

# 국방부 디지털 정체성 관리를 위한 실시간 인터넷 문서 자동검색 및 분석 프로그램 발명에 관한 연구

장상훈\* · 이석준\*\*

## 《국문초록》

인터넷 미디어가 발전함에 따라 개인뿐만 아니라 정부도 디지털 정체성(인터넷 디지털 미디어에 의하여 만들어진 이미지)을 가지게 되었으며, 인터넷 미디어가 발전함에 따라 실제의 이미지 역시 디지털 정체성에 의하여 좌우될 가능성이 증가하고 있다. 따라서 기업들의 경우에는 디지털 정체성을 관리하고, 관련한 데이터를 분석하여 고객이 어떠한 이미지를 선호하는지에 맞추어 향후 정책적 의사결정 과정에 포함하는 선제적 활동을 진행하고 있다. 이러한 조치는 대민 신뢰도가 조직의 영향력으로 크게 작용하는 국방부와 같은 정부 기관에도 필요하다. 인터넷에서 현재 디지털 정체성이 어떻게 형성되어 있는지 실시간으로 확인하고, 특정 정책을 추진할 때의 반응은 어떤지 등의 활동을 통하여 최소한 가짜 뉴스로 인한 이미지 손실과 대응을 위한 인적 물적 손실을 막을 수 있고, 국방정책에 대한 실시간 국민들의 반응을 확인함으로써 국방의사결정에 중대한 참고자료를 확보함으로써 국민이 원하는 모습으로 성장하거나, 불가피한 경우 효과적인 대국민 설명 등이 가능할 것으로 판단된다. 이러한 필요성을 충족시킬 방안으로 실시간 인터넷 문서 자동 검색 및 분석 프로그램을 개발 및 발명하였고, 국방부 정책에 대한 실험을 통해 실효성을 함께 연구하였다.

주제어 : 디지털 정체성, 자동검색, 실시간, 인터넷 언론, 국방정책, 대변인

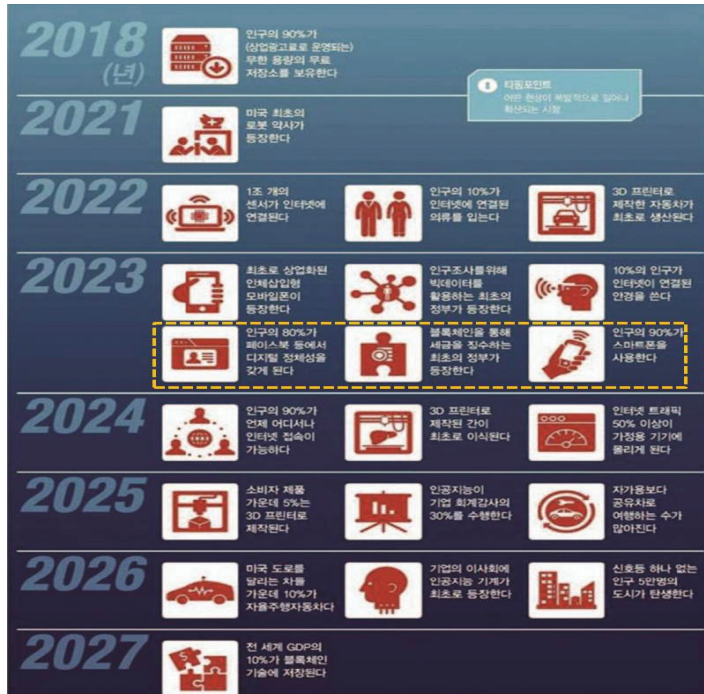
\* 광운대학교 방위사업학과 박사과정(방위사업청, 사업감사담당관), [jangsanghoon@korea.kr](mailto:jangsanghoon@korea.kr)

\*\* 광운대학교 경영학부 부교수, [sjlee@kw.ac.kr](mailto:sjlee@kw.ac.kr), <https://orcid.org/0000-0001-5608-2656>

# I. 서론

## 1.1 디지털 정체성 관리의 중요성

2016년 세계경제포럼은 향후 인구의 대부분은 디지털 상의 또 다른 자아를 가지게 될 것으로 판단된다고 발표하였다. <그림 1>의 기술적 티핑 포인트(tipping point)와 사회 충격 서베이 리포트에 따르면, 2023년 인구의 80%는 Social Network Service(이하 SNS) 등 인터넷에서 디지털 정체성을 가질 것으로 예측하고 있다.<sup>1)</sup>



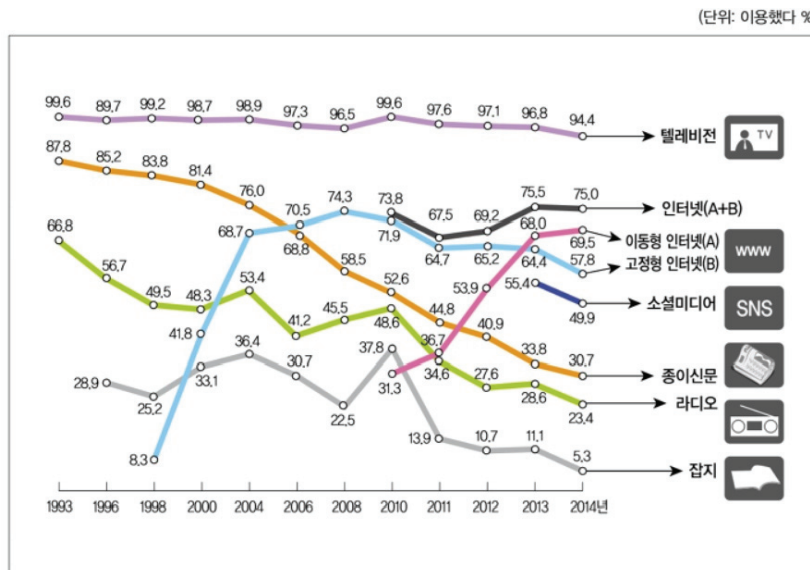
<그림 1> 기술적 티핑포인트(tipping point)와 사회 충격 서베이 리포트, 세계경제포럼

이를 디지털 자아라고도 하는데, ‘나는 누구인가’에 대한 스스로의 해답이 아니라, 인터넷과 현실이 연결된 사회에서 제3자에 의하여 생성된 정보 등에 의하여 재구성된 ‘자아’를 지칭한다. 즉, 다양한 디지털 미디어에서 본인과 관련된 정보의 조각들을 통합·분석하여 제3자가 제시한 자아로서, 디지털 데이터 기반의 자아(digital data-based self)라고 명명하기도 한다.<sup>2)</sup> 인터넷의 보급과 발달

1) WEF (2015. 9). Deep Shift: Technology Tipping Points and Societal Impact. Retrieved from [http://www3.weforum.org/docs/WEF\\_GAC15\\_Technological\\_Tipping\\_Points\\_report\\_2015.pdf](http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf).

로 디지털 기반의 자아는 개인을 넘어 기업 및 정부조직도 갖추고 있거나, 갖추게 될 것으로 판단된다. 즉, 기업 및 정부조직의 실제 모습뿐만 아니라, 디지털 미디어에서의 제3자로부터 제공된 정보와 함께 융합된 디지털 정체성을 이미 가지고 있거나, 가지게 될 것이다. 특히, 이들 기업 및 정부조직을 디지털 미디어로만 주로 접하는 대부분 국민 입장에서 이렇게 형성된 디지털 정체성만을 보고 기업 및 정부조직의 실체를 판단하게 될 가능성이 높다. 따라서 디지털 이미지의 중요성을 이미 깨달은 민간 기업에서는 디지털 정체성을 활용하여 고객과 교감하는 방식을 개선하고 있다. 즉 디지털에서의 좋은 이미지를 부각시키고 고객들이 선호하는 채널과 서비스를 제공하는 등의 활동을 통해서 기업의 이미지를 효과적으로 향상시킬 뿐만 아니라, 고객에게 제공하는 서비스를 효율적으로 발전시키기 위해 노력하고 있다.<sup>3)</sup> 또한 국방정책에서도 빅데이터 등을 활용하여 이를 부분적으로 적용시키기 위한 시도도 진행 중이다(이종하, 2017).

