(fact)의 확인을 자동으로 하려는 시도로도 이어지고 있다. 앞서 2절에서 살펴본 바와 같이 좋은 기사로서 특정 요건들을 갖춘다면 좋은 기사라 판단할 수 있고, 좋은 기사는 다른 기사들에 비해 사실에 근접해 있다고 평가할 수 있기 때문이다. 현재 수많은 뉴스와 정보들이 거의 실시간으로 전달되고 있다. 이러한 가운데 가짜 뉴스 등 허위 및 거짓 정보가 확산되면서 전달되고 있는 정보들에 대한 검증 요구가 높아지고 있다. 가짜 뉴스 등 허위정보로 인한 문제와 관련해 법률을 통한 규제, 팩트체크(fact checking) 시스템의 구축, 뉴스 생산 및 유통의 투명성 강화, 저널리즘 품질 고양, 미디어 교육 확대, 알고리즘 등 기술의 힘을 통한 가짜 뉴스 검증 등 다양한 대응책들이 제시되고 있다(오세욱·정세훈·박아란, 2017). 하지만, 법률 규제의 경우 표현의 자유 등을 침해할 우려가 있으며, 미디어 교육 확대, 저널리즘 품질 고양, 기술적 대응 등은 그 효과가 나타날 때까지 시간이 상당히 소요된다는 문제가 있다. 이에 따라 즉각적인 효과를 발휘할 수 있는 대응 방안으로서 '팩트 체크'가 주목받고 있다.

문제는 기준과 같이 소수의 전문화된 인력만으로는 수많은 정보들을 모두 확인하기 어렵다. 디지털화에 따라 수많은 뉴스와 정보, 발언 등이 기록되어 나타나면서 무엇이 사실인지를 사람이 일일이 확인한다는 것은 거의 어려운 일이기 때문에, 정보량 증가에 따라 사실을 자동으로 확인(automate factchecking)하는 기술이 등장한 것이다. 오세욱(2017)은 사실 자동 확인 기술의 등장 배경을 세 가지로 제시한다.

첫째, '사실 확인'을 해야 할 텍스트 자체의 급격한 증가다. 유명인, 정치인 등 사실 여부를 확인해야 할 주요 인사들의 발언 창구가 2010년대 이후 신문, 방송 등 기존 언론을 넘어 트위터, 페이스북 등 SNS와 팟캐스트 등으로까지 확대됐다. 또한, 이들의 발언 내용을 임의로 편집한 이용자 생성 콘텐츠들까지 기하급수적으로 늘어났기 때문에 기계의 힘을 빌리는 것이다. 둘째, 확인하려는 사실을 선정함에 있어서 개입되는 편견의 배제다. 그 많은 정보 중 왜 꼭 특정 사실에 대해서만 확인하려고 했는지에 대해 설명할 수 있어야 하기 때문이다. 왜 이 발언을 확인하는지 왜 이 사실을 확인하는지에 대한 합리적, 객관적 설명이 없다면 '사실 확

인'은 그 근거를 잃는데, 사실 자동 확인 기술은 그러한 우려에서 벗어날 수 있기 때문이다. 셋째, 뉴스 기사 유통의 플랫폼화다. 플랫폼들은 뉴스를 직접 생산하기보다 유통에만 주력하기 때문에 사실 확인 등을 수행할 전문 인력을 보유하지 않고 있다. 기술기업으로서 플랫폼들은 '사실 확인'에 있어서도 사람보다는 기술의 힘을 선호하고 있다. 가짜뉴스 등 허위 및 거짓 정보에 대한 우려가 커짐에 따라 사실 확인을 위한 플랫폼의 책임을 강조할수록 사실을 자동으로 확인하려는 플랫폼의 욕구도 높아지기 때문에 사실 자동 확인 기술이 점점 확산되고 있다는 것이다.

하지만, 사실 자동 확인 기술들은 사람과 같이 복합적 맥락을 고려하여 사실 여부를 판단하기보다는 특정 기술적 요건의 충족 여부에 따라 이를 판단하고 있다. 어떤 정치인이 특정 발언을 했다고 한다면, 이 발언을 하게 된 맥락을 고려하기보다는 이 발언이 과거 발언과 일치하는지 여부, 기존에 쌓여 있던 통계 등 정보와의 일치 여부 등을 판단한다. 이는 지식 기반(knowledge based) 방식이라고 할 수 있는데, 가능한 많은 정보 및 문서들을 확보한 후 새로 등장한 사실과 비교해 사실 여부를 확인하는 것을 의미한다. 지식 기반 방식은 기록된 데이터를 바탕으로 하기 때문에 비교적 정확하다는 장점이 있지만, 방대한 양의 정보를 데이터베이스화해야 하며, 기록되지 않은 데이터 및 기록할 수 없는 데이터는 확인 불가능하다는 한계가 있다. 특정한 정보의 전파 패턴을 분석해 자동으로 사실 여부를 확인하기도 한다. 예를 들어, 완전히 틀린 내용은 제대로 된 내용보다 확산 속도에 있어서 차이를 보인다는 연구들(Mocanu et al., 2015, Kwon et al., 2017)이 대표적이다.



[그림 1] 루머(위)와 비루머(아래)의 트위터 전파 패턴 비교(Kwon et al., 2017, p. 7)

[그림 1]처럼 이러한 자동 확인 사실 확인 기술은 사실 정보가 확산되는 과정과 거짓 정보가 확산되는 과정을 분석해 어떤 정보가 확산되면서 이용자들이 어떠한 반응을 보이고, 그 확산의 과정이 어떤 패턴을 보이는지에 따라 사실 여부를 확인하는 방식이다. 이는 SNS 등에서 정보 전파 유형을 통해 사실을 확인하는 것으로 사실 자체에 대해 확인하는 것은 아니라고 할 수 있다. 진짜 정보이지만 너무나 놀라운 내용이라 루머처럼 빠르게 확산되는 정보의 경우 진짜 정보임에도 루머로 판단할 수 있는 한계가 있다.

아직까지 사실 자동 확인 기술은 완성된 기술은 아니다. 그보다는 사실 여부 확인이 필요한 정보를 빠르게 찾아주는데 유용한 기술이라고 할 수 있다(오세욱, 2017). 최종적인 사실 여부 판단은 아직까지는 인간의 몫으로 남아 있다. 이야기로서 정보는 인간의 창작 영역이기 때문에 유형화하거나 숨겨진 맥락까지 데이터화하는 현재의 기술 수준으로는 어렵기 때문이다. 사람의 사실 여부 판단 내용을 기계에게 학습시켜 이를 해결하려는 '클레임버스터

(ClaimBuster)’ 등과 같은 사실 자동 확인 기술도 있지만, 클레임버스터도 최종적인 사실 여부를 확인하기보다는 사람들의 판단 경향에 따라 ‘팩트’일 확률만을 계산해 낼 뿐이다(Hassan et al., 2015).

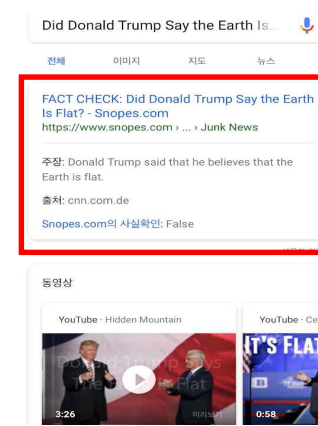
이러한 현실적 한계로 인해 현재 사실 자동 확인 기술과 관련해 가장 많이 쓰이고 있는 방식은 형식 기반(style based) 방식이다. 어떤 콘텐츠가 사전에 정의한 혹은 통계적으로 추론된 형식을 준수하면 그 내용을 사실과 가깝다고 판단하는 방식이다. 예를 들어, 루빈 등(Rubin et al., 2015)은 풍자 뉴스를 자동으로 식별해 냈다. 이들은 180개의 뉴스 기사를 분석한 뒤 이 중 사람이 풍자 뉴스로 구분한 뉴스 기사들에서 자주 쓰이는 단어들의 유형을 추출했다. 이후 새로운 뉴스 기사들을 대상으로 이러한 단어들이 등장할 경우 풍자 뉴스로 자동 판별했다. 비슷한 방식으로 사실 정보와 거짓 정보의 유형을 추출해 ‘팩트’ 여부를 확인하는 시도들도 있지만, 기존 유형을 추출해 비교하는 방식은 기존 유형을 벗어나는 새로운 유형이 등장할 경우에는 오류를 범할 가능성이 크다. 이러한 오류를 방지하고자 사전에 정의한 형식의 충족 여부를 기준으로 판단하기도 한다. 예를 들어, 구글은 특정 콘텐츠가 ‘클레임리뷰(ClaimReview)’라는 구조화된 데이터 형식을 준수하고 있을 경우 ‘팩트체크’ 기사로 간주하여 검색 결과에서 별도로 표기한다. ‘클레임리뷰’에 대해서는 다음 장에서 구체적으로 살펴본다.

