

웹 미디어 기반 실시간 국방 이슈 탐지 및 그룹핑

오효정

전북대학교 대학원 기록관리학과
문화융복합아카이빙 연구소

Realtime National Defense Issue Detection and Grouping based on Web Media

Hyo-Jung Oh

Graduate School of Archives and Records Management
Institute of Culture Convergence Archiving
Chonbuk National University

요 약 디지털 공간에서 개인과 조직의 활동기록이 축적되면서 활용할 수 있는 정보의 양도 폭발적으로 증가하고 있는 환경에 있어 사용자에게 가장 절실한 요구사항은 ‘유용한’ 정보의 ‘효율적인’ 획득에 있다. 본 논문에서는 실시간 이슈 탐지 기술을 활용하여 국방 정보 서비스로의 접목, 디지털 융복합을 시도한다. 국방 분야와 관련된 웹 상에 공개된 대량의 콘텐츠와 국방 보고서 및 DB를 연계하여 무기체계, 정책 및 훈련, 기관 및 조직 등에 대한 국방 이슈를 실시간으로 탐지하는 특화 기술을 개발하고자 한다. 이를 위해 실시간 국방 이슈 탐지 및 그룹핑 기술에 대한 요구사항을 수렴하고 분석하여 핵심 기능을 도출하고자 한다. 또한 ‘국방’ 분야에 특화된 언어자원을 보강하고 이슈 성 지수 자질을 발굴함으로써 도메인 특화 이슈 탐지 방법의 차별화를 꾀한다. 더불어 실제 수집된 웹 뉴미디어에 기반한 국방 분야 이슈 탐지 결과와 기존 이슈 탐지 결과와 비교함으로써 그 차이를 규명하고자 한다.

주제어 : 이슈 탐지, 국방 정보 분석, 실시간 탐지, 웹 미디어, 디지털 융복합

Abstract Because mass of activity records of individuals and organizations are accumulated in the digital space and amount of information is also increasing exponentially, the most urgent requirements of users is the tool for ‘efficient’ acquisition of ‘useful’ information. This paper try digital convergence to combine a domain specific technology with real time issue detection and grouping based on Web media. To derive core functionalities, we collect and analyze user requirements of national defense issue detection services. By utilizing linguistic resources specialized on national defence area and discovering features for measuring issue importance, we try to seek differentiation domain specific issue detection method. Finally we compare our detection results based on the development outputs of prototype.

Key Words : Issue Detection, National Defence Information Analysis, Realtime Detection, Web Media, Digital Convergence

* 이 논문은 2015년도 전북대학교 연구기반 조성비 지원에 의하여 연구되었음

Received 1 June 2015, Revised 7 July 2015

Accepted 20 August 2015

Corresponding Author: Hyo-Jung Oh (Graduate School of Archives and Records Management, Chonbuk National University)

Email: ohj@jbnu.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

모바일 디바이스의 보편화와 클라우드 기술의 비약적인 발전으로 인해 디지털 공간에서 개인과 조직의 활동 기록이 축적되면서 활용할 수 있는 정보의 양도 폭발적으로 증가하고 있다. 이러한 컴퓨팅 환경 변화에 있어서 사용자에게 가장 절실한 요구사항은 ‘유용한’ 정보의 ‘효율적인’ 획득에 있다. 특히 최근에는 실시간으로 생성되는 엄청난 양의 콘텐츠와 기존에 구축되어 있던 다양한 형태의 데이터 정보를 대상으로 한 정보 분석 및 통합에 대한 요구가 급증하고 있는 실정이다.

‘빅데이터’라는 광활한 정보의 바다에서 사용자가 원하는 정보를 실시간으로 추출하기 위한 연구는 국내외에서 매우 활발히 진행되고 있다. 해외에서는 블로그가 활성화된 2000년도 이후부터 소셜미디어 등과 같이 웹 상에 공개되어 있는 대량의 콘텐츠로부터 사용자의 니즈(needs)를 파악하기 위한 연구가 진행되어 왔다[1]. 특히 비즈니스 인텔리전스(Business Intelligence: BI) 분야에서는 특정 브랜드의 모니터링, 사용자의 피드백 감성분석 및 네트워크 분석 기반 구전 마케팅 및 의사결정을 지원하는 제품들이 개발되었다[2]. 나아가 각국에서는 다양한 웹 콘텐츠로부터 질병과 재해 관련 정보를 탐지하거나, 글로벌 동향을 분석하고 모니터링하는 서비스를 제공하는 등 공익을 목적으로 한 연구도 활발히 진행 중이다[3].