기업은 디지털 미디어 채널을 확인하여 좋은 이미지의 데이터 및 일반적인 문서 관리를 통해 기업의 인터넷 정체성 이미지를 좋은 방향으로 생성할 수 있다. 또, 고객에게 제공하는 서비스의 질에 대한 평가를 인터넷 문서들을 통해 확인하고, 니즈를 충족시키도록 개선할 수 있다. 오늘날 디지털 미디어 채널은 주로 인터넷으로 판단할 수 있고, 특히 정식적인 언론 외 대중의 의견까지도 블로그 및 홈페이지, SNS를 통하여 상당히 판단할 수 있다. <그림 2>와 같이 텔레비전의 정규방송이 아직



<그림 2> 미디어 이용률 추이(1993년~2014년), 한국 언론진흥재단

2) 정보통신정책연구원 (2013), 초연결 사회에서 디지털 자아의 정체성 연구(A Study of Digital Self Identity in Hyper-Connected Society), 미래창조과학부.  
 3) Rencher, B (2012). ADOBE DIGITAL MARKETING SUMMIT. <https://www.adobe.com/aboutadobe/pressroom/pressreleases/201203/032112AdobeDigitalMarketingSummit.html>

도 가장 영향력 있는 것으로 조사되었지만, 인터넷의 활용도가 상당히 높은 비율을 차지하고 있음을 알 수 있다. 특히 <표 1>과 같이 연령대를 고려했을 때, 20대와 30대는 절반 이상이 뉴스 미디어를 인터넷으로 접한다는 것을 알 수 있다. 이것은 시간이 지날수록 인터넷 언론의 파급력이 더 증가할 것으로 예측할 수 있다는 것을 의미한다. 뿐만 아니라, 10% 이상의 20대가 SNS를 이용하여 뉴스를 접하면서 정규방송 이외 개인들의 의견 등에도 노출되어 있을 것으로 판단된다.<sup>4)</sup> 또한, 현재는 대부분의 언론사가 텔레비전을 통해 방송한 내용을 인터넷 동영상 서비스 등 인터넷 채널을 통해 구독할 수 있게 해두었기 때문에 실질적으로 대부분의 언론정보는 인터넷에 존재한다고 볼 수 있다.

<표 1> 연령대별 뉴스 미디어별 이용시간 점유율

(단위 : %)

미디어 구분	전체 (n=5,061)	20대 (n=929)	30대 (n=968)	40대 (n=1,059)	50대 (n=989)	60대 이상 (n=1,116)
텔레비전	48.5	25.7	34.4	46.3	54.6	74.5
인터넷(A+B)	31.7	57.5	50.4	32.9	19.8	6.1
이동형인터넷(A)	17.5	33.9	28.2	18.2	9.8	2.6
고정형인터넷(B)	14.2	23.6	22.3	14.7	10.0	3.5
종이신문	10.9	4.9	6.2	11.7	16.4	13.0
라디오	4.6	1.0	3.0	5.2	7.1	5.7
소셜미디어	4.1	10.8	5.8	3.5	2.0	0.3
종이잡지	0.2	0.1	0.2	0.4	0.1	0.1
합계	100.0	100.0	100.0	100.0	100.0	100.0

인터넷에서 보도되거나 특정한 사건에 대한 인터넷의 의견들이 결국 디지털 정체성을 형성하는 가장 큰 데이터로 작용되므로 온라인에서의 평가는 매우 중요하다. 그렇게 형성된 평가가 다시 인터넷을 통해 대중들에게 노출되기 때문에 인터넷의 평가가 곧 전체 평가와 직결되어지기 쉽기 때문이다. 표 1을 보면 연령별 이동형 인터넷의 사용을 고려했을 때 온라인의 평가가 정체성에 직접적으로 영향을 주게 되는 증상은 점차 증가할 것으로 판단된다. 향후 국민들은 이동형 인터넷 등을 통해 인터넷에 지속적으로 노출되므로 대상에 대한 인터넷의 이미지가 더욱 중요해질 수 있어 정부조직 차원에도 동일하게 적용할 필요가 있다.

## 1.2 디지털 정체성 관리의 한계

정부조직의 디지털 정체성 관리가 현재에는 물론, 향후에는 더욱 중요해질 것으로 판단한 결과 우리나라는 2011년부터 온라인 대변인제도를 시행하여 정부부처는 온라인과 미디어 수단을 통해

4) 한국 언론진흥재단 (2014). 2014 언론수용자 의식조사. <https://www.slideshare.net/girujang/2014-45645888>



부처의 활동을 홍보하고 잘못된 보도에 해명하는 등의 활동을 수행하고 있다. 그런데도 인터넷에서 디지털 정체성에 대한 효과적인 관리가 쉽지 않다. 일례로, 그림 4와 같이 2017년 기준 연간 가짜뉴스의 추정발생 수는 13만 건이며, 이에 의한 피해는 연 30조로 추산되었으며, 특히 SNS의 영향력이 커지면서 그 심각성은 더욱 커지고 있는 것으로 판단된다(표 2)(정민, 백다미, 2017).

<표 2> 가짜 뉴스의 경제적 비용 추정

구분		피해금액
당사자 피해금액		22조 7,700억 원
	개인	5,400억 원
	기업	22조 2,300억 원
사회적 피해금액		7조 3,200억 원
합계		30조 900억 원

이렇게 가짜뉴스 등의 인터넷 언론의 부작용이 나타나는 이유와 대응의 한계점은 다음과 같이 예측된다. 첫째, 인터넷 언론이 발전되면서 인터넷 언론사의 수익 중 인터넷 광고비가 증가했고, 이를 위해 조회수 유도를 위한 무분별하고 자극적인 가짜 뉴스가 생산될 가능성이 높아졌다. 둘째, 가짜 뉴스가 생산되었거나 비슷한 소식을 접한 개인이 소셜 미디어(social media)를 통해 이 뉴스를 크게 확산시킬 수 있다. 셋째, 가짜 뉴스의 대상이 된 입장에서는 자신에 대해 인터넷 언론이 어떻게 형성되는지 24시간 감시하는 것이 불가능해서 초기에 대응이 지연된다. 넷째, 인터넷에서 새롭게 형성된 자신에 대한 문서를 발견해도 이를 분석하고 필요 시 해명자료를 작성하는데 걸리는 시간 동안 실시간으로 인터넷에서 해당 가짜 뉴스는 퍼져나간다. 또한, 관련 법률 역시 인터넷에서 피해를 입었다고 판단되는 경우 스스로 소명자료를 준비하도록 되어있다.<sup>5)</sup> 따라서 분석, 대응 중에 이미 이미지가 훼손 등 디지털 정체성에 피해를 입게 된다.

디지털 정체성 관리의 한계 중 큰 요소는 앞서 서술한 것과 같이, 가짜 뉴스 등에 대한 감시와 대응속도이다. 온라인 대변인실을 운영한다 하더라도 사람에 의해 인터넷 문서를 검색하여 관련된 내용을 찾고 대응하는 구조에 불과하다. 따라서 실시간으로 생산되고 순식간에 퍼져나가는 가짜뉴스의 속도를 감당할 수 없다. 또한 MIT의 연구에 따르면, 가짜 뉴스의 경우 진짜 뉴스에 비해 약 20배의 전파속도를 가지고 있다(윤인아, 2018). 따라서 프로그램에 의한 자동화 기술로 24시간 실시간 검색하고 내용을 분석하는 등의 방법이 그 대응방향으로 판단된다. 하지만 이것은 단순한 가짜뉴스 등에 의한 피해를 예방하는 것에 그치지 않는다. 기업의 경우에는 신제품을 출시하기 전에 사전 발표로 고객들의 반응을 확인할 수 있고, 정부조직의 경우에는 정책 방향에 대하여 국민들이 인

5) 정보통신망 이용촉진 및 정보보호 등에 관한 법률 (2017). 제44조의2(정보의 삭제요청 등).