3. ‘팩트’의 형식적 구성 요건 사례

자동 사실 확인 기술들은 완전하게 사실 여부를 판단하기보다는 사람의 최종적인 사실 확인이 필요한 내용을 빠르게 찾아주는 수준에 있다고 할 수 있다. 사실이라는 것은 종합적 진실이기 때문에 고정되어 있지는 않기 때문이다. 사전 등을 활용한 지식 기반 방식, 기계 학습 방식, 정보 전파 유형을 분석하는 맥락 분석 방식 등 모두 이로 인해 근본적인 한계를 가질 수밖에 없다. 이러한 한계를 감안하며, 특정한 형식을 충족하고 있을 경우 보다 사실에 근접한 것으로 판단할 수 있을 것이라는 가정하에 형식 기반 방식이 현재 자동 사실 확인 기술에서 가장 효율적으로 활용될 수 있는 모델로 보인다. 2장에서 살펴본 바와 같이 품질 높은 기사들이 가져야 할 요건들을 기술적으로 미리 규정하고 이를 충족하고 있는지를 확인해 보다 사실에 근접한 내용이라고 판단하는 것이다. 이와 관련해 대표적으로 참고할만한 사례들은 ‘클레임리뷰 마크업(ClaimReview Markup)’, BBC의 ‘스토리라인 온톨로지(Storyline Ontology)’, ‘투명성 선언(Transparency Initiative)’의 ‘hnews’ 마이크로 포맷 등이다.

1) 클레임리뷰 마크업

[그림 2]는 “도널드 트럼프가 세상은 평평하다고 말했나?”(Did Donald Trump say the earth is flat?)라는 내용을 구글 모바일 앱에서 검색한 결과다. 검색 결과 최상단에는 빨간 상자 속 내용처럼 미국의 팩트체크 전문 매체인 ‘스노프스(Snopes)’의 기사 링크와 함께 ‘주장’, ‘(주장의)출처’, 사실 확인 결과가 함께 제공된다. 일반적인 검색 결과에서는 검색어에 맞는 문서의 간략 정보와 링크가 제공되는 것에 반해 [그림 2]에서는 주장, 출처, 사실확인 결과 등의 내용이 더해져서 제공되고 있다. 또한, 팩트체크 기사가 최상단 검색 결과로 노출되고 있다. 구글이 검색어와 관련한 기사의 내용을 분석해 이러한 정보들을 자동으로 추출해 최상단 검색 결과로 제공한 것일까? 정답은 “아니”다. 구글은 이 팩트체크 기사의 내용을 분석한 것이 아니라, 스노프스가 정해진 팩트체크 형식(style)을 준수했기 때문에 해당 검색어에 대한



[그림 2] 구글 모바일 앱 검색결과

결과로 [그림 2]와 같이 자동 배열된 것이다.

구글 검색 가이드의 사실 확인 안내 페이지⁶⁾는 “웹 페이지에 ClaimReview 구조화된 데이터 요소가 포함되어 있으면, 해당 주장에 관한 검색결과에 내 페이지가 표시될 때 내 페이지의 사실 확인을 요약한 내용이 함께 표시됩니다”라고 제시한다. 클레임리뷰라는 마크업(markup)을 기술적으로 준수하면 ‘팩트체크’ 기사로 기계가 자동 인지하는 방식이다. 마크업은 웹상에서 문서(document) 처리를 지원하기 위해 문서에 추가되는 정보, 문서의 논리적 구성 등 포맷, 문서 내용의 찾아보기 정보, 다른 문서와의 연결 정보 등을 정의한다. 클레임리뷰 마크업은 이 문서가 특정 주장에 대해 평가를 하는 내용을 담고 있음을 입증하기 위해 지켜야 할 요건을 제시한 것이다. 이 마크업이 정의한 방식대로 특정 요건의 정보를 기술적으로 담고 있을 경우 ‘클레임리뷰’ 즉, ‘팩트체크’ 정보라고 인식한다. ‘클레

임리뷰 마크업’은 주장을 검증하기 위해 지정된 일종의 컴퓨터 언어 형식이다. 한편, ‘클레임리뷰 마크업’은 스키마닷오알지(Schema.org)에 등록돼 있다. 스키마닷오알지⁷⁾는 지난 2011년 구글, 마이크로소프트, 야후가 정형 데이터 검색 기능을 높이기 위해 협력해 만든 기구다.

클레임리뷰 마크업은 각각의 요건들을 [그림 3]과 같이 속성(property), 기대 유형(expected type), 설명(description)으로 정의한다. 클레임리뷰 1개, 리뷰 3개, 크리에이티브워크(CreativeWork) 82개, 부가정보(Thing) 11개 등 4개 분류에 총 97개의 요건을 제시한다. 클레임리뷰는 문서 내에 검증하는 주장의 요약이 담겨 있어야 함을 정의한 것이며, 기대 유형은 그 요건을 제시하는 방식이 텍스트, 이미지 등 어떤 유형이면 좋을지를 정의한 것이다. 크리에이티브워크는 사건 발생 시간, 출처 정보, 등장인물, 후원자, 수정 내역 등 주장을 검증하는 내용으로 갖춰야 할 요건들을 정의한 것이며, 부가정보는 식별자(identifier), url 등 해당 내용과 관련한 추가 정보들의 제시 방식을 정의한 것이다. 클레임리뷰와 리뷰는 클레임리뷰 마크업을 준수하고 있음을 밝히도록 하는 장치로 반드시 준수해야 한다. 크리에이티브워크는 이 내용이 주장 검증, 즉, 팩트 체크 내용을 담고 있음을 기술적으로 확인하기 위해 정의한 요건들이다. 아직까지 컴퓨터가 내용을 읽고 판단하는 수준에 이르지 못했기 때문에, 특정 형식적 요건을 미리 정의해 놓고 이를 준수할 경우 ‘팩트 체크’임을 자동으로 인식하는 방식이다.

6) <https://developers.google.com/search/docs/data-types/factcheck>

7) <https://en.wikipedia.org/wiki/Schema.org>

[그림 3] 클레임리뷰 마크업 내용

Property	Expected Type	Description
Properties from ClaimReview		
claimReviewed	Text	A short summary of the specific claims reviewed in a ClaimReview.
Properties from Review		
itemReviewed	Thing	The item that is being reviewed/rated.
reviewAspect	Text	This Review or Rating is relevant to this part or facet of the itemReviewed.
reviewBody	Text	The actual body of the review.
reviewRating	Rating	The rating given in this review. Note that reviews can themselves be rated. The <code>reviewRating</code> applies to rating given by the review. The <code>aggregateRating</code> property applies to the review itself, as a creative work.
Properties from CreativeWork		
about	Thing	The subject matter of the content. Inverse property: <code>subjectOf</code> .
accessMode	Text	The human sensory perceptual system or cognitive faculty through which a person may process or perceive information. Expected values include: auditory, tactile, textual, visual, colorDependent, chartOnVisual, chemOnVisual, diagramOnVisual, mathOnVisual, musicOnVisual, textOnVisual.

* 출처 : <https://pending.webschemas.org/ClaimReview>

클레임리뷰마크업은 스키마닷오알지의 목적처럼 정형 데이터 검색에서 '팩트 체크'와 관련된 더 많은 정보를 더 효율적으로 전달하기 위해 만든 형식이다. 검색 효율성을 위해 만든 형식적 요건이지만, '팩트'를 검증하는 팩트 체크 기사의 특성과 관련해 의미를 지닌다. 사건 발생 시간, 출처 정보, 등장인물, 후원자, 수정 내역 등은 사실 여부를 확인하는 팩트 체크 내용에 반드시 담겨야 할 정보라고 할 수 있다. 또한, 이는 기사의 품질을 측정하는 요인이기도 하며, 이러한 내용을 기술적으로 준수하게 함으로써 투명성도 높일 수 있다. 검색 효율을 높이기 위한 방안이었지만, '팩트'를 다루는 정보의 객관성, 투명성을 높이기 위한 형식적 요건을 제시한 것과 같다.

2) 스토리라인 온톨로지⁸⁾

BBC는 지난 2013년 “뉴스 스토리라인을 표상하는 온톨로지(An ontology to represent News Storylines)”⁹⁾를 발표했다. 온톨로지란 철학적 의미에서는 존재론이지만, 컴퓨팅적 의미에서는 실제 세계의 현상을 정보자원화하면서 컴퓨터가 해석할 수 있는 의미로 변환한 메타 데이터를 의미한다. 즉, “세상에 대해 서로 보고 듣고 느끼고 생각하는 바에 대해 합의를 이룬 후 이를 개념화하여 컴퓨터에서 다룰 수 있는 형태로 표현한 모델”¹⁰⁾을 말한다. 이때, 개념의 유형과 제약 조건들은 명시적으로 정의되어야 한다. 스토리라인 온톨로지는 이야기로서

8) 스토리라인 온톨로지 관련 내용은 “김선호·박대민·오세욱 (2015). 스트럭처 저널리즘, 데이터 저널리즘을 넘어서. <해피미디어동향> 2015-06.”의 33-37쪽 내용을 재구성하였다.

