

SNS에 제보되는 재해정보 추출 매시업 설계 및 구현

서태웅[†], 박만곤^{**}, 김창수^{***}

요 약

급작스런 폭우로 침수피해를 예측하기 힘든 실정에서, 재해 현장에서의 신속한 제보 및 정보 전파의 중요성이 커지고 있다. 재난 상황에서 실시간으로 정보를 수집하는 수단이 많지 않은 점에 착안하여, 본 논문에서는 소셜 네트워크 서비스를 통한 정보 수집 시스템을 설계하였다. 그리고 현재 소셜 네트워크 서비스를 융합한 여러 매시업 시스템들이 정보 확산을 목적으로만 활용하는 점을 개선하여, 기존과는 반대로 소셜 네트워크 서비스의 정보를 수집, 분석하여 재해 상황에서 최대한 활용 할 수 있도록 하는 방법을 연구하였다. 최종적으로는 트위터에 실시간으로 등록되는 재해 정보를 통해 수많은 매체 중에서 가장 신속한 재해 정보를 추출하기 위한 연구이다.

Design and Implementation of the Extraction Mashup for Reported Disaster Information on SNSs

Tae-Woong Seo[†], Man-Gon Park^{**}, Chang-Soo Kim^{***}

ABSTRACT

The quick report and propagate information are increasingly important because nowadays it is hard to predict the damages of flooding by unexpected heavy rain. In addition, there are not many ways to receive disaster information in real time. Accordingly, we designed the system which can earn information from a lot of messages on twitter. Above all, our system can extract and deploy disaster information by comparison with erstwhile social network service mash-up system as only broadcast media. Significant objective of this paper is to design the fastest extract disaster information system of mass media.

Key words: 소셜네트워크서비스(Social Network Service), 트위터(Twitter), 방재(Disaster Prevention), 재해지도(Disaster Map), 매시업(Mashup), 키워드분석(Keyword Analysis)

1. 서 론

2010년과 2011년에는 이상기후로 인한 급작스럽고 동시다발적인 폭우가 쏟아져 그 피해가 심각했는데, 당시 타임라인은 일정시간동안 재해 상황을 전파하는 트윗으로 가득 찼다. 사진과 위치 등의 정보를 포함한 주요 재해 정보 트윗은 리트윗을 통해 수차례

전파되기도 했다. 재난상황에서의 SNS활용은 트위터가 전 세계적으로 확산된 계기이기도 한데, 그 대표적인 것이 2009년 '뉴욕 허드슨강' 항공기 추락사건을 한 트위터 사용자가 제보한 경우이다. 국내에서는 해운대 고층아파트 화재사건 등 유사한 경우가 많고, 행원 풍력 발전기 화재의 경우 동영상으로 현장 상황을 전달했다[1].

※ 교신저자(Corresponding Author) : 김창수, 주소 : 부산광역시 남구 대연 3동 599-1 부경대학교 웅비관 1301호 (608-737), 전화 : 051)629-6245, FAX : 051)629-6230, E-mail : cskim@pknu.ac.kr
접수일 : 2013년 3월 22일, 수정일 : 2013년 5월 28일
완료일 : 2013년 10월 8일

[†] 부경대학교 IT융합응용공학과
(E-mail : efisode@pknu.ac.kr)

^{**} 부경대학교 IT융합응용공학과
(E-mail : mpark@pknu.ac.kr)

^{***} 부경대학교 IT융합응용공학과

※ 본 연구는 부경대학교 석사학위 논문을 바탕으로 연구된 결과를 재구성한 논문임.