터넷 상에 제공한 정보를 수집할 수 있다. 따라서 국방부는 국방정책에 대한 인터넷 여론을 확인함으로써 정책개선의 방향을 검토하는 등의 추가적 노력을 통해 국민이 진정으로 원하는 국방업무를 수행하기 위한 적절한 자료로 사용될 수 있다.

### 1.3 디지털 정체성 관리의 자동화

디지털 정체성 관리의 한계 중 큰 요소는 앞서 서술한 것과 같이, 가짜 뉴스 등에 대한 감시와 대응속도이다. 온라인 대변인실을 운영한다하더라도 사람에 의해 인터넷 문서를 검색하여 관련된 내용을 찾고 대응하는 구조에 불과하다. 따라서 실시간으로 생산되고 순식간에 퍼져나가는 가짜뉴스의 속도를 감당할 수 없다. 또한 MIT의 연구에 따르면, 가짜 뉴스의 경우 진짜 뉴스에 비해 약 20배의 전파속도를 가지고 있다(Lazer, et al., 2018). 따라서 프로그램에 의한 자동화 기술로 24시간 실시간 검색하고 내용을 분석하는 등의 방법이 그 대응방향으로 판단된다. 하지만 이것은 단순한 가짜 뉴스 등에 의한 피해를 예방하는 것에 그치지 않는다. 기업의 경우에는 신제품을 출시하기 전에 사전 발표로 고객들의 반응을 확인할 수 있고, 정부조직의 경우 정책 방향에 대하여 실시간으로 여론을 확인할 수 있다.

## II. 제안 모델

### 2.1 인터넷 문서 자동검색 및 분석 모델의 개요

디지털 정체성의 자동화된 관리를 위한 제안 모델의 핵심기능은 다음과 같이 정리할 수 있다. 첫째는 인터넷에서 사용자와 연관된 게시 문서를 실시간으로 찾는 기능이며, 둘째는 검색된 문서를 설정한 기준에 의해 분류하는 기능이고, 셋째는 검색 및 분류결과를 사용자에게 실시간으로 알려주는 기능이다. 세 가지가 핵심기능들이 어떠한 기술로 구성하고, 기능을 추가하느냐에 따라 그 기본 기술의 정확도 및 목적이 달라질 수 있다. 예를 들어, 사용자와 관련된 어떤 문서가 검색되었는지를 알기 위해서는 원하는 온라인 사이트 등을 크롤링(crawling)하고 그 내용을 자연언어처리(natural language processing) 기법을 통해 사용자와 관련된 언론 및 웹문서 등을 탐색할 수 있다. 또한 키워드의 분류를 통해 검색된 문서의 성향을 분석하고, 이 결과를 통신기능을 통해 사용자 단말기에 전송하여 실시간으로 보고를 받을 수 있다. 반면에 소명자료 작성을 위한 프로그램 측면에서는 성향을 분석한 내용에 대해서 로봇저널리즘(Robot journalism) 알고리즘을 활용하여 단시간 내 해명 자료를 작성할 수도 있다(윤인아, 2018). 실제 예로는 2017년도에 발표된 전자대변인 프로그램을 들 수 있다.

### 2.1.1 전자대변인 시스템

전자대변인 프로그램은 인터넷 오보 및 악성 댓글을 실시간으로 수집하고 대응하여 인터넷 상에서 명예훼손 등의 문제를 해결하기 위한 프로그램이다(장상훈, 2017). 따라서 온라인상에서 발생하는 관련 게시 문서를 탐색하고, 그 문서의 문장 및 키워드를 통하여 성향을 판단하여, 이를 저장하거나 상태에 따른 대응정보를 응답문서로 자동으로 게시한다. 또한 이 대응정보에 대한 피드백 정보를 사용자에게 전송하며, 대응된 내용을 다시 성향을 판단하는 근거가 되는 키워드로 업데이트하여 사용할수록 정확도를 향상시킬 수 있는 프로그램이다. 즉 탐색부, 판단부, 저장부, 응답부, 통신부, 업데이트부로 구성되며, 작동원리는 다음과 같다.

첫째, 탐색부에서 사용자가 설정한 게시 문서를 자연언어처리 기법을 통해 크롤링한다. 예를 들어 사용자가 국방부와 관련된 소식을 실시간으로 확인하기 위해서는 크롤링하는 키워드를 ‘국방부’로 설정한다. 이때 탐색이 시작되면 현 시각부터 해당하는 온라인 사이트에서 ‘국방부’라는 자연어가 포함된 모든 단어를 검색한다. 둘째, 판단부는 탐색부에서 크롤링된 문서의 웹페이지 내부의 키워드들을 분석하여 성향을 판단한다. 이 키워드들의 분석을 위해 사용자는 저장부에 성향을 판단하는 근거가 되는 키워드들을 사전에 저장해야 한다. 예를 들어 단순히 국방부에 대한 최신 여론과 긍정적, 부정적 성향의 분류일 수도 있고 또는 구체적인 다른 키워드에 대한 새로운 문서를 확인할 수도 있을 것이다. 이는 영향력을 고려하지 않은 제1가중치로 반영하고 최신 여론과 긍정 및 부정 등 성향분석의 키워드는 <표 3>과 같이 구성할 수 있다.

<표 3> 전자대변인 시스템 제1가중치 선정 예시

구분	부호	가중치	키워드 예시(단어 또는 문장)	비고
중립	+	0	규모, 예산, 조직, 병력, 위치, ~밝혔다...	단순정보
긍정		0.8	고마운, 즐거운, 도움이 되었습니다...	긍정표현
		1	항상 감사합니다, 가장 보람찼던 경험...	큰 긍정표현
부정	-	0.8	수사를 받았다, 혐의가 있다...	부정표현
		1	역량부족, 비리, 부패, 경질을 요구...	큰 부정표현

또한, 가중치에 따라 성향을 분석하기 위해서는 단순한 키워드의 분류뿐만 아니라 그 문서가 가지는 영향력도 고려되어야 한다. 왜냐하면 아무도 보지 않는 다량의 유사한 문서는 큰 의미를 가지지 못하기 때문이다. 따라서 온라인상에서의 조회수 등을 고려하여 <표 4>와 같이 제2가중치를 반영하였다. 이는 모집단의 규모에 따라 해당 문서의 노출량이 달라지고, 이에 따른 영향력의 차이가 존재하므로 구분한 것으로, 사이트의 크기나 조회수에 따른 가중치 산정은 예시로써 표현하였다.

<표 4> 전자대변인 시스템 제2가중치 선정 예시

구 분	조회수(좋아요)	제2가중치
100만 회원 사이트	100회 이하	1.1
	101~300회	1.2
	301~400회	1.3
1만 회원 사이트	50회 이하	1.1
	51~80회	1.2
	81~100회	1.3

상기 <표 3>과 <표 4>의 제1가중치와 제2가중치를 기반으로 성향분석 결과를 감성지수(E)로 나타낼 때, 제1가중치에서 긍정, 부정 키워드의 합에 의해 결정되는 부호와 가중치의 평균점수, 제2가중치에서 영향력을 고려한 값을 전체 검색된 키워드로 표현할 수 있다. 즉, 제1가중치에 적용된 양의 상수를  $a$ , 음의 상수를  $b$ , 제1가중치의 키워드의 평균값을  $A$ , 제2가중치를  $B$ , 제2가중치 키워드의 평균값을  $C$  라고 한다. 이때 긍정(중립) 키워드의 전체 개수는  $n_1$ , 부정의 키워드 개수는  $n_2$ , 전체 키워드의 총 개수는  $n_3$ 이다. 이에 따라  $E$ 는 식 (1)과 같이 표현할 수 있다.

$$E = \frac{a(A \cdot B)n_1 + b(C \cdot B)n_2}{n_3} \quad (1)$$

뿐만 아니라, 특별한 키워드는 별도의 이벤트로 관리될 수 있다. 예를 들어 ‘국방개혁 2.0’에 관한 내용이라면 소개페이지 링크를 제공하는 등의 판단도 가능하며 순환 비방이나 혐오성 내용에 대해서는 저장부로 IP 및 해당 페이지를 저장하여 필요 시 법적대응을 위한 근거자료로 활용될 수 있다.