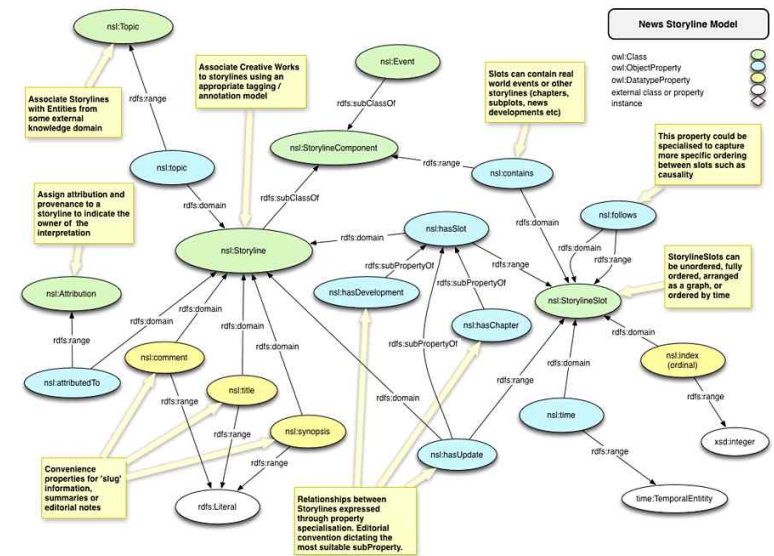
9) <http://www.bbc.co.uk/ontologies/storyline>

10) <http://ko.wikipedia.org/wiki/온톨로지>

뉴스가 디지털 영역에서 어떻게 존재해야 하는지를 개념화한 것으로 뉴스가 스토리가 되기까지 각 단계에 존재하는 구조요소들을 개념화하여 각각 하나의 데이터가 될 수 있는 개념으로 제시했다. 즉, 디지털 영역에서 뉴스로 존재할 수 있는 각각의 형식적 요건들을 정의한 것이다.

BBC는 ‘스토리’를 두 가지 측면으로 구분한다. 한국어로는 이야기로 번역되지만, 뉴스의 관점에서 스토리는 기사와 리포트같이 사실 중심인 콘텐츠의 개별 조각인 동시에, 발생한 사건에 대한 해당 뉴스 조직의 편집 관점이 될 수도 있다. 저널리스트들은 벌어진 사건의 중요성을 설명하기 위해 정보, 사실, 의견, 인용, 그리고 데이터를 끌어오며, 내러티브를 만들어내기 위해 맥락을 끌어온다. “남대문이 화재로 전소했다”는 팩트이지만, “남대문 화재는 관리 당국의 문제였다”는 관점이 되는 것이다. BBC는 창작 작업(작성된 기사)과 사건에 대한 편집 관점 사이의 모호함을 제거하기 위해 스토리라인이라는 용어를 사용하고 있다.

[그림 4] BBC의 뉴스 스토리라인 온톨로지



* 출처 : <http://ichef.live.bbci.co.uk/images/ic/912xn/p01r1pm1g.png>

[그림 4]처럼 BBC의 뉴스 스토리라인 온톨로지는 스토리로서 뉴스가 만들어지기까지 각 과정을 분석한 후 그 과정들을 각각 하나의 개념으로 제시한다. 이를 컴퓨터와 커뮤니케이션이 가능할 수 있도록 데이터화할 것을 제안하고 있다. 뉴스 스토리라인 온톨로지는 뉴스 조직이 말하고자 하는 이야기를 설명하고 조직하기 위한 총괄적인 모델이다. 또한, 다양한 뉴스 제작 조직을 위해 유연하게 구성되어 있다. 각각 하나의 개념으로 제시되는 과정들이 '팩트'를 전달하는 뉴스의 형식적 요건이 된다.

온톨로지 속 각각은 개별적으로 존재하지만, 이들은 항상 더 큰 집합 속에 속한다. BBC 온톨로지에서 이 각각을 포괄하는 집합이 '슬롯(slot)'이다. 슬롯은 각 지표 속성에 따라 인과 관계 혹은 순서를 갖고 있으며, 시간적으로 나열될 수도 있고, 그래프 형태로도 표현이 가능하다. 스토리라인을 구성하는 요소들은 실제 세계의 사건 혹은 다른 스토리라인들(채터, 서브플롯, 업데이트, 뉴스 전개 등)에서 명백하게 나타나는 것들이기 때문이다. 또한, 스토리라인은 몇몇 지식 영역(사람, 장소, 조직 등)에서 다양한 주제들로 묶일 수 있다. 물론 뉴스의 스토리라인 온톨로지에 존재하고 주요 층위(class)는 속성(Attribution), 사건(Event), 스토리라인(Storyline), 스토리라인 구성요소(StorylineComponent), 스토리라인 슬롯(StorylineSlot), 토픽(Topic) 등이다.

먼저, 스토리라인의 속성은 스토리라인의 해석을 몇몇 다른 행위자(agent)들에게 돌리기 위해 사용된다. 일반적으로 출처의 의미로 사용된다. 사건은 뉴스 가치가 있는 사건으로 논쟁의 여지가 없는 실제 세계의 사건을 의미한다. 스토리라인(Storyline)은 뉴스의 전체적인 스토리라인을 말하며, 스토리라인 구성요소는 사건, 스토리라인 등과 같은 스토리라인 구성요소들의 포괄하는 추상화된 상위 클래스(parent class)다. 스토리라인 슬롯은 스토리라인 구성요소들을 포괄하는 용도로 사용되며, 토픽은 시간, 장소 등 일반적인 지식 영역에서 사용되고 있는 몇몇 개념에 대한 인터페이스다. 즉, 스토리라인 구성요소는 스토리라인, 사건, 속성 등을 포괄하는 상위 층위이며, 이 스토리라인을 포괄하는 상위 층위가 스토리라인 슬롯이다. 이 슬롯들은 컴퓨팅적 의미가 아닌 일반적 의미에서 사용되고 있는 토픽에 따라 상호 묶이거나 관련되어 배열될 수 있다. 이 모든 요소들을 포괄하는 상위 층위가 슬롯이 되며, 슬롯들이 모여 하나의 관계를 형성하면서 주제에 따라 시계열적 배치를 하거나, 병렬적인 그래프 배치 등이 가능하다.

하지만, 스토리라인 온톨로지는 개별 기사를 염두에 두기보다는 개별 기사를 향후 구조적으로 활용하는 것을 목적으로 하고 있다는 점에서 클레임리뷰 마크업과는 다르다고 할 수 있다. 즉, 클레임리뷰 마크업이 특정 기사가 팩트체크 기사임을 입증하기 위한 형식적 요건을 규정했다면, 스토리라인 온톨로지는 특정 기사의 형식적 요건들이 향후 새로운 사실 관계 구축의 한 요소로 활용될 수 있도록 구조적으로 정의한 것이다.

3) 투명성 선언

미디어 스탠다즈 트러스트(Media Standards Trust, 이하 MST)는 뉴스의 품질(quality), 투명성(transparency), 책무성(accountability)을 증진하기 위한 방안을 연구하기 위한 공익단체로 지난 2006년 영국에서 설립됐다. MST는 언론의 자유에 헌신하고 특정 정당이나 미디어 기업과 제휴하지 않으며, 영국 국민을 대신해 뉴스 미디어의 향상된 기준(high standards)을 조성하는 것을 목표로 하였지만, 현재는 운영되지 않고 있다. MST는 출범과 함께 월드 와이드 웹(WWW)의 기반을 설계한 것으로 유명한 컴퓨터 과학자 팀 버너스 리와 공동으로 '투명성 선언(Transparency Initiative)'을 제안했다. 온라인 뉴스의 출처나 변경 사항을 메타데이터(metadata)로 관리하여 뉴스 기사의 투명성을 제고하자는 목적이었다.

'hNews'¹¹⁾라 불리는 이 메타데이터는 [그림 5]의 상단처럼 기사의 HTML 소스 상단에 해

11) <http://microformats.org/wiki/hnews>

당 기사의 작성자, 언론사명, 작성 시각, 언론사의 윤리 강령(principles), 위치정보(geo), 출처 정보 등을 기록한 것이다. 기존의 기사 형식에 단순히 추가(add-on)되기 때문에 마이크로 포맷(micro-format)이라 불리기도 한다. 마이크로 포맷이란 웹 페이지의 메타데이터 표현을 위해 통일된 규격을 뜻한다. 이 규격을 통해 기계적인 작업을 쉽게 하거나, 브라우저에 여러 메타데이터를 제공해줄 수 있다. 기존 HTML 포맷에 마이크로 포맷을 추가하는 것은 어려운 작업이 아니며, HTML 속성을 추가적으로 지정해주기만 하면 가능하다. 대표적인 마이크로 포맷에는 'hAtom', 'hCard', 'hCalendar' 등이 있다. 'hNews'는 'hAtom'에서 파생된 마이크로 포맷으로서 뉴스 기사의 메타데이터를 제공하기 위한 전용 포맷이라고 할 수 있다.