실시간 정보 분석 분야에서 가장 선두를 달리고 있는 연구진은 역시 구글(Google)로, 전 세계에서 매일 수집되는 대량의 웹 사용자의 검색 로그를 기반으로 한 동향 분석 서비스[4]를 제공하고 있다. 뿐만 아니라 사용자의 관심 주제의 시간상의 변화 관찰, 지역별 사용자 관심 추이 관찰이 가능하다. 나아가 구글에 입력된 사용자 검색어가 특정 산업의 현 경제 활동 상황과 상관관계가 있다고 가정, 다양한 예측 모델 제시하는 서비스도 제공하고 있다[5].

이러한 추세에 발맞춰, 국내에서는 2012년 총선을 기점으로 기존 전통적인 여론 조사 방식에서 탈피, 웹 공개 콘텐츠로부터 실시간 여론을 수렴하고자 하는 서비스가 출시되고 있다. 특히 기존의 한국어 검색서비스 업체들이 언어분석 기술을 접목하여 대용량 소셜미디어 분석으로 방향을 잡고 한국어 분석 기반 서비스를 출시하고 있다. 국내에서 언어분석 기반으로 소셜미디어 분석 기능

을 제공하고 있는 대표적인 서비스들을 소개하면 다음과 같다[1].

소셜메트릭스[6]는 국내 텍스트마이닝 전문 업체인 다음소프트가 제공하는 서비스로 자연어처리 기술과 텍스트마이닝 기술을 바탕으로 블로그, 트위터 문서를 분석한 모니터링 정보를 제공한다. 기업용 검색 솔루션 개발 업체인 솔트룩스는 정치인 혹은 정치적 관심을 받는 인물들을 대상으로 소셜미디어 상에서의 추이를 제공하는 트루스토리 서비스를 제공한다[7]. 한국전자통신연구원(이하 ETRI)에서 개발한 소셜미디어 이슈 탐지 및 모니터링 플랫폼 WISDOM[8]은 소셜미디어(뉴스, 블로그, 트위터) 문서를 수집/저장하고, 차별화된 심층 언어분석을 기반으로 추출한 정보를 이용하여 이슈를 탐지하고 모니터링하는 기능을 제공한다.

또 다른 한편으로는 실시간 이슈 탐지 기술을 정치, 경제, 안보, 복지 분야 등 사회 각 분야에 응용하여 활용하고자 하는 시도들이 각국에서 진행되고 있다. 2009년 유엔에서는 인간의 복지와 관련된 다양한 변화와 영향을 정확하게 이해하고 위기에 대응하기 위해 데이터 분석을 토대로 한 전략을 수립하기 위해 UN Global Pulse라는 기관을 창설하였다[9]. 실시간 데이터 스트림 분석을 통해 사회 및 개인 안전 관련 주요 이슈를 도출하거나 8년간의 뉴스 정보를 분석하여 식품 안정성 관련 이머징 이슈를 도출하는 연구를 진행하고 있다. 그밖에도 영국, 싱가포르 등 해외 주요국들은 향후 국가 및 사회의 위협요소에 대한 효과적인 선제적인 탐색을 위해 이슈 스캐닝(Horizon Scanning) 기법을 활용하고 있다[10].

우리나라에서도 공공정보의 적극적인 개방을 통해 국민 맞춤형 서비스를 제공하기 위한 정부 운영 패러다임의 일환으로 정부 3.0이 출범하였다[11]. 이를 통해 공중보건 분야에서는 전국의 의료 데이터를 연계하여 전염병 발생과 같은 긴박한 순간에 빠른 의사결정을 지원하거나 [12], 교통 분야에서는 빅데이터에 기반한 교통 흐름 모델링과 예측, 최적화 시스템 구현, 교통 신호 체계와 유지 보수 정책에 반영하고자 하는 시도들이 일어나고 있다.

특히 대외적으로 ‘국방’ 분야에 대한 빅데이터 기반 이슈 분석에 관한 대규모 프로젝트가 전략적으로 진행되고 있다. 미국의 경우, 방위고등연구계획국(DARPA)를 비롯한 6개 연방 부처 및 기관이 ‘빅데이터 R&D 이니셔티브’ 프로젝트에 참여, 빅데이터를 활용해 인지, 지각, 결

정을 제공하는 자율시스템을 구축하여 스스로 결정하고 군사 행동을 수행하도록 추진하고 있으며, 싱가포르 역시 텍스트 데이터에 기반하여 국가안전을 위협하는 요소에 대한 평가와 주변 환경변화를 탐지하여 새로운 기회를 발굴하는 RAHS 프로그램을 2004년부터 운영하고 있다[13].