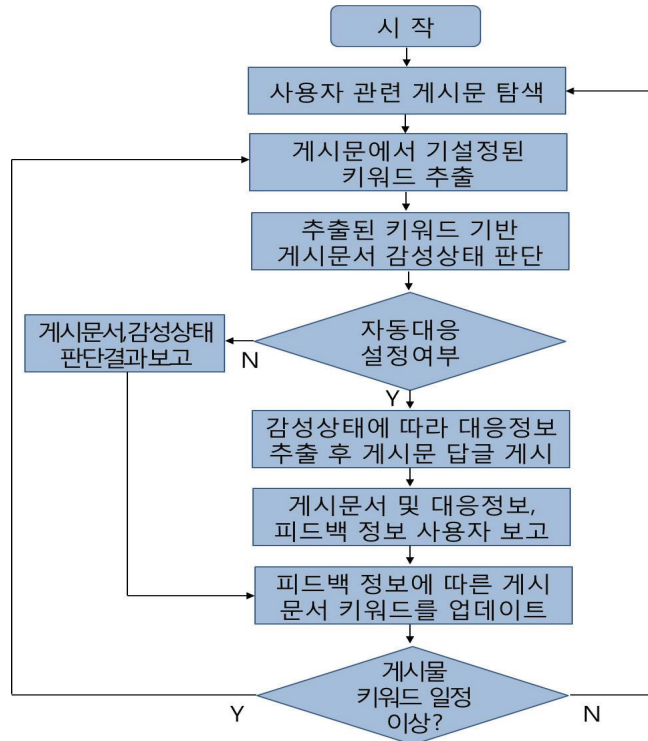
셋째, 저장부는 <표 3>과 <표 4>의 예시와 같이 키워드를 저장하고 시스템의 작동에 따라 업데이트되는 키워드를 자동으로 저장할 수 있으며 판단부의 판단에 따른 별도의 이벤트를 저장할 수 있다.

넷째, 응답부는 감성지수에 따라 분석된 문서를 사용자에게 알려주고 설정에 따라 기 저장된 대응문서를 탐색된 문서에 답글 등의 형식으로 게시할 수도 있으며 로봇저널리즘 기능을 활용하여 대응문서를 자동으로 작성하여 대응할 수도 있다.

다섯째, 통신부는 탐색된 게시문서와 감성지수 결과, 대응결과 등의 모든 사항을 사용자가 등록한 단말기에 전송하고 사용자의 피드백 정보를 다시 전송받아 이를 저장부에 저장할 수 있다.

여섯째, 업데이트부는 상기 피드백 정보를 저장부에 저장하여 사용자의 대응에 따라서 대응방법을 패턴으로 학습하게 된다. 또한 다시 탐색된 유사문서에 새로운 키워드 등을 자동으로 업데이트하여 탐색된 문서가 새로운 성향의 문서인지, 과거와 동일한 문서인지를 판단하여 대응방법을 단순히 사용자에게 내용을 알려줄 것인지, 기 저장된 문서를 게시할 것인지, 자동으로 문서를

작동할 것인지 등을 결정하는 기반 정보가 된다. 전자대변인 시스템의 동작 체계는 <그림 3>과 같다.



<그림 3> 전자대변인 시스템 동작 순서도

### 2.1.2 전자대변인 시스템 프로그램의 제약 조건

국방부의 디지털 정체성 관리를 위한 프로그램을 현실점에 실제로 운용하기 위해서는 몇 가지 갖추어야 할 조건이 있다. 첫째는, 사전에 시도하는 내용으로 전담부서 생성 및 서버 등의 관리가 불필요하다. 그 효용성이 사전연구를 통해 입증되기 전에 예산을 투입하는 것은 부적절하기 때문이다. 둘째는 무엇보다 국민정서에 반하거나 위법할지도 모르는 사항이 되어서는 안 된다. 이러한 관점에서, 전자대변인 시스템은 프로그램 상 다음과 같은 한계사항이 있다. 첫째, 제1가중치를 기반으로 감성지수를 분류할 때 자연어 처리를 기반으로 설정하기 때문에 전체 문서의 의미를 파악하는데 한계점이 발생한다. 예를 들어 ‘비리의 혐의가 있었으나 이는 사실이 아닌 것으로 밝혀졌다.’ 등의 문장에 있어서는 문서의 성향을 부정적으로 인식할 수 있지만 실제로는 긍정 또는 중립의 내용인 경우도 발생할 수 있다. 따라서 방대한 양의 기사와 문장을 저장하여 빅데이터 및 패턴분석의 기능이 필요하며, 이를 위해서는 별도의 서버관리 등의 예산이 필요하다. 둘째는 매크로 기능에 대

한 국민적 인식의 한계점이 있다. 공인된 정부기관에서 비록 공적이고 정치 중립적인 내용에 한정해서 사용할 것이라고 하더라도, 자동으로 댓글을 다는 등의 행동이 현 시점에서 좋은 방향으로 인식되기에는 어려워 보인다.

## 2.2 인터넷 문서 자동검색 및 분석 모델의 구현

전자대변인 시스템이 현시점에서 실용화되는 데의 한계점은 프로그램의 알고리즘에 있는 것이 아니라 오히려 방대한 기능에 있으므로 기능을 축소하고 프로그램의 주 사용방법을 다르게 적용하면 즉시라도 국방부의 디지털 정체성 관리를 위한 프로그램으로 활용할 수 있을 것으로 판단된다. 정부기관에서 인터넷이라는 세계에 접근하여 자동화된 댓글로 대응을 하는 모습이 불법 매크로 프로그램과 닮아 있기 때문에 이 기능을 제거하고 담당자에 의한 정식적인 대응으로 전환한다. 즉 현재 모든 것을 사람이 담당하는 온라인 대변인실의 기능에서 실시간으로 문서를 확인하고 성향을 분석해주는 기능만 더 한 형태로 개발하였다. 크롤링 기술자체가 불법은 아니기 때문에 최소한 사용하는 것을 기반으로 하였고, 업데이트 부의 삭제, 감성지수의 단순화, 빅 데이터 기능을 삭제하여 윈도우 7 이상 사양이 가능한 모든 PC에서 사용가능하도록 개발한 점에서 전자대변인 시스템과 차이점이 있다.<sup>6)</sup>