그러나 현재 SNS는 일방향의 정보 전파 수단으로만 활용되거나, 재난상황의 실시간 중계, 보도자료 활용 이후에는 유용한 정보로서의 가치를 잃는다. 따라서 본 논문에서는 소셜 네트워크 서비스의 클라우드소싱 방식¹⁾으로 재해 발생시간, 위치, 영상 등의 다양한 재해 정보를 추출하여 활용하고자 하는 것이 목적이다[2]. 재난, 재해 목격자 혹은 피해자가 SNS를 활용하여 관련 부처에 재해 정보를 제보하고 이 정보를 모니터링 담당자가 재가공하는 방식은 이미 많이 사용되고 있다[3]. 그러나 이런 방식은 SNS를 매개체로 사용 할 뿐 기존 유선 신고 혹은 재난 어플리케이션을 통해 신고하는 것과 정보 확산 정도나 신속함이 크게 개선되지 않는다. 중요한 것은 재해 지역에 있는 누군가가 SNS상에 업로드 한 재해 시간, 위치, 영상 등의 정보를 신속히 IT 기술로 판별해 내는 것인데, 이때 매일 수억 개의 트윗 중에서 재해 정보만을 찾아내는 것이 본 연구의 핵심이다. 본 논문에서는 재해정보 검색의 정확도를 높이기 위해 최적의 검색 키워드를 선정하였고, 트위터 유력자 활용, 유용한 계정의 '리스트' 기능, 트위터 확산을 통해 중요도를 가늠하는 리트윗 가중치 등을 이용하여 재해정보를 추출하는 시스템을 설계 하였다.

이후에 추출된 재해정보, 위치정보, 영상정보 등은 시민들에게 제공할 수 있으며, 데이터베이스에 저장된 이 정보들은 다른 시스템과 연동되어 다시 소셜미디어로 재전송되거나, 국가 및 지자체의 재해이력 데이터로 활용 될 수 있을 것이다.

2. 재난상황에서 기존 소셜미디어의 활용 사례

현재 국내외 공공기관에서는 SNS 계정을 만들어 해당 지역의 재난정보를 서비스 하고 있다. 확산에 제한이 있거나, 소셜미디어를 재난 전파 수단으로만 활용할 뿐, 재해정보를 예측하거나, 인지하는 수단으로 활용하지 못하고 있다.

2.1 소방방재청 CBS 재난 문자 방송

소방방재청은 2004년부터 KTF와 함께 '긴급 재난

문자방송 서비스'를 제공하고 있다. CBS(Cell Broadcasting Service)방식의 이 서비스는 실시간 정보를 다수의 가입자에게 동시에 발송하는 방식으로, 휴대폰에 특정 ID를 입력해 해당 기지국 반경 내 속해있는 모든 CBS 가능 휴대폰으로 한글 60~230자의 문자 메시지를 발송한다. 이는 번호나 가입 통신사와 관계없이, 재난이 발생한 특정 지역의 휴대폰 소지자에게 발송이 되는 것이다[4]. 그러나 기술적 문제로 인해 3G망을 사용하고 있는 국민의 65%는 이 서비스를 제공받지 못하고 있는 실정이다. 최근 있었던 2011년 9월 15일 대규모 정전 사태에서도, 2G 휴대폰 서비스 가입자 1700만 명만 문자를 받고, 3G망을 이용하는 3311만 명은 문자를 받지 못했다. 2010년 폭설 및 2011년 7월 중부지역 폭우 때도 마찬가지 상황이었다²⁾.

2.2 Daum 폭우피해지도

2011년 7월 26일 중부에 폭우가 내릴 때는 Daum 아고라 게시판에 '폭우피해지도'가 개설 되다. 이는 해당 관리자 3명이 트위터를 통해 침수 제보를 받아 '구글'에서 제공하는 '내 장소'라는 기능을 이용한 서비스이다. 이 서비스는 트위터로 제보한 트윗을 읽고, 해당 위치를 지도에 표시하고, 해당 위치를 선택하면 피해 사진과 내용을 누구나 볼 수 있도록 한 것이다. 폭우 당시 중부지역을 중심으로 약 80여 곳의 침수상황을 제공하여, 약 70만 회의 조회 수를 기록해 인근 시민들이 이동시에 큰 도움을 받았다. 이 서비스는 새로운 기능이나 기술을 활용한 것은 아니지만, 시민들은 서울시나 기상청의 대처보다 우수하다 평가를 했다.