본 논문에서는 실시간 이슈 탐지 기술을 활용하여 국방 분야의 뉴스/블로그/트위터 등 웹 상에 공개된 대량의 콘텐츠와 국방 관련 보고서 및 DB를 연계하여 무기체계, 정책 및 훈련, 기관 및 조직 등에 대한 국방 이슈를 실시간으로 탐지하는 특화 기술을 개발하고자 한다. 이를 위해 실시간 국방 이슈 탐지 및 그룹핑 기술에 대한 요구사항을 수립, 분석하여 핵심 기능을 도출하고자 한다. 또한 ‘국방’ 분야에 특화된 언어자원을 보강하고 이슈성 지수 자질을 발굴함으로써 도메인 특화 이슈 탐지 방법의 차별화를 꾀한다. 더불어 실제 수집된 웹 뉴스미디어에 기반한 국방 분야 이슈 탐지 결과와 기존 이슈 탐지 결과와 비교함으로써 그 차이를 규명하고자 한다.

2. 국방 분야 이슈 탐지 요구사항 분석

실시간 국방 이슈 탐지 및 그룹핑 기술은 웹 콘텐츠 내에서의 정보를 기반으로 국방 관련 이슈를 탐지하고 이슈의 시간에 따른 전개 과정을 모니터링함으로써 국방 기술 분야에서의 데이터 기반(data-driven) 의사 결정 지원 체계를 확립하기 위한 기반 연구이다. 본 장에서는 이를 위한 세부 기능 요구사항을 살펴보도록 한다.

실시간 국방 이슈 탐지 및 그룹핑 시스템은 먼저 1) 대량의 웹 미디어로부터 ‘국방’관련 콘텐츠를 수집 및 선별하는 기능과 2) 수집된 콘텐츠로부터 고정밀 언어분석을 수행하고, 3) 해당 날짜의 주요 이슈를 파악, 관련있는 이슈들끼리 묶어 그룹핑하고 4) 마지막으로 시계열에 기반하여 해당 이슈들이 어떻게 전개되었는지 제시하는 기능이 필수적이다.

세부적인 요구사항을 살펴보면, 수집 대상이 되는 웹 미디어는 국내외 국방 분야 주요 일간지, 온라인 잡지, 홈페이지, 블로그, 트윗 등으로 사용기관 운영자가 관심 사이트에 대한 정보를 등록 및 관리할 수 있는 기능이 필요하다. 특히 국방망 내에서 관심 홈페이지, 인가된 정보서

비스, 국방기술정보통합관리 체계(DTiMS, [14]) 내의 사이트, 전자도서관 등 정보가 수집되어야 한다. 수집된 대량의 웹 미디어로부터 양질의 국방 관련 콘텐츠를 선별하기 위해서는 콘텐츠의 신뢰도 및 정보성에 따라 필터링하는 기능이 필요하다. 또한 정보의 유형에 따른 관심 정보 필터링 기능이 요구된다.

수집된 콘텐츠 내에서 주제를 파악하여 핵심 이슈를 파악하기 위해서는 고정밀 언어분석 기술이 필수적이다. 특히 국방 분야 내에서만 사용되는 무기 이름이나 조직 체계, 군사 작전 명령어 등 해당 분야 내의 전문용어에 대한 이해가 선제되어야 한다.

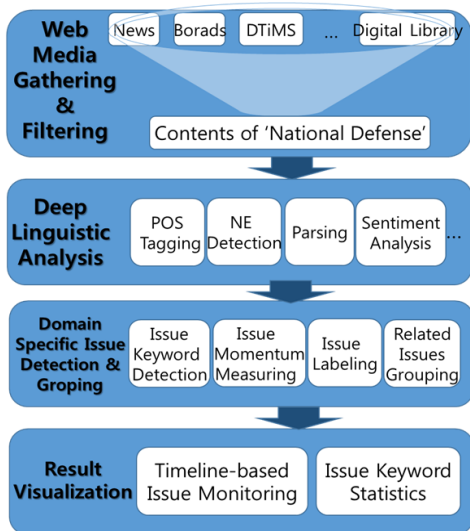
해당 날짜별로 주요 이슈를 파악, 같은 주제를 같은 이슈들끼리 묶어 그룹핑하는 기능은 대량의 텍스트로부터 언어분석을 통해 인식된 다양한 주제어 중 사용자가 관심이 있는 특정 정보를 추출하는 기술을 필요로 한다. 예를 들면 국내외 방산 거래, 가격, 예산, 신기술 개발, 신무기체계 개발, 전력화 정보 등 추출하거나 국방 키워드 관련 트윗, 블로그, 뉴스에서 이슈 추출 및 출처별 연관성 파악하는 방법등이 그것이다.