### 2.2.1 인터넷 문서 자동검색 및 분석 프로그램의 구성

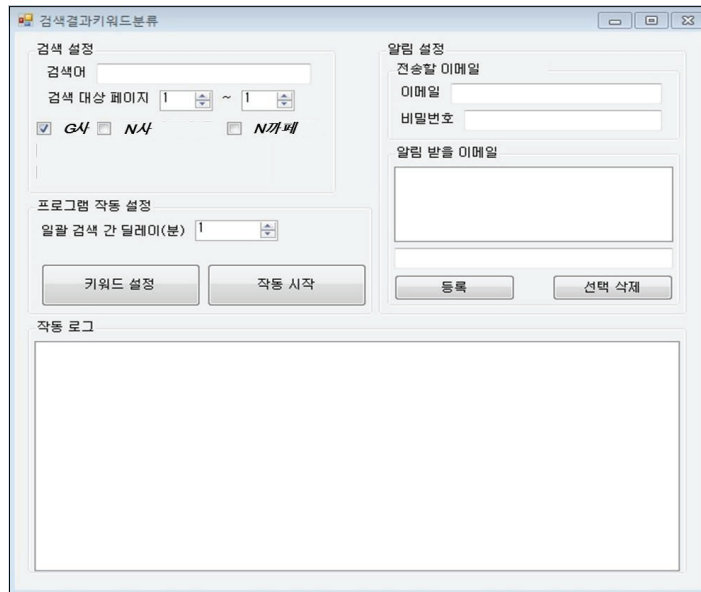
본 연구를 위해 자체적으로 개발한 이 프로그램의 구성은 탐색부, 판단부, 저장부, 통신부로 구성되며, 인터넷이 가능한 컴퓨터에서 작동할 수 있다. 탐색부에서는 사용자의 관심사를 자연어로 입력하여 관련된 모든 문서를 실시간으로 크롤링하여 판단부로 전달해주는 기능을 한다. 판단부는 전자대변인 시스템의 제1가중치 키워드 분류의 기능만 구성하여 크롤링된 문서가 어떤 주제인지, 어느 정도의 감성상태인지 부분적으로 구현이 가능하다. 저장부는 사용자가 설정한 판단부의 키워드를 저장하며, 통신부는 사용자가 등록한 이메일 주소로 검색된 문서의 링크와 및 키워드 분류결과를 알려주는 기능을 수행한다. 이를 수행할 수 있는 프로그램의 디자인은 <그림 4>과 같이 구성하였다.

<그림 4>의 좌측 상단의 검색어 칸에 사용자의 관심사를 키워드로 입력할 수 있다. 검색 웹사이트는 일반적으로 많이 사용되고 영향력 있는 사이트를 위주로 임의로 작성하였다. 그 하부 일괄 검색 간 딜레이를 통해 프로그램 작동 시간을 설정할 수 있는데 이는 제안 프로그램의 사용목적에 따라 필요하기 때문이다. 예를 들어, 실시간 검색을 목적으로 한다면 검색 간 딜레이를 최소한으로 하고 정책 등에 관한 여론조사를 위한 검사기능을 위해서는 길게 하여 한 번에 많은 분류결과를

6) 디지털데일리 (2017. 01. 18). 불법적인 크롤링, 그 대응방안은?, IT·IP법 바로알기. [http://www.ddaily.co.kr/m/m\\_article/?no=151940](http://www.ddaily.co.kr/m/m_article/?no=151940)



언을 수 있기 때문이다. 우측 상부의 전송할 이메일은 탐색 및 분류결과를 전송해주는 주소이고 알림 받을 이메일은 보고 대상의 이메일이다. 스마트폰으로 이메일을 실시간으로 받을 수 있기 때문에 통신부를 이메일로 구성하였다.



<그림 4> 인터넷 문서 자동검색 및 분석 프로그램의 디자인

### III. 실증분석

#### 3.1 국방부 디지털 정체성 관련 실시간 정보 확인

이 연구의 목적은 우선, 사용자가 원하는 검색주제 중 새로운 문서가 발견되거나, 원하는 기간 범위 내의 문서가 발견되는 경우에 자동검색을 통해 실시간으로 알려주는 것이다. 다음은 원하는 키워드의 조합을 통해 해당 검색주제에 대하여 어떠한 성향을 가진 문서가 검색되었는지 1차적으로 분류해 주어 사용자가 보다 정제된 정보를 획득하게 하는 것이다. 실증분석을 위해 로봇 배제 표준에 의한 접근 제한 페이지(robots.txt)는 제외토록 하였다. 크롤링과 관련한 판례에 따르면 상업적인 용도 및 저작권에 대한 침해에 대해서만 불법의 여지가 언급되어 있지만 사이트에서 원하지 않는 정보를 무단으로 추출하는 것은 데이터베이스 권리의 침해 여지가 있기 때문이다.<sup>7)</sup>

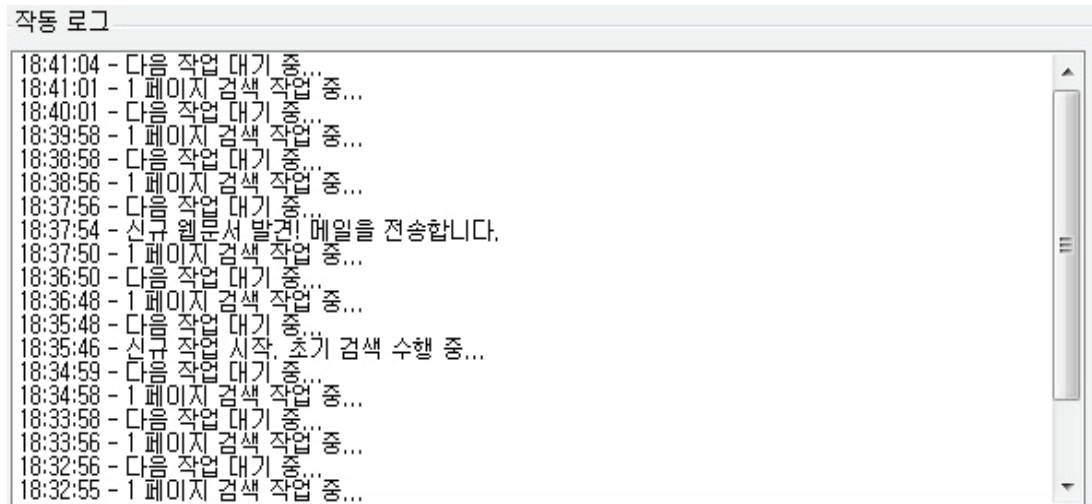
7) 법률신문 판결큐레이션 (2016. 3. 3). ‘웹사이트 무단 크롤링’ 소송 .... 잡코리아, 사람인에 승소.  
<https://www.lawtimes.co.kr/Case-Curation/view?serial=98844>

국방부 디지털 정체성 관련 실시간 정보 확인을 위한 실험은 다음과 같이 진행되었다. 첫째, 탐색 대상으로 하는 웹사이트는 소셜 미디어나 인터넷 동영상 서비스 등을 포함한 G 사의 검색엔진을 활용하였다. 둘째, <표 5>과 같이 키워드는 국방부 이하 국방부, 육군, 해군 그룹으로 설정한 후, 그 아래에 긍정 및 부정의 키워드를 입력하여 사용자에게 실시간으로 검색된 뉴스를 알려주도록 하였다. 검색 간 딜레이는 최소값인 1분으로 설정하였다. 또, 최근에 생성된 사건만 확인하면 되기 때문에 검색 대상의 웹페이지는 1장으로 설정하였다.

<표 5> 국방부 디지털 정체성 관련 실시간 정보 확인을 위한 실험의 키워드

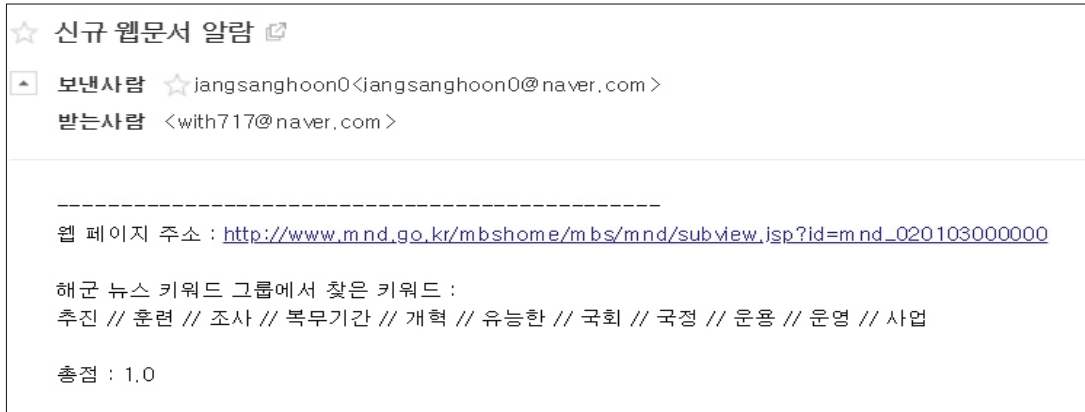
검색단어	키워드 그룹	키워드	가중치
국방부, 육군, 해군	국방부 뉴스	추진, 훈련, 조사, 철수, 합의, 복무기간, 교체,...	0
		못했다, 실패, 실수, 비리, 난타, 엇박자, 손해, 무능,...	-1
		고마운, 유능한, 합의, 협의, 완료, 의견 수렴, 우수함,...	+1
	육군, 해군 뉴스	(동일 키워드)	

실험은 2018년 10월 27일 18:30:46에 시작하였다. <그림 5>와 같이 주제인 검색단어에서 키워드 그룹과 키워드를 설정하고 1분 간격으로 검색작업을 수행하도록 작업한 결과, 신규 웹문서는 7분 8초가 경과된 18:37:54에 발견되었다.