[그림 5] '투명성 선언'의 메타데이터 소스와 라이선스

```
<dl class="profile">
  <dt>class</dt>
  <dd>
    <span rel="help" href="http://www.w3.org/2004/10/struct/globel.html#del-class">
      HTML4 definition of the 'class' attribute.</span>
      This meta data profile defines some 'class' attribute values (class names)
      and their meanings as suggested by a
      <a href="http://www.w3.org/2004/10/struct/globel.html#del-link">draft of
      "RDF-based Links in HTML"</a>.</span>
    </dd>

  <dt>news</dt>
  <dd>Used to describe semantic information associated with news stories.</dd>

  <dt>source-org</dt>
  <dd>The originating organization for the news story.</dd>

  <dt>date-line</dt>
  <dd>represents the location where the news story was filed.</dd>

  <dt>geo</dt>
  <dd>represents geographic coordinates of relevant locations in the story.</dd>

</dl>
</dt>
</dd>
<dt>relic</dt>
<dd>
  <dt>license</dt>
  <dd>represents the license for the story.</dd>

  <dt>principles</dt>
  <dd>represents the statement of principles and ethics used by the news organization that produced the news story.</dd>
</dd>
</dt>
</dd>

Renewing US ties with Latin America, Obama says he wants to lead, not lecture, on democracy
by BEN FELLER,Associated Press - 19 April 2009 18:17 GMT
PORT-OF-SPAIN, Trinidad (AP) -- Defending his brand of world politics, President Barack Obama said Sunday...
Both Graham and McCaskill spoke on "Fox News Sunday." Ensign was interviewed on CNN's "State of the Union."
* 출처 : http://microformats.org/wiki/hnews
```

언론사들은 자신들이 작성하던 형식으로 기사를 작성한 뒤 'hNews' 마이크로 포맷을 추가해도 웹 브라우저에 표현되는 기사의 형태는 변화가 없어 적용이 편리하다는 장점이 있다. 또한, 여러 신문사들이 공통의 포맷으로 개별 기사에 대한 정보를 기록하면 언론사에 관계없이 기사를 한 시스템으로 관리할 수 있어 기사의 출처와 생성 및 변경 사항 파악이 용이하다는 장점이 있다. 즉, 해당 기사의 작성자, 언론사명, 작성 시각, 언론사의 윤리 강령(principles), 위치정보(geo), 출처 정보, 수정내역 등을 기사 내용과 상관없이 코드 형식으로 별도 기록하게 함으로써 기사의 투명성을 높이고 사실 여부 확인을 위한 기준을 부여한 것이다. 투명성 선언 측은 'hNews' 포맷을 통해 언론사들이 출처가 어디인지를 명확히 밝히면 그 출처를 더욱 명확히 할 수 있고, 링크 정보가 제공되면 그 정보도 더욱 확실할 수 있으며, 블로그 등과 같은

정보에도 이용자들이 쉽게 접근할 수 있을 것이라고 강조했다. 몇 가지 필수적인 정보를 기사 내용과 별도로 형식적으로 강제함으로써 기사 자체에 대한 신뢰도를 높이는 것이다. 투명성 선언은 6가지지만 '팩트'를 전달하는 기사가 반드시 포함해야 할 형식적 요건을 규정했다.

단, 투명성 선언은 현재 재원 부족을 이유로 더 이상 운영하지 않고 있다. 투명성 선언의 활동이 지속되기 위해서는 언론사 기사 입력 소프트웨어 등에 접목할 플러그인 등 도구의 개발이 필요한데 이러한 작업은 현재 중단된 상태다. 이전에 개발했던 내용만이 깃허브 등에 소스코드 형식으로 남아 있다.

4. SNU 팩트체크 정보의 메타데이터 분석

SNU 팩트체크 또한 검증 내용의 투명성을 추구하고자 클레임리뷰마크업 및 투명성선언 등의 사례를 참고하여 사건 발생 시간, 출처 정보, 수정 내역 등 팩트체크 콘텐츠가 반드시 포함해야 할 정보들을 구성하여 메타데이터를 구축했다. SNU 팩트체크의 메타데이터는, 각 제휴 언론사에서 SNU 팩트체크 홈페이지¹²⁾에 등록하는 팩트체크 콘텐츠를 기초 데이터(raw data)로 하며, 이를 가공하여 축적하는 방식으로 구성된다. 제휴 언론사는 자체적으로 수행한 팩트체크를 통해 작성된 기사를 SNU 팩트체크의 검증 형식에 맞춰 정리하여 게시하고, SNU 팩트체크는 게시된 콘텐츠에 기초한 모니터링 과정을 통해 검증 내용을 세분화하여 메타데이터로 축적하고 있다. 즉, 참여 언론사는 기사 형식으로 작성된 내용을 SNU 팩트체크가 제시한 형식적 요건에 맞춰 SNU 팩트체크 정보로 등록하고 이 내용을 SNU 팩트체크의 운영 요원들이 검증하여 메타데이터를 추가하는 방식으로 관리하고 있다. 3장에서 분석한 사례들처럼 SNU 팩트체크는 팩트체크 기사의 형식적 요건을 정한 후 그 요건에 맞도록 메타데이터를 구성해 관리한다.

1) SNU 팩트체크의 콘텐츠 등록 절차

SNU 팩트체크의 콘텐츠 등록 절차는 크게 2단계(무엇을 검증할 것인가와 관련한 '팩트'의 추가, 어떻게 검증할 것인가와 관련한 '팩트체크' 방식의 추가)로 구분할 수 있으며, 세부적인 페이지의 이동까지 모두 고려했을 때 5단계(팩트체크 참여-검증할 팩트 목록-팩트 추가-검증 참여-검증 결과 노출)로 나뉘볼 수 있다. SNU 팩트체크에서는 검증하고자 하는 대상 및 관련 내용을 '팩트'라고 부르며, 검증이 완료된 하나의 콘텐츠를 '배지'로 부른다.

① 팩트체크 참여

제휴 언론사에서 팩트체크를 수행하는 첫 단계는 팩트체크 참여로, 홈페이지 우측 상단에 위치한 '팩트체크 참여' 버튼 클릭을 통해 검증 과정을 시작하게 된다. 해당 메뉴는 팩트체크를 수행하는 언론사 계정으로 접속하였을 때만 활성화되는 메뉴로, 이후 메인 페이지에서 검증할 내용을 등록할 수 있는 '검증할 팩트 목록' 페이지로 이동하게 된다.

② 검증할 팩트 목록

12) <http://factcheck.snu.ac.kr/>

제휴 언론사는 해당 페이지에서 팩트체크가 완료된 콘텐츠뿐만 아니라 현재 제휴 언론사에서 검증을 진행 중인 팩트 또한 확인할 수 있으며, 이를 통해 다른 언론사에서 이미 관련 팩트체크를 수행하고 있는지를 확인할 수 있다. 해당 페이지에는 언론사가 검증하고자 하는 대상인 '팩트'가 등록 시간 순서대로 나열되어 있으며 구체적인 검증(팩트체크)이 완료되기 전까지 배지 하단에 위치하게 되는 언론사 로고는 공란으로 표시된다.

제휴 언론사에서는 이러한 정보를 확인 후, 검증할 콘텐츠를 다른 언론사에서 다룬 적이 없을 때는 '팩트 추가' 버튼을, 검증할 콘텐츠를 다른 언론사에서 이미 팩트체크를 실시해 교차 검증을 하고자 하면 '참여' 버튼을 클릭하여 목적에 맞게 세부 페이지로 이동하게 된다.

③ 팩트 추가

검증할 콘텐츠가 해당 언론사에 의해 처음 검증되는 내용일 때, 언론사에서는 총 7개의 정보를 입력하여 팩트를 추가하게 된다.

- 1) 검증대상 : 검증하고자 하는 사실 또는 주장을 제기한/발언한 주체(인물, 기관 등)를 입력한다. 인터넷에서 유포되고 있는 루머, 또는 언론사에서 자체적으로 제기한 사항, 또는 특정한 발언인/기관이 없는 경우 등의 검증 내용에 대해서는 '인터넷 루머', '언론 보도', '발언인 없음' 등의 구분을 두어 검증대상으로 등록하게 된다.
- 2) 팩트 테마 : 선거 및 특정 기간 동안 한시적으로 적용되는 메뉴로 관련 배지들만 정리하여 한 곳에서 확인할 수 있도록 구성된다. 예를 들어 검증하는 내용이 6.13 지방선거와 관련이 있는지에 따라 '6.13 지방선거' 또는 '관련 없음' 중 선택할 수 있도록 나뉘어 있다.
- 3) 팩트 유형: 검증하고자 하는 내용이 어떠한 성격을 지니고 있는지 선택할 수 있는 영역이다. 관련 영역에는 1) 정치인(공직자)의 발언, 2) 정치인(공직자)과 관련된 사실, 그리고 3) 기타가 있다.
- 4) 토픽 : 검증하고자 하는 영역이 뉴스의 섹션과 같은 구분을 둘 때 어떠한 내용을 담고 있는지를 복수로 선택할 수 있다. 섹션의 구분은 총 7개로, 정치, 경제, 국제, 사회, 문화, IT/과학, 기타 중 검증 내용에 따라 선택할 수 있다.
- 5) 헤드라인 : 검증하는 내용의 제목 역할을 하는 정보로, 구체적인 발언 문장, 또는 주장 내용을 서술한다. 120자 제한을 두어 명료한 문장으로 정리될 수 있도록 하였다.
- 6) 검증 대상 보충설명 : 7가지 정보 중 유일한 선택 입력 사항으로, 검증하는 내용의 상향적 정보를 입력하는 공간이다. 해당 발언, 또는 사실이 발생한 장소, 시간 등 사용자가 맥락을 쉽게 파악할 수 있도록 돕기 위해 만들어진 항목이다.
- 7) 검증 대상 출처: 검증 내용이 근거가 되는 출처를 입력하는 공간으로, 검증 내용의 발생 장소 및 날짜, 루머의 근원지 등을 구체적으로 입력하는 영역이다.