시계열에 기반하여 해당 이슈들이 어떻게 전개되었는지 제시하는 기능을 구현하기 위해서는 추출된 이슈들을 분야별, 주제별, 출처별 등 다양한 관점에 따라 저장하고 관리하는 기술이 수반된다. 또한 국방 및 민간 분야에서 정의된 온톨로지 정보 관리 기능과 사용자의 질의에 해당하는 키워드에 대한 이슈를 탐지하는 기능이 제공되어야 한다[14]. 특히 외국어 검색시 한국어 키워드를 입력하면 국방과학기술용어사전의 대역어를 찾아 검색하는 기능이 필요하며 대역어가 다수인 경우에는 후보어를 제시하여 사용자가 선택 후 검색 수행을 수행하는 편의가 제공되어야 한다. 또한 사용자 질의에 대하여 시간대별로 정보를 정렬하는 기능과 시대의 흐름에 맞게 그래픽 등을 통하여 가시화 하는 시각화 기능에 대한 요구가 높았다.

그 밖의 요구사항으로는 사용자 조회 빈도, 자료의 빈도 및 양 등을 종합해 이머징 핫키워드 추출하거나 일일, 주간, 월간, 분기, 반기, 연간을 이전 기간과 대비하여 통계 현황을 제시하는 등의 통계정보 제공 기능과 주간, 월간, 분기, 반기, 연간으로 축약해서 제시하거나 표 혹은 카렌다(calendar) 형태 등의 다양한 사용자 인터페이스가 필요한 것으로 분석되었다.

3. 실시간 국방 이슈 탐지 및 그룹핑 시스템

이번 장에서는 앞 장에서 분석된 요구사항을 기반으로 웹 미디어 기반 실시간 국방 이슈 탐지 및 그룹핑 시스템을 구축하기 위한 세부 구성 요소에 대해 기술하기로 한다. [Fig. 1]은 본 논문에서 제안하는 실시간 국방 이슈 탐지 및 그룹핑 시스템의 구성도이다. 2012년 6월 ~ 2012년 12월에 수집된 웹 미디어를 대상으로 실제 구축된 시범서비스 결과를 통해 제안 시스템의 특징을 설명하고자 한다.



[Fig. 1] System Overview of Realtime Issue Detection and Grouping

3.1 ‘국방’ 분야 웹 미디어 수집 및 선별

본 논문에서 수집대상으로 삼은 웹 미디어는 국방일보 및 조선일보, 연합뉴스 등을 포함한 10대 언론사를 대상으로 뉴스 기사를 수집하였다. 문서 수집 및 저장 관련은 ETRI에서 개발한 대용량 문서 수집기 및 저장구조[8]를 활용하였다.

수집된 기사 중 ‘국방’, ‘안보’ 등과 같이 섹션 정보가 명시된 경우와 주제 분류기를 통해 ‘국방’ 주제로 인식된 경우만을 대상으로 선별하였다. <Table 1>은 2012년 6월 한달 동안 수집된 전체 뉴스 건수와 ‘국방’ 분야로 선별된 건수를 비교한 표이다. 전체적으로 10대 언론사 대

상 전체 신문기사 중 ‘국방’관련 기사는 1/50 수준으로, 일일 평균 3,500여건 중 약 70 여건이 수집되었다. 반면 국방관련 사건이나 정치적 이슈가 많았던 날은 1/30 수준으로 수집되는 것을 알 수 있었는데, 2012년 6월의 경우, 6월 18일은 “연평해전 10주년”이 되는 날 이였고, 19일부터 21일 사이에는 이명박 대통령이 “G20 회의로 인한 브라질 방문”한 이벤트가 있었던 관계로 평소보다 2~3배 많은 기사가 수집되었다. 특히 6월 25일을 기점으로 “6.25 기념 장갑차 훈련”, “한일군사협정”과 관련된 이슈들이 회자된 것으로 분석되었다.