<그림 5> 국방부 디지털 정체성 관련 실시간 정보 확인을 위한 실험 작동로그

스마트 폰으로 실시간으로 등록된 주소로 이메일을 받을 수 있었으며, 알람 내용은 <그림 5>과 같이 ‘해군 뉴스’ 그룹이고, ‘추진’, ‘훈련’, ‘유능한’... 등의 키워드로 총점이 1점인 것으로 긍정적인 내용으로 분류할 수 있었다.



<그림 6> 국방부 디지털 정체성 관련 실시간 정보 확인을 위한 실험 결과 알람

링크를 클릭하여 해당 웹페이지를 방문한 결과, 국방부 국방소식에서 해군교육사령부 전투병과 학교가 ‘2018년 전국 우수 도서관 운영평가’에서 병영도서관 부문 최우수 부대로 선정되어 국무총리 표창을 수상했다는 소식이었으며 <그림 6>과 같이 웹페이지 작성날짜는 2018년 10월 26일로 기록되어 있었다. 결론적으로 현재 프로그램의 경우 G사의 웹페이지 1장을 기준으로 원하는 검색 주제에 대하여 문서를 검색하여 분류 및 전달하는 과정에 7분 8초가 소요되었다.



<그림 7> 국방부 디지털 정체성 관련 실시간 정보 확인을 위한 실험 결과 확인내용

이는 G사의 웹페이지 1쪽을 기준으로 최신의 소식으로 확인되었다. 지속적으로 프로그램을 작동할 시에는 작성된 문서를 실시간으로 검색 및 분류하여 사용자에게 보고할 수 있는 것이 가능할 것으로 판단되어 가짜뉴스나 오보에 대해서 담당자가 즉각적으로 인지할 수 있을 것으로 기대된다.

### 3.2 국방부 정책에 관한 여론조사

국방부 정책에 관한 여론조사를 위한 실험은 실시간 정보 확인의 실험과 달리, 특정 정책에 대한 많고 다양한 의견을 분류하는 것이 더 중요하다고 판단하였다. 따라서 그 예로 ‘국방개혁 2.0’에 대하여 긍정적인 내용과 부정적인 내용을 앞서의 실험보다 더 긴 기간, 즉 웹페이지 수를 늘려 검색하고 이를 종합하여 분류하는 실험을 하였다. <표 7>와 같이 키워드를 설정하였고 여기의 가중치는 어감에 따라서 양과 음으로 분류하였다. 예를 들어 단순한 어문의 경우 긍정의 단어나 문장을 이용하여 긍정의 표현을 사용할 수 있으나, 자연어의 경우 긍정의 단어나 문장에 다시 부정의 단어나 문장을 이용하여 부정의 표현을 사용하는 경우도 있다. 즉, ‘신뢰할 수 있는’을 긍정의 표현으로 볼 때, ‘신뢰할 수 있는 군을 만든다더니...아쉬워’ 등의 부정의 문장으로 만들 수도 있다. 이 경우에 ‘신뢰할 수 있는’을 긍정의 가중치 +0.7로 두고, ‘아쉬워’를 부정의 가중치 -0.7로 산정하면 해당 문장은 부정의 표현임에도 불구하고 가중치가 0인 중립적인 문장이 된다. 그러나 해당 문장이 들어간 문서의 경우에는 표현하고자 하는 내용이 부정적인 내용으로, 긍정이나 중립의 단어 및 문장에 비하여 부정의 단어와 문장이 많을 것으로 예상된다. 또한 해당 프로그램은 해당 문서 전체의 키워드에 대한 가중치를 계산하므로, 이 문서 자체는 부정적으로 가중치를 계산할 가능성이 높다고 판단하였다.

<표 6> 국방개혁과 관련된 여론 검색 및 분석 실험의 키워드

검색단어	키워드 그룹	키워드	가중치
국방개혁	긍정적 의견	신뢰할 수 있는, 올바른, 적절한, 전시작전권 환수, 바람직한, 마땅한, 강한군대, 희소식, 다행히, 다행한,...	+0.7
		바른, 우수한, 고마운, 늘 잘하는, 믿음이 가는, 믿을 수 있는 군대, 신뢰도 향상, 정확한 청사진, 기대되는,...	+1
	일반정보	강한이지, 새로운 군대, 추진, 추진방향, 예산, 육군, 해군, 공군, 방위사업청, 병무청, 병력, 계획, 국회,...	0
	부정적 의견	난해한, 우려되는, 비현실적인, 탁상공론, 맥락이 안 보여, 아쉬운, 아쉽다, 곤란한, 뜬구름, 어려워,...	-0.7
비리 우려, 비난받아 마땅한, 멍청한, 바보 같은, 무식한, 실무를 모르는, 실패, 부족한, 역량 부족, 교체 필요,...		-1	

따라서 <표 7>은 국방개혁을 검색주제로 두고, 키워드 그룹은 긍정, 일반, 부정으로 분류한 다음, 강력한 지지나 긍정적인 표현은 +1의 가중치를, 그 반대의 경우에는 -1의 가중치를 산정하고,

그 보다 덜 적극적인 긍정이나 부정의 표현은 절대치를 0.7로, 일반적인 단어나 문장의 경우는 0으로 산정하였다. 여론조사를 위해서 검색 대상 웹페이지는 1쪽에서 30쪽까지 약 1년간 정보를 모두 검토하도록 설정하였으며, 이번에는 G사를 제외하고 N사 및 다른 웹페이지를 포함하였다. 검색 간 딜레이는 검토시간 및 실험 시간 등을 고려하여 10분으로 설정하였다. 실험은 2108년 10월 27일 20:23:27에 시작하였고, 20:29:41초인 6분 14초가 소비되었다. <그림 7>과 같이 총 135건 문서가 검출되었고 이를 총점으로 분류한 결과는 <표 8>과 같다.



<그림 8> 국방개혁과 관련된 여론 검색 및 분석 자료

<표 8>에서 긍정적 의견, 일반정보, 부정적 의견의 키워드 그룹 중 총점과 사이트는 대표적인 예시이며, 각 키워드 그룹의 소개는 그룹별 총점의 합을 의미한다. 여기서 추출된 키워드를 통하여 어떤 측면에서 검토된 긍정적 의견인지, 일반정보인지, 부정적 의견인지를 내용을 보지 않고 일부 예측이 가능하다. 또, 합계는 전체 총점의 합으로 전반적으로 인터넷에 형성된 여론의 분위기를 파악할 수 있을 것으로 보인다. 종합하면, 인터넷에서 형성된 ‘국방개혁’ 관련한 웹페이지 135개 중 11개는 긍정적으로, 27개는 중립적으로, 97개는 부정적인 여론이 형성되어 있었다. 중립적인 내용은 국방부 및 신문사에서 일반적인 소식전달의 내용이 대다수이지만, 긍정적 의견과 부정적 의견의 키워드를 분석하면 어떤 부분에서 지지를 받고, 어떤 부분에서 우려하는지 참고할 수 있을 것으로 판단하였다. 그러나 긍정적 의견은 대다수 국방부 자체에서 발표한 내용이거나 홍보 블로그가 11건 중 9건으로 이를 제외하고 종합하였다. 부정적 의견은 97건 모두 뉴스나 블로그의 내용이었으며, 이들의 키워드 순위로 주요 내용을 요약한 결과는 <표 9>와 같다.