2.3 호주의 '이머전시 2.0 오스트레일리아'

호주 정부는 트위터, 페이스북, 메시업 등 소셜미디어 사이트를 연계해서 홍수경보, 대피경로와 같은 재난정보를 실시간으로 전파하려는 목표로 '이머전시 2.0 오스트레일리아'를 계획 중이다. 단 이 시스템은 본 연구와 달리, 전용 온라인 클라우드소싱 어플리케이션을 이용해야 하는데, 시민들이 차단된 도로, 비화, 목격한 연기 등을 사회 구성원들이 직접 보고

※ 본 논문은 2012년 부경대학교 석사학위 논문의 주요 내용을 발췌하여 재편집 및 보완한 논문입니다.

1) 클라우드소싱은 '대중(crowd)'과 '외부자원활용(out-sourcing)'의 합성어로, 기업이 제품이나 서비스 개발과정에 일반 대중이 참여할 수 있도록 하는 의미이다.

2) 박홍두, " '3G 휴대폰' 3311만명 재난문자 못 받아," 경향신문, 2011년 9월 23일

하여 데이터를 축적하고, 실시간으로 활용 할 수 있도록 한 것이 특징이다[5].

3. SNS의 재해정보 추출에 관한 연구

3.1 트윗 특성 연구

재해정보는 신속히 전파되어야 함에도 불구하고, 기존의 재해정보 제공 서비스들은 SNS의 장점인 신속성을 살리지 못하고 있다. 그래서 본 논문에서는 서비스 관리자의 개입을 최소화 하여 신속하게 재해정보를 취득하여 확산시키기 위한 재해정보 추출 메시지 시스템이 트위터의 재해정보를 추출하는 방법을 연구하였다. 우선 소셜미디어의 재해 글을 인식하고 추출하기 위해서 트위터에 제보되는 글인 ‘트윗’의 3가지 특성을 분석하였는데, 다음과 같이 3가지로 정리 할 수 있다[6].

① 140자 이내의 짧은 분량

일반 문서를 다루는 다양한 텍스트 분석 연구들이 이루어지고 있지만, 트윗은 140자 이내의 짧은 분량으로 구성되어 있어 글의 특성을 분석하기 어렵다.

② 정보 수준이 낮은 트윗 많음

트위터 사용자의 대다수는 개인 신상에 대한 내용을 주로 게시하기 때문에 이러한 트윗을 제거하는 것이 본 연구의 관건이다.

③ 분석 대상이 되는 문서의 수

전 세계에 5억 명 이상, 매일 3억 4천개의 트윗 등록과 16억 개의 검색 쿼리 발생(Wikipedia, 2012년), 우리나라는 642만 명 이상(OikoLab, 2012년)의 트위터 사용자 존재한다.

3.2 재해 키워드 검색을 이용한 재해정보 추출

트윗에 제보되는 재해 정보를 추출하기 위해서는 가장 먼저 재해를 종류별로 분류하고 그 재해에 해당하는 관련 키워드를 선정해야한다. 본 논문에서는 소

2011년 침수 트윗 분류

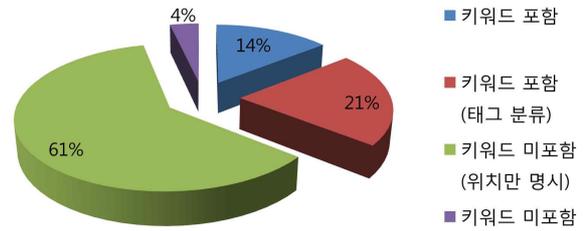


그림 1. 2011년 7월 26일 중부지역 침수 제보

방 방재청의 2008년 ‘사이버 자연재해 기술정보 통합센터 구축’ 보고서의 ‘한국형 자연재해 분류방법’과 ‘행정표준코드 관리 시스템’의 재난상황 관리 항목 중에서 호우 등으로 인한 침수피해 부분을 정리하였다[7].