<Table 1> Web Media Collection in June, 2012

Date	# of News of All domains	# of News of 'Defense' domain
20120601	3,691	67
20120602	2,021	30
20120603	2,312	62
20120604	3,201	67
20120605	3,171	61
20120606	2,072	50
20120607	3,298	58
20120608	3,150	53
20120609	1,941	23
20120610	2,019	38
20120611	3,468	60
20120612	3,257	49
20120613	3,675	52
20120614	2,501	65
20120615	4,722	73
20120616	2,133	20
20120617	2,965	93
20120618	4,964	149
20120619	5,362	147
20120620	5,274	141
20120621	5,298	125
20120622	4,561	61
20120623	2,008	33
20120624	3,069	93
20120625	4,746	141
20120626	5,153	85
20120627	5,234	79
20120628	4,910	81
20120629	4,819	35
20120630	1,984	38
Total	106,979	2,129
Average	3,565.97	70.96

3.2 ‘국방’ 분야 고정밀 언어분석

한국어 텍스트에 나타난 다양한 어휘를 분석하여 주

제어를 추출, 이슈어를 선별하기 위해서는 고정밀 언어 분석 기술이 필요하다. 본 논문에서는 ETRI에서 개발한 언어분석기[14]를 기반으로 국방기술품질원에서 제공한 ‘국방’ 분야 특화 언어자원을 보강하여 활용하였다. 국방 온톨로지[13]는 120만 어휘에 해당하는 국방용어를 포함하고 있으며, 국방 분야의 실체에 대한 속성, 사건, 프로세스, 관계들을 분류하여 공통적이고 공유가능한 정보로 구축하여 정보의 통합과 새로운 지식의 발견을 촉진하는 도구로 활용되고 있다. 특히 기존 ETRI 개체명 태그[15]에 ‘국방’분야 전문용어 사전을 매핑하는 방식으로 전문용어 사전을 보강하였는데, 대상이 되는 개체명은 주로 군사장비, 군용 운송장비, 조직, 직위, 군사조직, 훈련, 정책/전략/작전, 군수품 등이었으며 ‘국방’ 분야 인물, 정책/훈련, 무기, 기관/조직을 세부 분류하여 매핑하였다. <Table 2>는 ‘국방’ 분야 용어 리스트의 예이다.

<Table 2> Examples of Domain Specific Terms

Category	Terms	Named Entity Tag
Weapon/ Munitions	K11 Rifle	AF_WEAPON
	Objective Individual Combat Weapon	AF_WEAPON
	OICW	AF_WEAPON
	LRF	TML_HW
	K-1	AF_TRANSPORT
	tank	AF_TRANSPORT
	BlackPanther	AF_TRANSPORT
Military operations	ACTD Project	TML_PROJECT
	Eulji Train	EV_OTHERS
	Aegis combat system training	EV_OTHERS
	ACMR	EV_OTHERS
	air combat maneuvering range	EV_OTHERS
Military organization	National Defense Committee	OGG_POLITICS
	Armed Forces Nursing Academy	OGG_EDUCATION
	Transportation Movement Office	OGG_MILITARY
	POG	OGG_MILITARY
	psychological operations group	OGG_MILITARY

3.3 실시간 국방 이슈 탐지 및 그룹핑

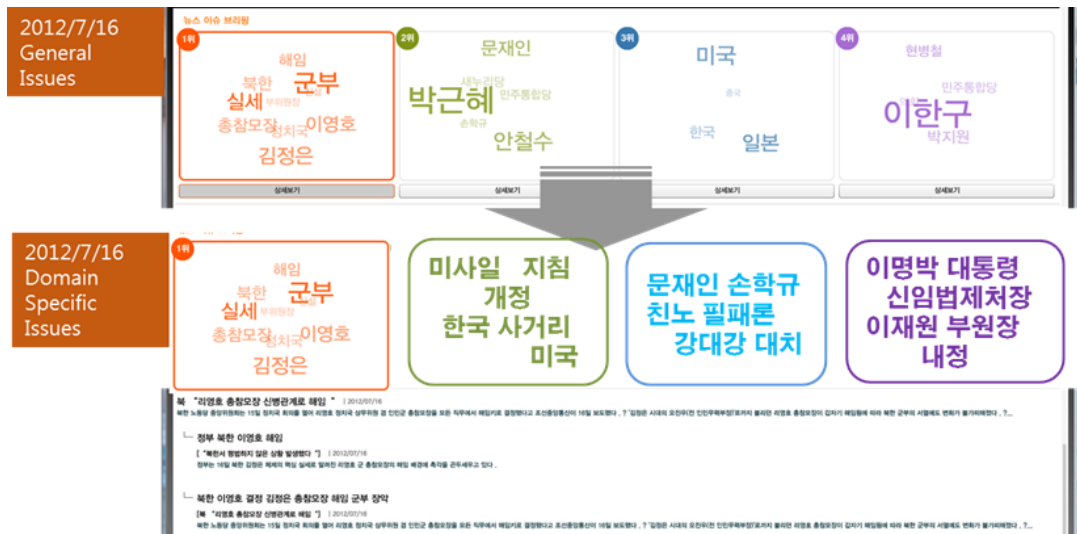
기존 웹 콘텐츠 분석서비스들에서는 대부분 빈도를 기반으로 이슈어를 선정하고 있다[6,7]. 따라서 항상 많이 나타나는 단어나 계절적, 이벤트 성으로 정기적으로 출