④ (검증) 참여

'팩트 추가' 페이지에서 무엇을 검증할 것인지에 대한 구체적인 정보를 입력하여 팩트를 추가하게 되면 '검증할 팩트 목록' 페이지 최상단에 해당 언론사에서 추가한 '팩트'가 배치되게

된다. 이때 다른 언론사에서 검증한, 또는 검증하고자 등록한 '팩트'가 겹치게 될 경우 언론사는 해당 팩트에 검증을 '참여'하여 교차 검증을 수행할 수 있으며 교차 검증의 경우 또한 아래와 같은 방식으로 검증 과정을 입력하여 진행된다. 추가된 팩트의 우측 '참여' 버튼을 클릭하면 어떻게 팩트체크하였는지를 세부적으로 입력할 수 있는 '검증참여' 페이지로 이동하게 된다. '검증 참여' 페이지에서 언론사는 구체적인 검증 내용을 위해 아래와 같은 4개의 정보를 입력하게 된다.

- 1) 본문 : 어떻게 검증하였는지를 구체적으로 서술할 수 있는 영역으로 텍스트, 하이퍼링크, 이미지 등을 첨부할 수 있다. 글자 수에 제한은 없다.
- 2) 평가점수 : 검증한 내용에 기초하여 최종 팩트체크 결과를 정리하는 영역으로, 총 6가지 보기 중 결과에 맞춰 언론사가 선택하게 된다. 최종 평가 지표는 사실성에 근거하여, 사실/대체로 사실/절반의 사실/대체로 사실 아님/전혀 사실 아님/판단 유보 중 선택하게 된다.
- 3) 검증기사 : 언론사에서 관련 내용으로 게시한 기사의 URL 링크를 복수로 입력할 수 있는 공간으로, 관련 URL을 입력하면 자동으로 인식하여 하이퍼링크를 생성하고 기사의 링크에서 제목을 추출하여 한글로 변환하여 제공한다.
- 4) 근거자료 : 선택적으로 입력할 수 있는 공간으로, 언론사의 팩트체크 검증 및 평가에 근거가 된 자료를 정리하여 입력하는 영역이다.

⑤ 검증 결과 노출

선택사항을 제외한 모든 항목을 작성한 다음 제출하면 모든 절차가 완료되며 SNU팩트체크 메인 페이지의 가장 상단에 해당 배지가 게시된다. 배지에는 검증대상, 헤드라인, 팩트유형, 토픽, 평가점수 및 검증 언론사 정보가 담긴 카드 형태의 검증 내용이 등록 순서대로 정렬된다. 등록 순서는, 검증 내용의 완료뿐 아니라, 교차 검증, 또는 검증 내용의 수정이 있을 경우에도 가장 상단으로 배치된다.

이를 통해 사용자들이 이미 확인하였던 검증 내용에 대해서 수정 및 추가 사항이 있을 때 쉽게 확인할 수 있다. (언론사에서 게시한 팩트체크 콘텐츠의) 세부 내용을 확인하고자 할 때는 '헤드라인' 문장을 클릭하면 팩트체크 상세보기 페이지로 이동하여 검증 내용 전문을 확인할 수 있다. 2개 이상의 언론사가 교차검증한 콘텐츠의 경우 배지 하단에 참여한 언론사의 로고가 복수로 나타나며 (상세보기 페이지에서 등록 시간에 따라 확인할 수 있다. 또한) 메인 페이지의 '핫이슈' 버튼을 클릭하면 교차 검증이 진행된 배지만을 따로 확인할 수 있다.

2) SNU 팩트체크의 메타데이터 구성

앞선 1절에서의 순서를 통해 SNU 팩트체크에 게시된 팩트체크 정보는 정해진 요건에 따라 메타데이터로 추출된다. 최종 검증된 정보는 그 담고 있는 형식적 특성에 따라 '팩트' 데이터와 '팩트체크' 데이터로 구분하여 [그림 8]과 같이 메타데이터로 정리된다.

[그림 8] SNU 팩트체크 정보의 메타데이터 추출 화면

팩트	팩트번호	2
	토픽(정치, 경제 등)	정치
	테마(19대 대선, 새정부 등)	19대 대선
	팩트유형(검증 유형)	정치인(공직자)의 발언
	기타성격	책임없음
	팩트발생시간	2017.01.03 00:00
	발언주체	이재명
	발언대상	미군주둔비
	발언내용	"전 세계 미군 진주지역에서 대한민국이 진주비 부담이 크다. 독일은 18%, 일본은 50% 선인데 우리는 77% 부담하고 있다"
	사실/타주체	책임없음
사실/타대상	책임없음	
사실/타내용	책임없음	
검증팩트와영향력	1. (내외비 실시간 검색어) 해당 없음 2. (SNU 게시일 기준 기사 수) 0개	
팩트체크	참여언론사	중앙일보
	최초등록시간	2017.03.29 11:03
	검증내용	(생략)
	검증방식	기사 활용, 관련인 발언 활용, 법률/합정 활용
	검증근거/출처	1. The New York Times, What the U.S. Gets for Defending Its Allies and Interests Abroad "(내용에는 15일, 실제는 16일)
		2. CNN, Full Donald Trump and Wolff Blitzer Interview: Part 1 (YouTube 링크, CNN 홈페이지에 녹취록도 있음)
		3. Vincent K. Brooks 주원미군사령관, 2016.04.19, 인준청문회의(Senate Armed Services Committee hearing on Gen. Brooks nomination)
		4. 주한미군지위협정 (SOFA) 대한민국과 아메리카 합중국간의 상호방위조약 제4조에 의한 시설과 구역 및 대한민국에서의 합중국 군대의 지위에 관한 협정
		5. 김동업 국동문제연구소 교수
	검증근거/출처URL	1. https://www.nytimes.com/interactive/2017/01/16/world/trump-military-role-treaties-allies-nato-asia-persian-gulf.html
2. https://www.youtube.com/watch?v=14cXVxCrqPc		
3. http://www.army.mil/article/166787		
4. http://mofa.go.kr/webmodule/hisboard/template/read/korboardread.jsp?typelD=24&boardid=11695&seqno=3		
5. 자체섭외		
검증기사제목	[팩트체크] 이재명 "77%로 미군 주둔비 부담 커... 일본은 50%"	
검증기사URL	http://news.joins.com/article/21367474?cloc=Joongangjipreelection2017/factcheck	
최초검증결과	대체로 거짓	
검증결과수정여부	No	
수정검증결과	책임없음	
검증결과수정시간	책임없음	

팩트 데이터는 검증된 내용 중 일어난 사실에 대해서만 정리한 13개의 관련 데이터를 말하며, 팩트체크 데이터는 검증된 내용 중 언론사에서 팩트체크한 내용을 분석하여 정리하는 14개의 관련 데이터를 의미한다. 이 중 언론사에서 입력한 부분이 자동으로 추출되는 부분이 있으며, 모니터링을 통해 SNU 팩트체크에서 수동으로 추출하여 정리하는 데이터가 있다. 테마, 토픽 및 최초 검증결과 등 자동으로 추출되는 데이터는 언론사가 검증할 내용을 추가하는 단계에서 선택하게 되는 항목 값으로, SNU 팩트체크에서 미리 구분해둔 정보이기 때문에 이를 추출해내는 것은 값의 정보만 받아올 때에도 확보가 가능하다. 하지만 검증 내용, 근거 자료 등과 같이 세부적인 유효화가 추가로 필요하고 맥락의 이해를 통해서만 추출될 수 있는 정보들은 현재의 기술 수준에서 데이터화하기 어렵다. 팩트 발생 시간, 검증 팩트의 영향력 등과 같이 직접적인 검색 과정을 요구하는 정보 또한 위와 같은 이유로 현재 모니터링 과정을 통해 수동으로 추출하고 있다. 메타데이터를 구성하는 팩트 데이터의 내용은 [표 1]과 같으며, 팩트체크 데이터의 자세한 내용은 [표 2]와 같다.