<표 7> 국방개혁과 관련된 여론 성향분석 결과

키워드 그룹	총 점	키워드	비 고
긍정적 의견	3.4	우수한, 그나마, 정확한, 덕분에,...	<a href="https://namu.wiki/w/%EA%B5%AD%EB%B0%A9%EA%B0%9C%ED%98%81%202020">https://namu.wiki/w/%EA%B5%AD%EB%B0%A9%EA%B0%9C%ED%98%81%202020</a>
	0.7	그나마, 국방개혁, 개혁, 발표, 공군,...	<a href="http://news.kbs.co.kr/news/view.do?ncd=3425496">http://news.kbs.co.kr/news/view.do?ncd=3425496</a>
	1	정확한, 개혁, 변화, 혁신, 추진, 해군,...	<a href="http://www.archives.go.kr/next/search/listSubjectDescription.do?id=006223">http://www.archives.go.kr/next/search/listSubjectDescription.do?id=006223</a>
	1.7	우수한, 덕분에, 우려, 부족한,...	<a href="http://rigvedawiki.net/w/%EA%B5%AD%EB%B0%A9%EA%B0%9C%ED%98%81%202020">http://rigvedawiki.net/w/%EA%B5%AD%EB%B0%A9%EA%B0%9C%ED%98%81%202020</a>
소 계		11.2	11건
일반정보	0	국방개혁, 국방부,...	<a href="https://www.facebook.com/justicedefense21">https://www.facebook.com/justicedefense21</a>
	0	개혁, 혁신, 예산,...	<a href="http://www.mnd.go.kr/mbshome/mbs/mnd/subview.jsp?id=mnd_010301010000&amp;titleId=mnd_010301000000">http://www.mnd.go.kr/mbshome/mbs/mnd/subview.jsp?id=mnd_010301010000&amp;titleId=mnd_010301000000</a>
	0	국방개혁, 개혁, 발표, 추진,...	<a href="http://www.asiae.co.kr/news/view.htm?idxno=2011030814265574495">http://www.asiae.co.kr/news/view.htm?idxno=2011030814265574495</a>
일반정보	0	국방개혁, 변화, 밝혔다, 추진, 공군,...	<a href="http://news.chosun.com/site/data/html_dir/2018/08/16/2018081601329.html">http://news.chosun.com/site/data/html_dir/2018/08/16/2018081601329.html</a>
소 계		0	27건
부정적 의견	-2.4	어려운, 어려, 경질,...	<a href="http://news.chosun.com/site/data/html_dir/2018/07/27/2018072700303.html">http://news.chosun.com/site/data/html_dir/2018/07/27/2018072700303.html</a>
	-2.1	난해, 우려, 곤란,...	<a href="http://news.donga.com/3/all/20100502/28041372/1">http://news.donga.com/3/all/20100502/28041372/1</a>
	-0.4	난해, 없는,...	<a href="http://shindonga.donga.com/3/all/13/110104/1">http://shindonga.donga.com/3/all/13/110104/1</a>
	-2.4	우려되는, 우려, 어려, 부족한,...	<a href="http://news.chosun.com/site/data/html_dir/2018/08/03/2018080302005.html">http://news.chosun.com/site/data/html_dir/2018/08/03/2018080302005.html</a>
소 계		-225.4	97건
합계		-214.2	135건

<표 8> 국방개혁과 관련된 여론 성향별 내용 요약

키워드 그룹	키워드 순위	대표내용 요약
긍정적 의견 (2건)	① 적절한 (1건)	• K55 자주곡사포의 개량...독일도 PzH2000...만큼 적절한 선택일 지도...(A)
	② 긍정적인 (1건)	• 국방개혁 2.0은 많은 긍정적인 부분이 있다...3군 균형발전을 추구했던...국방부 문민화는 대단히 긍정적인 부분이다...(B)
부정적 의견 (97건)	① 우려, 우려되는, 우려의, 걱정되는, ... (77건)	• 전력증강 방향이 북한 위협 대비에만 일방적으로 치우칠 것을 우려하는 목소리도...(C)
		• 이미 하고 있는 국방비 절감 노력을 굳이 정책에 반복한 것을... 관련 예산이 추가될 수 있을지 우려의 목소리...(D)
		• 병력 숙련도 문제에 대한 우려를 이미 하고 있으면서 이를...(E)
	② 난해, 맥락없는, 어려운... (16건)	• 개념과 과제의 맥락이 안 보여..., 전체 맥락이 분명하게 잘 드러나지 않아 아쉽다...(F)
		• 지난 수십 년 동안 추진해온 방향에서 벗어난 국방개혁이 얼마나 어려운 일인지를 보여주는 사례로 보인다...(G)
③ 탁상공론, 비현실적인, ... (4건)	• 절대로 드론이나 전자감시 장비가 대체할 수 없는 부분이다... 한마디로 조선의...탁상공론의 냄새가 짙게 풍기는...(H)	
	• 비현실적으로 예산 책정이 되어있는 것이고...하지만...우리가 못 살걸 산 건 뭐가 있냐...(I)	



보다 정밀한 분석이 필요하겠지만, 순수하게 상위 키워드를 순위로 정리한 결과에서 볼 수 있듯이, 긍정적인 의견에서 나온 것은 독일의 사례와 유사하게 무기체계를 개량하는 사안(A)과 균형발전 및 문민화(B)에 관련된 내용으로 보인다. 반면에 부정적인 요소는 전력증강의 방향(C), 예산책정과 관련된 사안(D, I), 병력 숙련도(E), 현실성 및 명확한 계획 설명(F, G, H) 등으로 분류할 수 있겠다. 모든 키워드에 대해서 국방개혁에 대한 전력증강 방향, 문민화, 예산, 계획 설명 등의 콘텐츠별로 전수 조사하여 정리하면, 정책에 국민들의 의견을 더욱 반영할 수 있다. 본 연구에서는, 부정적 의견 97건 중 77건을 차지하는 ‘우려 등’ 키워드에 관련된 내용을 분야별로 정리했다.

<표 9> 국방개혁과 관련된 우려 여론의 분야별 정리내용

키워드 그룹	해당분야	대표내용 요약
우려, 우려되는, 우려의, 걱정되는,... (77건)	전력증강 방향 (37건)	· 육, 해, 공 균형 있는 발전가능성
		· 북한 외 잠재적 위협국가에 대한 대책
		· 예산획득의 가능성
	병력 감축, 전문성 감소 (22건)	· 병력 감축을 대처할 기술의 현실성
		· 복무기간의 축소에 따른 전문성 부족 대책
	비리관련 (16건)	· 첨단 무기의 개발 예산 비리방지 대책
		· 예산 운영의 투명성 관련
기 타 (2건)	· 국방개혁 과제의 모호함	
	· 국방개혁 추진 인물 관련	

<표 9>의 예시와 같이, 정책에 직접적으로 반영할 수 있을 뿐만 아니라 나아가 정책을 설명함에 있어서 어떤 부분에서 세분화하고 보충설명을 해야 할지 등을 정리하는데도 의미 있는 참고자료를 확보할 수 있을 것으로 기대된다. 이렇게 국민을 위한 군으로서 노력하는 모습 자체가 결국 국민의 지지를 받고, 신뢰받는 국방부의 디지털 정체성을 갖추게 될 것이다.