현 빈도가 높아지는 단어에 대해서는 진정한 이슈성을 계산하기 힘들다. ETRI에서 개발한 WISDOM 서비스에서는 이러한 취약점을 보완하는 이슈 모델링을 제시하고, 이를 바탕으로 빈도와 별도로 이슈성을 반영한 이슈 탐지 방법론을 적용했다[8]. WISDOM에서는 이슈 대상어들의 신규성, 중요성, 과급력, 관심도 및 미디어별 신뢰도 등을 통합하고 수치화하여 이를 이슈성 지수로 사용한다. 본 논문에서는 이와 더불어 ‘국방’ 분야에 특화된 이슈성 지수로 콘텐츠의 ‘국방’ 분야 적합도와 추출된 용어의 국방 용어 가중치를 반영하여 이슈성 지수를 계산하도록 개선하였다.

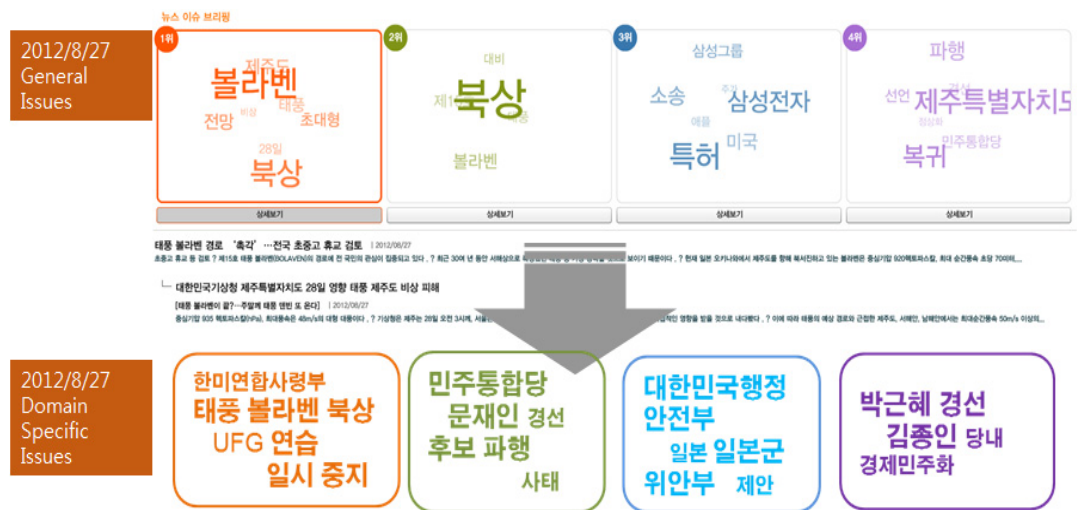
[Fig. 2]의 상단은 2012년 7월 16일 수집된 뉴스 전체를 대상으로 추출된 이슈 탐지결과를 나타낸다. 해당 날짜를 대표하는 이슈로는 “북한의 ‘리영호’ 총참모장 해임” 사건이 1등으로, “박근혜, 문재인 대선 지지율” 관련 이슈, “한국과 미국, 일본 환태평양 경제협정 가입” 관련 주제 등이 추출되었다. 해당 박스는 추출된 이슈어를 기준으로 이슈 중요도에 따라 시각화 한 것으로, 같은 주제를 갖는 뉴스 기사들이 하단에서와 같이 그룹핑되어 제공된다. 사용자는 이를 통해 유사한 주제를 갖는 기사를 묶어서 살펴봄으로써 언론사별 관점에 따른 차이와 향후 기사의 전개 방향을 파악하기 용이하다.