[표 1] SNU 팩트체크 정보 메타데이터 중 팩트 데이터

항목	설명	자동 추출 여부
서버 번호	· 게시된 배지의 서버 등록 번호	○
도판	· 검증 내용이 해당되는 섹션 (예: 정치, 경제, 사회 등)	○
테마	· 검증 내용이 속하는 특정 카테고리 (예: 19대 대선, 6.13 지방선거, 관련 없음 등)	○
팩트 유형(발언/사실)	· 검증 내용 성격이 특정 인물/기관의 발언 또는 사실과 관련되었는지 여부	○
팩트 유형(기타)	· 검증 내용 성격이 기타일 때의 세부적 구분 (예: 루머, 외신, 수치, 기사, 관련 항목 없음, 해당 없음 등)	×
팩트 발생 시간	· 검증 내용의 구체적 발생 시간	×
발언 주체	· (검증 내용이 발언일 경우) 해당 발언을 한 주체의 이름	○
발언 대상	· (검증 내용이 발언일 경우) 해당 발언을 한 주체가 지목한 대상의 이름	×
발언 내용	· (검증 내용이 발언일 경우) 발언 주체의 정확한 발언 문장/단어	×
사실/기타 주체	· (검증 내용이 관련 사실 또는 기타일 경우) 해당 사실에 문제를 제기한 주체의 이름	○
사실/기타 대상	· (검증 내용이 관련 사실 또는 기타일 경우) 해당 사실과 관련된/지목된 대상의 이름	×
사실/기타 내용	· (검증 내용이 관련 사실 또는 기타일 경우) 관련 사실 또는 기타에 해당하는 내용	×
검증 팩트의 영향력	· 검증 내용의 영향력 지표(실시간 검색어 순위, 게시 후 1주일간 생성된 관련 기사수)	×

[표 2] SNU 팩트체크 정보 메타데이터 중 팩트체크 데이터

항목	설명	자동 추출 여부
참여 언론사	· 검증을 수행한 언론사의 이름	○
최초 등록시간	· 언론사가 검증 내용을 게시한 시간	○
검증 내용	· 언론사의 검증 내용 전문	○

검증 방식	· 검증을 위해 근거로 제시한 자료의 유형(예: 기사, 관련인 발언, 법률/협정, 통계, 보고서, 논문, 기타)	×
검증 근거/출처	· 검증에 사용된 자료의 세부 명칭	대부분 수동 추출
검증 근거/출처 URL	· 검증 근거/출처로 삼은 자료를 확인할 수 있는 URL주소	대부분 수동 추출
검증 기사 제목	· 출처로 등록된 기사의 제목	○
검증 기사 URL	· 출처로 등록된 기사의 URL	○
최초 검증 결과	· 언론사의 검증 결과(예: 사실, 대체로 사실, 절반의 사실, 대체로 사실아님, 전혀 사실아님, 판단 유보)	○
검증결과 수정여부	· 검증 결과의 수정 여부	○
수정검증 결과	· 수정 후 변경된 검증 결과	○
검증결과 수정시간	· 검증 결과가 수정된 시간	○
검증결과 수정이유	· 검증 결과가 수정된 이유	○
팩트검증 기사 영향력	· 검증 기사로 등록된 뉴스의 포털에서의 반응 정도(게시 후 1주일)	×

3) SNU 팩트체크 메타데이터 분석

SNU 팩트체크에 등록된 정보들의 메타데이터에 기초하여 데이터가 어떻게 축적되고 있는지를 분석했다. SNU 팩트체크 정보의 메타데이터의 경우 3장의 사례들처럼 코드화해 등록하기보다는 모니터링 팀이 별도의 파일로 구축하고 있다. 모니터링 팀이 구축한 파일 내용을 바탕으로 어떠한 데이터가 축적되었는지를 분석했다. 분석 대상은 SNU 팩트체크 서비스가 시작된 2017년 3월 29일부터 2018년 6월 30일까지 740개의 정보였다. 언론사명, 서버번호, 테마 등 강제로 할당돼 등록돼 관리되는 데이터를 제외하고 총 10개의 메타데이터를 추출해 현황을 분석하였다. 각 메타데이터를 전수 수집한 후 이를 분석하였는데, 구체적인 분석 대상 및 결과는 [표 3]과 같다.

[표 3] SNU 팩트체크 정보 메타데이터 분석 결과

항목	자동 추출 데이터			수동 추출 데이터
	(언론사) 작성 여부	총 수	누락 배지 여부 및 갯수	수동 추출 결과
검증대상	○	740(총 검증 수)	×	×
팩트유형	○	740(총 검증 수)	×	×

토픽	○	74(총 검증 수)	×	×
검증기사	○	740(총 검증 수)	×	×
근거자료	○	625(총 근거자료 수)	○ (346)	근거/출처:2,242개 근거/출처 URL:1,513개
하이퍼링크	○	117개	○ (804)	×
결과 수정여부	○	153번	×	×
팩트 발생 시간	×	×	×	740
검증팩트영향력	×	×	×	740
팩트검증 기사 영향력	×	×	×	740

분석 기간 동안 언론사에서 게시한 배지 수는 총 675개로, 65번의 교차 검증을 포함하여 전체 검증은 총 740번 이뤄졌다. 팩트 유형은 정치인(공직자)의 발언이 362개, 정치인(공직자)과 관련된 사실이 107개, 기타가 271개였으며, 검증된 내용의 토픽은 중복 내용을 포함하여 정치 370건, 사회 126건, 경제 79건, 국제 68건, IT/과학 45건, 기타 28건, 문화 14건 순서였다. 근거자료 입력에서는 346번의 언론사 팩트체크에서 근거 자료를 입력하지 않은 것으로 확인되었으며 총 625개의 근거자료가 입력되었다. 검증 내용 내에 입력되는 하이퍼링크는 총 117개의 배지에 입력되었으며, 전체 하이퍼링크의 수는 804개였다. 자동 추출한 메타데이터들은 언론사들이 정보 등록 당시 입력한 데이터들이다.

수동 추출된 메타데이터 정보의 경우 검증 방식은 총 1,253건으로 집계되었으며, 관련된 발언 활용 377건, 법률 활용 208건, 보고서 활용 153건의 순서로 팩트체크가 수행되었다. 검증 근거/출처는 총 2,242개가 확인되었으며, 이를 확인할 수 있는 URL은 총 1,506개가 확보되었다. 근거자료의 URL 도메인을 분석해본 결과, 정부 관련 부처의 웹사이트인 'go.kr'이 550개, '.com'이 343개, 'co.kr'이 278개의 순서로 많았으며, 마지막으로 사실, 대체로 사실, 전혀 사실 아님 등으로 언론사가 검증 내용을 사실성에 기초하여 최종으로 판단하게 되는 검증 결과를 언론사에서 변경한 검증 결과수정 여부는 총 153번으로 확인되었다.

SNU 팩트체크에 게시된 검증 내용의 메타데이터 분석 결과, 대부분 기사에서 팩트를 뒷받침하는 근거자료, 원문 출처 URL 등의 표기를 제대로 하고 있지 않은 것으로 확인되었다. 언론사가 입력한 근거자료 입력 수를 분석한 결과, 총 740번의 검증 중 625개(중복계산)였으며, 입력된 검증 수만 계산하면 약 394개에 불과하였다. 전체 검증 중 346개, 약 47%가 검증 내용으로 제시한 자료의 출처를 근거자료 항목에 입력하지 않았음을 의미한다. 또한 특정 법률 및 통계 등을 근거자료를 사용한 사례 중 해당 법률 및 통계의 정확한 명칭이 아닌 '소방안전법', 'UNDP 통계' 등과 같이 제시하여 어떠한 자료를 근거로 하였는지 정확히 확인할 수 없는 경우 또한 확인되었다.

검증 투명성 등을 위해 강조되는 원문 출처 URL의 경우도 740개의 검증 본문에 입력된 URL의 수는 117개의 검증에서 총 804개에 불과했다. 전체 팩트체크 검증의 84%를 차지하는 623개의 검증에서 원문 출처 URL를 제공하지 않았음을 의미한다. 즉, SNU 팩트체크에 언론사가 검증한 팩트의 경우 그들이 편견이 있을 수 있다는 점을 인정하고, 더 찾아볼 수 있는 서류나 정보원, 그리고 증거들에 대해서 링크를 제공함으로써 하여 확보되는 투명성 측면에서

매우 미약한 것으로 분석됐다.

한편, 모니터링을 통해 수동으로 분석해 찾아낸 검증 근거와 출처 수는 2,242개, URL의 수는 1,506개로 나타났는데, 이를 언론사에서 입력한 625개의 검증 근거 수와 비교해보면 1,617개나 많았으며 약 3.59배나 많은 양을 수동으로 확보할 수 있었다. 이러한 수치는 SNU 팩트체크에 게시된 팩트체크 정보들이 질차적 투명성에 대해 상당히 무관심하다는 점을 의미한다. 이는 검증 본문 내에 투명성과 관련해 아무런 정보를 제공하지 않는 것이 아니라, 기사 입력 당시 정의된 항목별 입력란에 해당하는 정보를 입력하지 않고 있음을 보여준다. 검증방식의 경우에서도 언론사가 입력할 수 있음에도, 대부분 입력하지 않고 수동으로 추출하고 있었다. 총 1,253건 중 관련인 발언 활용 377건, 법률 활용 208건, 보고서 활용 153건 등을 모두 모니터링 팀이 수동 추출했다. 또한, 검증결과를 수정할 경우 수정 이력과 '오탈자 수정'과 같은 단순한 내용만 남아 있을 뿐 어떤 내용의 수정이 구체적으로 이루어졌는지는 공개되지 않고 있었다. 즉, 수정 여부만 기록되어 있을 뿐 어떠한 내용이 수정되었는지 등은 전혀 등록돼 있지 않아 모니터링 팀의 수동 추출로도 이를 확인할 수 없었다.