한편, 일반 시민이 트위터를 사용할 때 사용되는 용어와 문장을 파악하기 위해 2011년 7월 26일 중부지역 침수 당시 트위터의 타임라인을 토대로 재해 관련 용어의 사용 정도를 분석해 보았다.

2011년 ‘침수피해지도’에 제보된 글의 대부분은 스마트폰의 위치정보와 정지 영상 이미지 정보를 포함하고 있다. 한편 당시 상황의 심각성을 비유적으로 표현하거나, 재해 분류 키워드가 아닌 혼합되거나 불확실한 용어를 사용하여 재해 정보를 알리는 트윗 등 다양한 방식의 표현으로 재해 상황을 알리고 있다. 이에 반해 ‘재해분류 키워드’를 포함한 트윗은 35%에 불과해 단순히 재해 분류에 사용된 단어를 사용할 경우, 수많은 트윗 속에서 유용한 정보를 추출해 낼 확률이 현저하게 떨어질 것이라는 것을 예측할 수 있었다.

표 2에서 보는바와 같이 일반 시민이 트윗을 작성하는 경우, 재해분류 용어나 일반적으로 예측 가능한 재해 관련 용어를 사용하는 경우가 드물었다. 따라서 기존의 재해분류 키워드보다, 실제로 트위터에 제보되는 글을 분석하여 관련 키워드를 확보하는 것이 중요하다는 것을 알 수 있다.

또한 키워드로 검색할 때 필요 없는 트윗을 제외시키기 위한 키워드 선정에 관한 연구도 필요하다. 예를 들어 표 2의 사용된 용어 중에 ‘잠겼다’라는 표현은 문이 잠긴 상황이 포함될 수 있기 때문에 이 ‘문’을 제외규칙어로 선정하면, 여닫는 물건을 열지 못하도록 잠그는 것과 같은 재해와 무관한 상황의 트윗을 제외시킬 수 있다.

표 1. 행정표준코드의 피해구분

분 류	피해 종류
인명피해	사망, 실종, 부상
시설피해	유실, 매몰, 침몰, 전파, 반파, 소파, 붕괴, 침하, 침수

표 2. 2011년 7월 폭우 때 제보되었던 트윗의 검색 키워드 추출결과

키워드 분류	사용된 용어	검색 수
표준 재해분류 키워드	침수, 호우피해, 붕괴, 토사붕괴, 유실현장, 인명피해	7
유사 키워드	안전, 폭우, 범람, 수해, 산사태, 수위, 대피, 물폭탄, 파도 몰아침, 물이 새는 상황, 물바다, 물난리, 물에 잠긴, 도로파괴, 하수구 역류, 비가 쏟아져, 난장판, 물살이 거세, 산 무너졌어요, 물바입니다, (한강)물이 불어나서, 잠기고 있다, 압사, 너무처참하다, 산에서 내려온 토사, 흙탕물, 잠겼네요, 비피해, 물에 잠겼네요, 난리, 한강은 미숫가루물색	37

3.3 트위터 사용자의 특성을 이용한 재해정보 추출

3.2절에서는 재해 키워드 검색을 통한 재해 트윗 추출만으로는, 다양한 규칙성을 가진 트윗을 정확하게 추출해 내지 못한다는 것을 보여 주었다. 그렇기 때문에 3.3절에서는 유용한 정보를 확산시키려는 트위터 사용자의 특성을 이용하여, 재해 정보 취득 방법을 보완하였다.

3.3.1 ‘트위터 유력자’를 이용한 재해 정보 추출

일반 트위터 사용자의 트윗은 한 두 단계를 거쳐 소멸되고 넓게 퍼지지 못하지만, 유력자(Influencers)를 통해 일반 트위터 사용자의 트윗이 널리 전파될 수 있다. ‘유력자’들은 사회적 여론 형성 과정에서 큰 영향력을 발휘하고 있는데, 이들 유력자가 정보의 생산에 치우쳐 있다면 ‘매개적 유력자’는 유력자의 역할보다는 유용한 정보를 전달하는 매개체로서의 역할을 수행한다. 특히 이 매개적 유력자들은 사건·사고와 관련된 트윗에 대해 많은 리트윗을 유발한다.