#### IV. 결론

여론은 단순한 기사가 아니라 국민의 목소리이자 생각을 뜻하며, 민주주의 사회에서 정부조직은 국민의 의견을 경청해야 한다. 또한 여론에 의해서 형성된 정부조직의 이미지는 국민의 신뢰와 지지에 영향을 미칠 수밖에 없다. 오늘날 여론은 텔레비전과 신문 등의 매체에서 인터넷으로 점차 이동하게 되었다. 이에 따라 정부조직 역시 인터넷의 문서화된 정보로 여론의 평가를 받게 된다. 이때, 그 평가된 문서의 기록들이 합쳐져서 그 정부조직의 디지털 정체성을 갖추게 된다. 그리고 국방부와 같이 대부분의 국민이 직접적인 접촉이 없는 정부조직은 실체보다는 디지털 정체성에 의해 판단하게 될 가능성이 크다.

따라서 정부조직의 디지털 정체성의 관리는 국민과의 신뢰도를 결정짓는 중요한 요소로 적절하게 관리하기 위해 다음과 같은 사안을 고려해야 한다. 첫째, 인터넷의 부작용으로 인해 이유 없이 디지털 정체성의 이미지가 훼손되지 않도록 관리하여야 한다. 인터넷이란 쉽게 의견을 남길 수도 있지만, 익명성과 전파성을 활용하여 거짓된 정보, 이른바 가짜뉴스를 생성하고 확산하는 것도 쉽다. 이 가짜뉴스로 인해서 이미지가 훼손되지 않기 위해 가장 중요한 것은 빠르게 인지하고 대응하는 것이다. 둘째, 정책을 수행함에 있어서 인터넷 여론의 의견을 확인할 필요성이 있다. 우직하게 밀고 나가는 것만이 장점이 아니라, 국민들의 의견을 중립적인 자세에서 듣고 유연하게 필요한 부분의 변화를 모색하는 부지런한 개선의 자세가 필요하다. 그리고 이러한 활동은 소통하고 투명한 정책의 시행으로 국방부의 디지털 정체성이 향상될 수 있다고 판단된다. 구체적으로 정책에 대한 인터넷 여론을 넓게 검색하여 분야별로 정리하여 격려와 응원이 많은 부분은 더 장려하고, 비판과 우려가 포함된 내용은 자세한 검토를 토대로 답변을 세분화하는 이원화 접근이 중요하다.

제안한 모델은 상기 두 가지 사안에 대해서 모두 응용이 가능한 것으로 사료된다. 한편, 자연어를 처리하는 과정에서의 문서의 복잡한 입장을 정확히 분류하거나 키워드를 선정하는 것은 어느 정도 숙련도가 필요할 것으로 판단된다. 또한, 적극적으로 개발 시 별도의 대형 서버를 구축하여 빅데이터 및 패턴분석 등의 기능을 추가하는 것이 그 정확도를 향상하는 방법으로 판단되며, 특정 사이트들을 사용할 경우 해당 회사와 분야에 대한 계약을 통해 국민의 사생활 및 데이터베이스 권리를 침해하지 않고 오로지 공익을 위한 방향으로 사용할 수 있도록 법적 절차를 마련하는 것도 연구되어야 할 것이다.

## V. 시사점

본 연구와 관련하여 궁극적으로 이루고자 한 것은 국방부의 사용자가 관심을 가진 주제에 대하여 실시간으로 인터넷 문서를 검색하여 그 성향 분석을 마친 후 사용자에게 결과를 제공하고, 필요한 경우 사용자가 적절한 대응을 할 수 있도록 돕는 것이었다. 첫째, 군에 대한 국민의 인식은 정신 전력에 직접적인 영향을 미치기 때문에 부정적인 여론 등이 인터넷 언론을 통해 빠른 속도로 번지는 것을 인지하고, 이것이 사실인지, 어떠한 오해가 있는지 등의 대응이 필요하다(김태성, 2017). 둘째, 군의 방향을 모색할 때, 징병제 국가로 군에 대한 일반 국민의 관심이 높아 인터넷 여론을 통해 많은 의견을 표출되고 있어 이를 자동적으로 분류하여 정책의 방향을 검토할 필요성에 대응할 수 있다. 그래서 본 연구는 새롭게 생성된 문서를 즉각적으로 분류하고 제공하는 실험과 비교적 긴 기간 동안 형성된 문서들을 분류하는 실험을 진행하여 전자대변인 프로그램의 실무적 기여점을 제시하였다. 현재 인터넷 여론을 포함하여 국방업무와 관련한 언론내용을 확인하고 대응하는 주체는 국방부 및 각 군, 방위사업청 등 각 기관 대변인실의 제한된 인원들이다. 특히, 방위사업청 대변인실

에서는 각 주요 요직의 인물들을 대상으로 일일 2회 방위사업청과 관련한 언론 모니터링 결과를 문자로 알려준다. 그러나 본 연구를 위해 실험적으로 개발한 현 프로그램의 경우 6~7분에 원하는 주제에 대한 신규문서를 지속적으로 검색하여 사용자에게 제공하며, 각 사용자 별로 맞춤형 주제를 선정할 수 있는 특징을 갖고 있다. 무엇보다 인터넷 언론은 초기 대응에 따라 그 확산방향과 속도가 달라질 수 있어 현행 대변인실 인력이 담당하는 모니터일 시스템 대비 효과를 개선하는 기대효과가 있다고 판단된다. 한편, 국방정책에 대한 폭넓은 분석과 분류에 대한 수작업이 향후에 지속될 필요성이 존재한다. 현재의 프로그램은 이는 아직 초기형태 그 한계가 분명하고, 국방분야에서 주로 사용하는 긍정과 부정의 단어 및 문장들을 면밀하게 분석하여 감정 사전을 작성하여, 보다 사용자가 원하는 성향을 정확하고 구체적으로 분석할 수 있는 미래 연구가 필요하다.

참고문헌

- 김태성 (2017). 군에 대한 국민의 인식 분석을 통한 정신전력 강화 방안. *정신전력연구*, 49, 167-216.
- 윤인아 (2018). 로봇저널리즘의 이해와 전망, 제4차 산업혁명과 소프트파워 이슈리포트, 2018-18, 정보통신산업진흥원, 1-14.
- 이종하 (2017). 국방 빅데이터 시범서비스 추진사례. 대한산업공학회 춘계공동학술대회 논문집, 1428-1447.
- 장상훈 (2017). 전자대변인 시스템(AN ELECTRONIC SPOKESMAN SYSTEM), 특허 제 10-1812933호(2017.12.05.).
- 정민, 백다미 (2017). 가짜 뉴스(Fake News)의 경제적 비용 추정과 시사점. *한국경제주평*, 17-11(736), 1-15.
- Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Schudson, M. (2018). The science of fake news. *Science*, 359(6380), 1094-1096. doi: 10.1126/science.aao2998

원 고 접 수 일 2019년 12월 06일  
원 고 수 정 일 2019년 12월 27일  
게 재 확 정 일 2019년 12월 30일

# Study on the Invention of Real-time Internet Document Automatic Search & Analysis System for MND's Digital Identity

Sang Hoon Jang\* · Suk Jun Lee\*\*

As Internet media develops, not only individuals but governments have digital identities (digital identity includes images made by digital media), in addition, as Internet media develops, chances are high that images will be determined by digital identity. Therefore, companies are meeting customer needs through digital image management and data analysis. This is also necessary for government organizations like the Ministry of National Defence (MND). At least we can prevent the damage caused by fake news and can see the Internet's response to defense policy. Therefore, we invented an automatic search and analysis program for real-time internet documents. Also, through experiments with the policy of MND, we studied the significance and possibility of the project.

**Keywords** : digital identity, automatic search, real time, Internet press, defense policy, spokesman

---

\* Kwangwoon University, Defense Acquisition Program, Ph.D. Candidate (Director for Programs Audit and Inspection, Defense Acquisition Program Administration)

\*\* Kwangwoon University, College of Business, Professor