[Fig. 2]의 하단은 같은 날짜에 수집된 뉴스를 대상으로 ‘국방’ 분야에 특정하여 이슈를 추출한 결과이다. “북한의 ‘리영호’ 총참모장 해임” 이슈는 7월 16일 전체 도메인을 대상으로 한 결과와 마찬가지로 1등 이슈로 추출되었으나 2등 이슈로는 “한미 ‘미사일 지침’ 효력”과 관련된 이슈가 추출되었고, 4등 이슈로는 “신입법제처장 내정” 등과 같이 청와대의 행보와 관련된 이슈가 탐지되었다. 이는 전체 뉴스 중에 국가 안보, 방위 및 국정 운영과 관련된 이슈를 자동으로 탐지, 중점적으로 관리할 수 있음을 의미한다.

[Fig. 3]는 2012년 8월 27일 “태풍 불라벤”이 발생한 날로 전반적인 주요 이슈는 대부분 태풍의 이동 경로, 피해 상황에 관련된 기사인 반면, ‘국방’ 분야 이슈 탐지 결과에서는 “태풍 불라벤”이 군사 작전에 어떤 영향을 끼쳤는지를 자동으로 분석했다. 이는 도메인에 따라 이슈성을 판단하는 지표를 달리함으로써 도메인 특화 이슈 추출의 정확도와 유용성이 향상될 수 있음을 의미한다.



[Fig. 2] Comparison with General Issues and Domain Specific Issues on July 16, 2012



[Fig. 3] Comparison with General Issues and Domain Specific Issues on August 27, 2012

3.4 국방 이슈 키워드 통계

실시간 국방 이슈 탐지 및 그룹핑 결과는 기본적으로 일단위로 제공된다. 특히 지정된 날짜에 가장 뜨거웠던 이슈 키워드들의 분야별(인물, 정책/훈련, 무기/군수품, 조직) 순위를 20위까지 제공한다. 핫키워드 추이를 보면, 분야별로 날짜별 이슈 키워드를 알 수 있고, 감성분석 결과와 전달 대비 등하락 정보도 같이 제공하므로 전반적

인 여론을 한 눈에 알 수 있다. 각 화면의 오른쪽 스크롤을 이용하면 1위에서 20위까지의 순위가 확인 가능하다. <Table 3>는 2012년 6월동안 이슈가 되었던 키워드들의 분야별 순위와 해당 이슈에 대한 뉴스 여론의 감성분석 결과이다. 해당 기간동안 가장 회자된 정책으로는 “4대강 정비사업”으로 대부분의 뉴스(약 88%)가 부정적인 여론을 보임을 알 수 있다.

<Table 3> Issue Keyword Statistics on June, 2012

Category	Rank	Keyword	Positive	Negative	+/-
Person	1	Chul-soo An	21.30%	78.70%	0
	2	Guen-hye Park	39.30%	60.70%	1
	3	Ji-won Park	4.10%	95.90%	48
	4	Jae-in Moon	28.00%	72.00%	0
	5	No-hyun Kwak	37.20%	62.80%	-3
Public Policy & Agreements	1	Four Major Rivers Project	11.80%	88.20%	8
	2	Gift Tax	0.00%	100.00%	3
	3	Acquisition Tax	75.80%	24.20%	4
	4	responsible prime minister	50.00%	50.00%	33
	5	Korea-U.S FTA	33.30%	66.70%	11
Weapon/ Munitions	1	MCRC	21.80%	78.20%	6
	2	K5	94.70%	5.30%	1
	3	Satellite Communications system (SCS)	41.30%	58.70%	121
	4	GRR-5K	27.70%	72.30%	340
	5	Surveillance Radar (SR)	82.00%	18.00%	44

4. 활용 방안 및 시사점

인터넷의 영향력이 증가됨에 따라 다양한 이슈분석 및 예측분석 플랫폼을 웹과 결합하여 개인, 기업 뿐 아니라 국가 차원에서의 활용에 대한 요구가 커지고 있다. 특히 국방 분야에 관한 이슈를 실시간으로 탐지하고 관련된 이슈들을 묶어 그룹핑하여 제공하는 기술은 국가 안보와 국정 운영과 관련된 효율적인 의사결정 지원을 위한 기반 기술이다.