재해정보 수집에 이러한 트위터 유력자의 행동 특성을 이용할 수 있다. 저수준의 재해 트윗을 인식하고, 재해정보를 해석, 배포 해줄 유력자를 선정하는 것이 중요한데, 2011년 이원태 외 2명은 소셜미디어의 유력자를 구분함에 있어 다양한 기준 제시했다 [8]. 이 논문에서는 단순히 많은 팔로어수, 기존 유명인사로 유력자를 구분하는 것이 아니라, 유력자가 올리는 유용 트윗의 리트윗 횟수를 고려하여 계정을 선정하였다.

이들의 특징은 팔로어 수보다는 트윗, 멘션 등과 같은 직접적 참여와 사회적 교류를 통해 뉴스미디어 기능을 수행하는 것이다. 유력자는 평균 15,000명 이상의 팔로어를 보유하고, 팔로잉 수, 트윗 수 모두 일정이상 되는 트위터에서 활발하게 활동하는 언론인, 기업인, IT전문가 등 다양한 분야의 전문 직업을

지닌 사용자였다. 이들은 직접 정보를 생산하거나 전달하는데 반해, 매개적 유력자는 일부 유력자도 포함 되어있으며, 고등학생, 대학생, IT관련업계 종사자, 쇼핑몰 운영자등 다양한 직업군을 형성했다. 유력자의 트윗은 많은 팔로워에게 타임라인을 통해 직접 전달되는 반면, 매개적 유력자는 리트윗을 통해 급속하게 전파하는데 일조를 한다. 오히려 본 연구에서 요구되는 재해 정보를 판단하기 위해서는 매개적 유력자의 리트윗 수가 더욱 유용하게 쓰일 수 있다는 결론을 낼 수 있었다.

3.3.2 유용한 계정의 리스트 구동

트위터의 ‘팔로우(Follow)’기능 이외에 재해정보를 얻기 위해 임시로, 특정 사용자의 글을 수신 받아야 할 때가 있다. 트위터에서는 ‘리스트(List)’라는 기능을 통해 사용자 그룹을 만들어 원하는 정보를 받아볼 수 있다. 재해지역의 정보를 트위터를 이용해 전파하는 ‘유용한 사용자’가 있기 마련인데, 재해 상황 일 때의 타임라인은 재난현장에 있는 사용자의 글, 그 정보를 반복적으로 리트윗하는 사용자의 글, 재난에 관해 언급은 하지만 불필요한 글 등 많은 정보들이 수집될 수 있기 때문에, 유용한 정보만을 취득할 수 있도록 방법을 개선해야 한다. 단, 이때 ‘유용한 사용자’는 유력자와는 다른 개념으로 유력자는 미리 계정을 확보하여 그 대상을 팔로우 하고 타임라인에서 정보를 확인할 수 있다.

사용자 특성을 활용하여 재해정보를 취득하기 위해 실시간으로 트위터의 리스트 기능을 활용하는데, 리스트에 포함될 재해 상황에서 주요한 계정은 다음 4가지 특성을 가지고 있다. 이 특성들은 2010년 중부권 폭우 당시, 실제로 리스트를 만들어 상황을 중계했던 한 블로거³⁾가 타임라인의 재해정보를 제공하

3) 김광현, 한국경제신문 IT전문기자 블로그(blog.hankyung.