5. 논의 및 결론

1) 논의

SNU 팩트체크는 '팩트' 데이터로서 토픽, 테마, 검증유형, 성격, 발생시간, 발언 주체/대상/내용, 사실 주체/대상/내용, 영향력 등을, '팩트체크' 데이터로서 언론사, 검증내용, 검증방식, 검증근거/출처, 검증근거/출처 URL, 검증결과, 수정여부, 수정이유, 영향력 등을 메타데이터로 관리하고 있다. 이러한 요인들은 팩트 체크 기사로서 갖춰야 할 형식적 요건이라고 할 수 있다. 하지만, 이러한 메타데이터의 수는 앞서 살펴본 '클레임리뷰' 등과 비교할 때 비교적 적다고 할 수 있으며, 그 입력 및 관리도 완전한 수준은 아니라고 할 수 있다. 입력의 경우 언론사들이 입력하는 경우보다 모니터링 팀의 수동 추출 수가 더 많았으며, 관리의 경우 추출된 메타데이터들을 기사 내 관련 정보로 포함하는 것이 아니라 별도의 파일로 관리하고 있어 통합적 관리가 이루어지지 못하고 있었다. 팩트 체크 기사에 있어서 투명성 윤리가 강조되고 있음을 고려할 때 이러한 메타데이터의 수를 더 늘려나가고, 입력 및 관리를 효율적으로 수행하기 위한 방법론의 개발이 필요할 것으로 보인다.

이를 위해서는 우선 SNU 팩트체크센터의 주도적 역할 뿐 아니라, '팩트' 데이터 관리 주체로서 참여 언론사들의 적극적인 형식적 요인 입력 작업이 필요하다. '팩트' 데이터 입력 작업을 별도로 수행하는 것을 업무 외 부가 작업으로 간주하는 경향이 존재하지만, 이 작업은 향후 '팩트'의 구성 요건을 확인하기 위한 기초 작업으로서 의미를 지닐 수 있다. 현재 SNU 팩트체크가 모니터링 팀을 통해 진행하고 있는 수동 메타데이터 추출 작업도 궁극적으로 팩트 체크 기사를 작성하는 언론사들이 기사 작성단계에서 염두에 두고 수행해야 할 작업이라고 할 수 있다. 디지털 저널리즘에서 핵심은 기자의 의무가 이제 단순한 게이트키퍼이거나 뉴스 보도 해설자가 아니라는 점이다. 어떤 정보가 믿을 수 있는지 확인하고, 그 정보를 다시 정리해 이용자들이 그것을 효과적으로 얻고 이해할 수 있게 해주는 것이다.

특히, 투명성의 원칙은 사실 확인의 규율에서 중요한 요소로, 취재원과 취재 방법에 대한 공개, 취재원들은 누구이며 그들이 가진 지식의 성격과 편견, 충돌하는 내용이나 모르는 내용이 무엇인지를 투명하게 드러내고 정직하게 대하면 객관성의 추구에서 발생하는 문제는 대체

로 해결될 수 있다(Kovach & Rosenstiel, 2007/2009, 146-155쪽). 투명성의 미덕은 기자가 수용자를 존중한다는 것을 보여준다는 점에 있으며, 또한 취재원을 드러내고 취재 과정을 공개하여 취재원의 오류나 의도적인 사기를 막을 수 있으며, 수용자가 이를 공정하게 판단할 수 있는 기회를 제공한다. 기자는 투명성을 추구하면서 공익적인 의무를 다하고 있다는 확인을 할 수 있다. 형식적 요건을 미리 정하는 것이 기자의 자유로운 기사 작성에 방해가 될 수도 있겠지만, 그보다는 변화한 환경에서 새로운 저널리즘을 시작하는 계기가 될 수도 있다.

또한, 이러한 형식적 요건의 준수는 디지털 환경에서 믿을 수 있는 정보의 제공을 위한 유용한 수단이 될 수 있다. 우리나라에서만 하루 수만 건의 기사들이 제공되는 상황에서 어떤 기사를 더 믿을 수 있는지를 이용자가 판단하는 것은 사실상 불가능하다. 또한, 대부분의 뉴스 이용이 플랫폼에서 이루어지고 있다. 플랫폼은 몇몇 언론사가 아닌 거의 대부분의 언론사들의 뉴스 기사를 모아 보여준다. 이러한 틀을 뉴스 형식의 허위정보들이 파고들면서 '페이스북 뉴스'와 같은 소동이 발생한 것이다. 어떠한 허위정보가 유통되고 있을 때, 이를 확인하기 위한 믿을 수 있는 뉴스 기사를 찾기 어렵기 때문이다. 지금까지 논의한 객관성을 대신한 투명성의 차원에서 뉴스 기사들이 지켜야 할 형식적 요건을 충족한 경우 이를 보다 신뢰할 수 있는 기사로 간주할 수 있을 것이다. 이러한 형식적 요건들을 지키기가 까다롭고 구체적으로 '팩트'에 대한 내용 확인이 이루어졌을 경우에 가능하도록 구성하면 된다. 단순히 기사 내용으로 모든 것을 말하는 것이 허위정보 등과 구분될 수 있는 형식적 요건의 준수가 새로운 대안이 될 수 있다. 형식적 요건의 준수를 코드화할 경우 플랫폼 기업 등에서는 자동으로 이를 확인할 수 있다. 형식적 요건을 준수했다는 기술적 신호를 플랫폼 기업이 받아서 처리하면 되기 때문이다. 플랫폼 기업들이 그 신호를 받아 해당 기사를 검색 순위나 기사 배열 과정에서 좀 더 높은 순위로 배치할 수 있으며, 특정 이슈에 대해서는 해당 내용을 최우선으로 노출할 수 있다. 일일이 확인할 수 없는 수많은 정보들 중에서 그나마 믿을 수 있는 정보로 판단할 기준을 갖고 있기 때문이다.

물론, 이를 위해서는 구체적으로 지켜야 할 형식적 요건이 무엇인가에 대한 이론적 논의와 함께 언론사들이 기사 작성 단계에서 이러한 요건을 지켜야 하는 실질적 적용 방안이 함께 논의되어야 한다. 박아란·이나연·정은령(2018)은 '팩트체크 저널리즘'의 기본 원칙으로서 다음과 같이 8가지를 제시한다. 1) 검증 대상은 사실 여부를 가릴 수 있는 것이어야 한다. 2) 조사(research)는 발언자로부터 시작해야 한다. 3) 조사의 출처는 편중되지 않아야 한다. 4) 증거는 물적 토대를 가진 것이어야 한다. 5) 최대한 익명 인용을 배제한다. 6) 자료의 출처를 독자들에게 투명하게 밝힌다. 7) 판정결과는 가장 나중에 밝힌다. 8) 오류는 공개적으로 즉각 수정한다 등이다. 이러한 내용을 형식적 요건으로 규정할 후 기술적으로 코드화할 수 있다. 검증 대상을 명확히 입력하도록 하고, 발언자를 명시하도록 하고, 조사의 출처를 반드시 명기하는 방식 등을 적용한 후 이를 기사 내용 외에 기사의 메타데이터로서 누구나 읽어 들일 수 있도록 반드시 입력하는 것이다. 이러한 입력을 현실화하기 위해 앞서 제시한 'hnews'와 같은 마이크로 포맷 형식의 코드를 적용한 뒤 이 코드를 '클레임리뷰마크업' 등과 같이 플랫폼 기업이 읽어 들이도록 하는 것이다.

2) 결론을 대신한 제안

SNU 팩트체크는 언론사들과의 제휴를 통해 국내에서 가장 많은 팩트체크 정보를 제공하고 있다. 이 연구는 SNU 팩트체크가 언론사들이 팩트체크 기사에서 지켜야 할 기본적 원칙을 제

시하고 이를 실질적으로 담을 기술적 도구를 개발해 다양한 기술 플랫폼 기업들과 협력 논의에 나설 것을 제안한다. 팩트체크 저널리즘이 지켜야 할 원칙을 형식적 요건으로 규정하고 이를 기술적 신호로 만들어 기술 플랫폼들이 기술적으로 읽을 수 있게 하는 방안이다.

이러한 제안이 실현된다면, 그 실질적 효과는 다음과 같이 세 가지 정도라고 할 수 있다. 첫째, 디지털 저널리즘 측면에서 투명성의 강화다. 사실 우리 언론들은 기사의 내용에만 주력할 뿐 다른 것에는 크게 신경을 쓰지 않는다. 하지만, 기사가 유통되는 환경은 디지털이 아니면 상상할 수 없고 디지털은 기록된 모든 것을 확인할 수 있게 한다. 사실의 확인은 사실 굉장히 논쟁적이다. '팩트'라는 것 자체가 복합적으로 구성되기 때문이다. 팩트체크에서 형식적 요건의 준수는 복합적으로 구성되는 '팩트'를 상호 검증이 가능하도록 유도하는 한 수단으로써 저널리즘 투명성 강화를 유도할 수 있다.

둘째, 허위정보 대응 수단으로서 가능성이다. 뉴스 형식의 허위정보가 확산되는 상황에서 믿을 수 있는 정보를 제시할 수 있다. 대부분의 허위정보들은 뉴스 기사의 형식을 차용할 뿐이다. SNU 팩트체크는 협력 모델을 구성한 후 어느 정도 검증된 언론사들과만 제휴를 맺고 있다. 이들이 검증한 내용을 형식적 요건에 따라 좀 더 까다롭게 구성할 경우 허위정보에 대응할 믿을 수 있는 정보로 제공이 가능하다. 또한, 이 요건을 기술적으로 구성해 플랫폼 기업들이 이를 자신들의 검색이나 배열에 활용하도록 협력이 이루어진다면 허위정보 대응의 실질적 방안이 될 수 있다.