본 논문에서는 실시간 국방 이슈 탐지 및 그룹핑을 위한 사용자 요구사항을 수립, 분석하였다. 이를 통해 세부 핵심 기능을 도출, 실제 웹 뉴스 미디어를 대상으로 시범 시스템을 구축하였다. 국방 분야에 적합한 이슈 탐지를 위해 국방 전문 언어자원인 국방용어 통합 DB를 활용하

였으며 도메인에 특화된 이슈성 자질을 추가 발굴하였다.

자동화된 텍스트 분석 기술은 기존 전문가 집단에 의한 고비용, 저효율의 수작업 분석작업을 대체가 가능하다. 특히 주요 온라인 뉴스, 블로그, 트위터, 전문 홈페이지 등에서 국방관련 이슈를 탐지하고 추적하는 기술을 활용해 국내의 방산 거래, 가격, 예산, 신기술 개발, 무기 체계 개발, 전력화 정보 등의 관심 정보를 추출하거나 관련 출처별 연관성과 신뢰도에 대한 분석이 가능하다.

또한 국방 행정, 인사, 동원, 군수, 전력, 과학기술, 군사 분야에 대하여 언어자원을 근간으로 하는 지식정보의 구축을 통하여 각 분야의 업무 생산성과 효율성의 향상이 기대된다. 나아가 빅데이터 분석 기반 위협의 전조를 실시간으로 탐지함으로써 신속 정확한 위기관리 및 온라인 여론 동향을 반영한 정치, 경제, 사회 전반의 혁신을 지원하는 밑거름이 될 수 있다.

ACKNOWLEDGMENTS

This research was supported by “Research Base Construction Fund Support Program” funded by Chonbuk National University in 2015

REFERENCES

- [1] C. H. Lee, et. al, Technology Trends of Issue Detection and Predictive Analysis on Social Big Data, Electronics and Telecommunications Trends, vol. 28, no. 12, 2013
- [2] C. White, Using Big Data for Smarter Decision Making, BI Research white paper sponsored by IBM, 2011.
- [3] Informatica, Harnessing Big Data, <http://www.informatica.com/us/vision/harnessing-big-data>
- [4] Google Trends, <https://www.google.com/trends/>
- [5] Google Prediction Analysis, <https://cloud.google.com/prediction/docs>
- [6] Daumsoft, SocialMetrics, <http://www.socialmetrics.co.kr>
- [7] Saltlux, TrueStory, <http://politician.truestory.co.kr>
- [8] C. H. Lee, et. al, Social WISDOM: A Issue

- Detection/Monitoring System, Proc. the 38th conference of the Korea Information Processing, vol.19, no.2 pp.431-434, 2012.
- [9] United Nations Global Pulse: <http://www.unglobalpulse.org/>
- [10] Beat Habegger, Horizon Scanning in Government, Center for Security Studies ETH Zurich, 2009.
- [11] NIA, The Future direction of National Strategies based Horizon Scanning, IT & Future Strategy, vol. 5, 2013.
- [12] S. S. Kim, Study on Big Data Utilization Plans of Medical Institutions, Journal of Digital Convergence, vol 12, no. 2. pp. 397-407, 2014.
- [12] H. N. Tan and T. B. Hoo, Thinking about the Future: Strategic Anticipation and RAHS, National Security Coordination Centre (Singapore), 2008.
- [13] G. S. Lee, Smart National Defence Information Service using Language Resources. Review of Korean Society for Internet Information vol.13, no.4, pp. 54-59, 2012.
- [14] S. J. Lim, Y. J. Bae, and H. K. Kim, Korean Semantic Role Labeling Using Domain Adaptation Technique, Proc. the 26th Annual Conference on Human & Cognitive Language Technology, pp. 56-60, 2014.
- [15] B. M. Ryu, H. J. Kim, H. K. Kim, and S. K. Park, Social Media Issue Detection and Monitoring based on Deep Linguistic Analysis, Communications of The Korea Information Science Society, pp. 47-58, 2012.

오 효 정(Oh, Hyo-Jung)



- 2000년 2월 : 충남대학교 컴퓨터과 학과(이학석사)
- 2008년 2월 : 한국과학기술원 컴퓨터공학과(공학박사)
- 2000년 2월 ~ 2015년 5월 : 한국전자통신연구원 지식마인딩연구실 책임연구원

- 2015년 5월 ~ 현재 : 전북대학교 기록관리학과 조교수
- 관심분야 : 정보검색, 질의응답, 빅데이터정보처리, 소셜웹마인딩
- E-Mail : ohj@jbnu.ac.kr