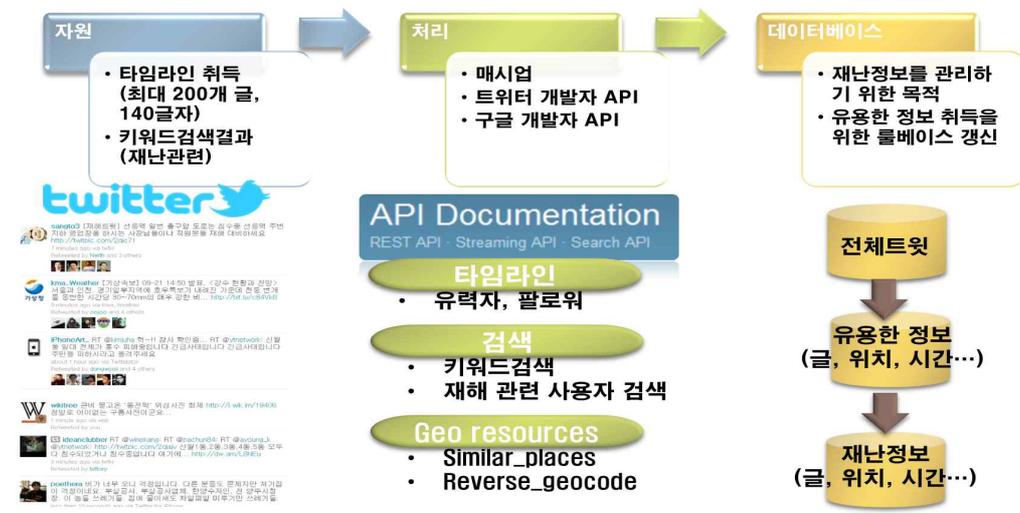


그림 2. 재해 정보 수집 절차

는 트위터 사용자의 행동들을 정리한 내용이다.

- ① 현장에서 계속 상황 전파
- ② 현장에서 단발성 상황 전파
- ③ 중요 트윗을 계속 확산(리트윗)
- ④ 중요 정보를 정리하여 새로 작성하여 전파

3.3.3 리트윗 특성을 이용한 중요도 평가

트위터 사용자들은 타임라인에서 중요한 정보를 인지하게 되면, 그 트윗 글을 리트윗을 하는데, 이때 리트윗 횟수를 통해 해당 트윗의 중요도를 판단할 수 있다. 사용자의 트윗 해석 능력을 리트윗을 통해 분석하는 것인데, 이를 활용하면 트윗의 정보 자체에 오류가 있거나 거짓정보를 포함 할 때 수많은 트위터 사용자들의 정보 자정 능력을 활용할 수 있는 장점이 있다.

실제 재해 및 사고 제보 트윗이 제보되면 신속하게 리트윗 되는데, 최근 발생한 대규모 침수, 전국적인 정전 사태, 당시에는 대부분의 소셜미디어 타임라인이 관련글로 넘쳐났고, 2010년 10월 부산 해운대구의 대형 아파트 화재 영상이미지는 조회 수 2위, 5위 8위 차지하여 즉각적인 반응을 소셜미디어에서 확인할 수 있었다(이원태 2011). 이러한 사례를 뒷받침하는 통계적 연구를 관련연구에서 찾아볼 수 있는데 이원태 외 2명의 논문에서는 트윗의 주제별로 리트윗 수와 리트윗 되는 주기(시간)을 분석하였다. 사건,

사고와 같은 사회적 이슈에 관한 리트윗의 주기는 평균 40분에서 약 1시간 정도에 걸쳐 짧게 나타난 반면, 기술과 정치에 관한 메시지는 평균 7시간에서 최대 1주일까지, 상대적으로 긴 주기를 나타냈다. 이러한 특징으로 볼 때, 짧은 시간동안 다발적으로 리트윗 되는 특정 트윗을 추출하여 재해 및 사건 사고의 규모를 예상 할 수도 있다[9].

리트윗 카운트의 임계치를 설정하기 위해서, 키워드 분석 검색 방법으로 검출된 트윗을 조사한 결과, 일부 트위터 사용자는 중요 정보를 인식하고 리트윗을 하거나, 원본 트윗의 일부를 인용하는 경우가 있었는데, 보통 3회 이상의 재배포 횟수를 나타냈다. 단, 재해 검색 키워드로 사용된 용어들의 사용 빈도와 용법에 따라 리트윗 횟수는 많은 차이를 보였기 때문에 재해 검색 키워드에 따른 리트윗 카운트의 임계치는 달리할 수 있을 것이다.