셋째, 자동 사실 확인 기술에 대한 경계다. 전술했듯이 자동으로 사실을 확인하려는 기술적 시도들은 앞으로 지속적으로 이루어질 것이다. 하지만, 사실은 복합적으로 구성되기 때문에 최종적인 사실 여부는 여전히 사람에게 의해 이루어질 가능성이 크다. 이러한 형식적 요건을 준수하더라도 그것이 사실임을 입증해 주지는 않는다. 사실을 자동으로 확인할 수 있다는 기술적 주장들에 대해 형식적 요건에 대한 논의는 많은 시사점을 줄 수 있다.

이 글은 새로운 저널리즘 윤리로서 투명성을 제시하고 사실 자동 확인 기술 중 형식적 요건을 활용하는 사례들을 제시하고 이와 관련해 SNU 팩트체크 정보의 메타데이터들을 분석했다. 이를 통해 SNU 팩트체크가 중심이 돼 형식적 요건을 정하고 이의 입력을 강제화할 수 있는 기술적 도구의 개발을 제안하고 이를 코드화해 플랫폼 기업에 제공하는 협의에 나설 것을 제안했다. 이루어질 경우 언론사의 입장에서 자신들이 공들여 만든 정보들이 플랫폼에서 좀 더 비중 있게 제공될 수 있는 효과를 누릴 수 있으며, 이용자들은 믿을 수 있는 정보들을 좀 더 쉽게 접할 수 있으며, 이는 궁극적으로 한국 언론의 전반적인 신뢰도 향상으로 이어질 수도 있다.

물론, 이러한 제안이 현실적으로 실현되는 과정은 어렵고 지날 것이다. 기사 중심 생산에 익숙해져 있는 기존 언론사들의 생산 관행을 벗어나 메타 데이터를 위한 별도의 입력 작업이 필요하기 때문에 이에 대한 반발이 있을 수 있다. 또, 언론사마다 상이한 기사 입력용 소프트웨어를 사용하고 있음을 감안하면, 표준 코드 삽입을 별도의 범용 도구를 개발해 제공할 필요가 있다. 'SNU 팩트체크'가 이러한 개발의 주체로 나서야 할 것으로 보이지만, 실제 개발까지 이르기 위해서는 언론사들과의 협업 또한 필요하다. 언론사들의 협력을 통해 이를 이루어낸다고 하더라도 플랫폼 기업들이 받아들이지 않는다면 아무 소용이 없을 수도 있다. 또한, 기술적 형식만을 충족시키며 확산만을 노리는 '어뷰징'이 나타날 가능성도 있다. SNU 팩트체크 메타데이터 분석에서 논의한 바와 같이 언론사들의 인식 전환과 그에 따른 실천이 우선되어야 하며, 플랫폼 기업이 받아들일 수 있을 만큼의 신뢰성도 확보해야 한다. 어려운 작업이지만 허위 정보 등으로 인한 사회적 문제를 고려할 때 시도해볼 만한 제안이라고 생각한다.

<참고문헌>

- 김경모·박재영·배정근·이나연·이재경 (2018). <기사의 품질: 한국 일간지와 해외 유력지 비교연구>. 서울: 이화여대출판문화원.
- 김선호·박대민·오세욱 (2015). 스트럭처 저널리즘, 데이터 저널리즘을 넘어서. <해외미디어동향> 2015-06.
- 뉴스트러스트위원회 (2018). <뉴스 트러스트 위원회 경과 및 뉴스트러스트 알고리즘 도출 결과 설명>. 서울: 한국언론진흥재단.
- 박대민 (2016). 뉴스 기사의 자연어처리: <뉴스소스 베타>를 중심으로. <커뮤니케이션이론> 12권 1호, 4-52쪽.
- 박아란·이나연·정은령 (2018). <팩트체크 저널리즘의 주요 원칙>. 한국언론진흥재단 연구서 2018-01.
- 박재영·이완수 (2010). <뉴스평가지수의 개발과 적용>. 서울: 한국언론진흥재단.
- 오세욱 (2017). 자동화된 사실 확인(fact checking) 기술(technology)의 현황과 한계. <사이버커뮤니케이션학보> 34권 3호, 137-180쪽.
- 오세욱·정세훈·박아란 (2017). <가짜 뉴스 현황과 문제점>. 서울: 한국언론진흥재단.
- 이민웅 (2002). 좋은 뉴스의 으뜸가는 조건으로서 진실 보도-사실, 사회적 구성, 진실 보도, 재귀성(reflexivity). <언론과사회>, 제10권 3호, 9-51쪽.
- 이준웅 (2010). 한국 언론의 경향성과 이른바 '사실과 의견의 분리' 문제. <한국언론학보> 54권 2호, 187-209쪽.
- Bogart, L. (1989). *Press and Public: Who Reads What, When, Where, and Why in American Newspapers*. Lawrence Erlbaum and Associates.
- Craft, S. & Heim, K. (2009). Transparency in Journalism: Meanings, Merits, and Risks. Wilkins, L. & Christians, C.(eds.) *THE HANDBOOK OF MASS MEDIA ETHICS*(pp. 217-228). Routledge: NewYork.
- Curtiss, M., Bharat, K., & Schmitt, M. (2012). Systems and methods for improving the ranking of news articles: Google Patents.
- Hassan, N., Adair, B., Hamilton, J. T., Li, C., Tremayne, M., Yang, J., & Yu, C. (2015). *The Quest to Automate Fact-Checking*. University of Texas Arlington.
- Ingersoll, G. S., Morton, T. S., & Farris, A. L. (2013). *Taming text: How to find, organize, and manipulate it*. New York: Manning.: 임혜연 (역)(2015). <자연어텍스트 처리를 통한 검색 시스템 구축: 아파치 솔라, 루씬, OpenNLP 등 오픈소스 활용>. 의왕: 에이콘.
- Kovach, B. & Rosenstiel, T. (2007). *The elements of journalism*. Danvers, MA: Three Rivers Press. 이재경(역). <저널리즘의 기본 원칙>. 서울: 한국언론진흥재단.
- Kwon, S., Cha, M., & Jung, K. (2017). Rumor Detection over Varying Time Windows. *PLoS ONE*, 12(1).
- Mcbride, K. & Rosenstiel, T. (2014). *The new ethics of journalism: principles for the 21st century*. London: Sage. 임영호 역(2015). <디지털 시대의 저널리즘 윤리>. 서울: 한국언론진흥재단.

- Merrill, J. C. & Lowenstein, R.(1971). *Media Messages and Men: New Perspectives in Communication*. David McKay Co: New York. 오인환 외 공역(1975). <매스컴과 사회>. 서울: 삼영사.
- Mocanu, D., Rossi, L., Zhang, Q., Karsai, M. & Quattrociocchi, W. (2015). Collective Attention in the Age of (Mis)Information. *Computers in Human Behavior* 51(Part B), pp. 1198-1204.
- Rubin, V., Conroy, N., Chen, Y. & Cornwell, S. (2016). Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News. In *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, pp. 7-17.
- Schudson, M. (2001). The objectivity norm in American journalism. *Journalism*, 2(2), pp. 149-170.
- Searle, J. (1995). *The construction of social reality*. New York, NY: The Free Press.
- Tuchman, G. (1978). *Making news: A study in the construction of reality*. New York: Free Press.
- Ward, S. (2004). *The invention of journalism ethics*. Montreal: McGill-Queens University Press. 이은택 역(2007). <언론 윤리의 재발견: 객관성에 대한 역사적 철학적 탐구>. 서울: 방송통신대학교 출판부.

[Abstract]

**An Exploratory Study on the Formal Requirements of 'Fact'
: Suggestion through Analysis of SNU FactCheck Information
Metadata**

Se-Uk Oh

Senior Researcher, Korea Press Foundation

Gu Hyun Hwang

Researcher, SNU FactCheck

As the misinformation spreads, 'fact check' is attracting attention as one of the possible countermeasures. However, it is difficult and time-consuming for a person to confirm a vast amount of information. Accordingly, attempts have been made to solve these problems through automated factchecking technology. But, Automated factchecking technologies are at the level of quickly finding the facts that need to be confirmed, rather than ultimately confirming the facts. In order to solve these problems, attempts have been made to establish the technically formal requirements for the facts and to confirm the facts automatically according to whether they meet the requirements. The study analyzed the ClaimReview Markup, which is used in the fact check on Google News, the Storyline Ontology on the BBC, and the 'hnews' microformat on the Transparency Initiative. And We analyzed the metadata of the 'SNU fact check' information, which can be called the database of 'fact check' information in Korea. As a result of analyzing a total of 740 pieces of information, most of the articles showed that the facts supporting the facts and the URLs of the original texts are not properly displayed. These figures indicate that the SNU fact-check information is quite indifferent to procedural transparency, and this study suggests the formal requirements of the facts that fit the Korean context and the realization of them.

keywords : fact, fact check, objectiveness, transparency, metadata, platform