4. 재해정보 추출 설계 및 구현

4.1 매시업 시스템 설계

3장에서 제시한 재해정보를 추출하는 방법(키워드추출, 트위터 유력자활용, 리트윗 중요도 평가)의 을 적용하여, 4장에서는 트위터 API로 매시업 프로토타입을 구현하였다.

그림 3의 좌측 상단 ‘재해 키워드 분석 적용’의 체크박스를 선택하면 본 논문에서 제시한 재해 트윗추출을 위한 방법들이 적용이 된다. 그리고 검색된



그림 3. 재해정보 추출 프로토타입의 구동 화면 및 검색 결과

트윗 중에서 위치정보(좌표)가 있을 경우 우측 구글 지도에 표현(지도표현-1)되도록 했다. 이 좌표를 기반으로 반경검색을 수행 할 수 있는데, 이 반경검색은 재해 인근 지역에서 발송되는 트윗을 추출하여 보다 더 정확한 재해 관련 정보를 추가적으로 추출할 수 있다. 만약 트윗에 좌표정보가 포함되지 않았을 경우에는 트위터 계정에 등록된 위치정보를 통해 지오코딩을 수행한 후, 지도에 표현(지도표현-2)이 가능하다. 주요 트윗의 위치정보 포함여부, 혹은 사용자 프로필에 포함된 위치정보의 정확도에 관한 상세 연구는 기 수행된 연구를 참조하여 재해정보 추출 및 활용 성능을 개선할 수 있으므로 본 논문에서는 다루지 않았다[10].

4.2 재해정보 검색 실험 결과

일반 키워드 검색과 본 연구의 재해 추출 방법의 정확도(Searching Accuracy)를 각각 수식 (1), (2)로 나타내었다. 소방방재청의 한국형 자연재해 분류방법, 행정표준코드의 피해 분류 키워드인 유실(Disaster Type : d=1), 매몰(d=2), 침몰(d=3), 전파(d=4), 반파(d=5), 소파(d=6), 붕괴(d=7), 침하(d=8), 침수(d=9), 호우(d=10) 등 10개의 키워드로 각각 100개의 트윗을 추출(General Retrieval) 하여 분석하였다. 추출 합계를 $\sum(\Sigma)$ 로 표현 했고, 위아래 첨자는 10개의 검색 키워드 1-10으로 나타내었다. 트위터 API를 사용해 일반 검색을 수행한 결과 총 987개의 트윗이 검색되었으며, 그중 본 연구의 재해 트윗 추출 방법을 적용하여 검색된 트윗은 330개 그

중 107개가 유용한 재해 정보(재해 분류, 위치, 발생 시간 등)를 포함하고 있는 트윗이었다. 수식 (1)은 일반 검색 결과(987개) 중에서 유용한 재해 정보(107개)의 비율이며, 수식 (2)는 본 연구의 재해 트윗 추출 검색 결과(330개) 중에서 유용한 재해 정보(107개)의 비율이다. 단, 동일한 시점에서 트윗 검색 결과를 비교하기 위해서 1000개 이상의 트윗을 무작위 검색하여 html형태로 저장한 데이터를 사용했음을 밝혀 둔다.

$$Acc_{gnr} = \frac{\sum_1^{10} Dis_d}{\sum_1^{10} Gnr_d} \times 100 = 107 \div 987 \times 100 = 10.84\% \quad (1)$$

where,

Acc_{gnr} (searching Accuracy by General retrieval) : 일반검색의 재해 트윗 검색율

Dis_d (Disaster tweets from disaster types) : 재해 분류에 따른 중요 재해 트윗 수

Gnr_d (General retrieval from disaster types) : 트윗의 일반 검색 결과

$$Acc_{anal} = \frac{\sum_1^{10} Dis_d}{\sum_1^{10} Anal_d} \times 100 = 107 \div 330 \times 100 = 32.43\% \quad (2)$$

where,

Acc_{anal} (searching Accuracy by Analyzed retrieval) : 본 연구의 재해트윗 검색율

$Anal_d$ (Analyzed retrieval from disaster types) : 본 연구의 검색방법을 적용 결과

본 연구의 검색 시스템을 통해, 987개의 전체 일반

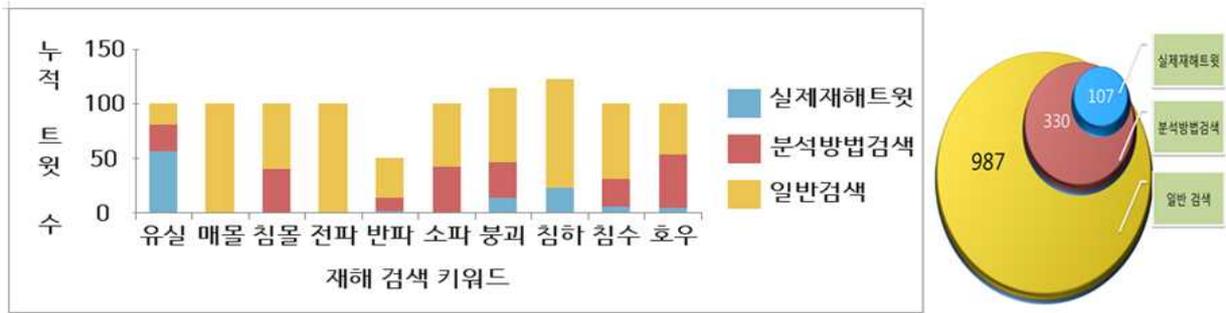


그림 4. 검색 방법에 따른 트윗 추출 실험 결과

키워드 검색 결과 중에서 657개의 데이터를 필터링 할 수 있었다. 재해 트윗 추출 방법이 기존의 일반 키워드 검색 결과와 비교하여 얼마나 정확도가 향상되었는지 개선도(Improvement Factor)를 수식 (3)을 통해 나타 내었다.

$$\begin{aligned}
 IFact &= \frac{\sum_1^{10} Gnr_d - \sum_1^{10} Anal_d}{\sum_1^{10} Gnr_d} \times 100 \\
 &= \frac{987 - 330}{987} \times 100 = 66.57\% \quad (3)
 \end{aligned}$$

where,

IFact (Improvement Factor) : 본 연구의 검색 방법 적용의 재해 트윗 검색 개선도

그림 4는 수식 (1), (2)와 같이, 일반 검색과 본 논문의 재해 검색 방법을 적용하여 매시업 프로토타입으로 검색을 실시하여 재해정보를 추출한 결과를 재해 검색 키워드별로 비교하여 나타낸 것이다. 노란색 부분이 일반검색을 수행하여, 분석 기반이 되는 데이터이며, 붉은색 부분은 최종적으로 검출 정확도 개선으로 사라져야 할 트윗이다. 침하의 경우 본 논문의 분석 방법을 통해 실제 재해 트윗만을 검색해 냈기 때문에 붉은색 부분이 없는 경우이다.

5. 결 론

기존 재난 관련 서비스들은 관리자가 확인하는 절차를 거치면서 신속성이 크게 떨어지기 때문에[10], 본 논문에서는 클라우드소싱 개념을 포함하여 재해 상황을 신속하게 추출하는 연구를 수행하였다. 재해 상황에서의 정보전달의 특성상 누가, 언제, 어디서, 어떤 매개체를 이용해 재해 정보를 게시하고, 이 게시 정보가 관련 부처나 불특정 다수에게 신속하고

효과적으로 전달 가능한지에 대한 계측 방법을 정의하기 어렵다. 정보 전파의 시작은 정의할 수 있지만, 정보 전달이 완전히 마무리 되었다는 시점이 불분명하기 때문이다. 관련부처에 신고접수 시점, 특정 인원의 정보 수신 시점, 특정 장소 도달 시점 등 한 가지를 특정할 수 없는 현실이기 때문에 본 논문에서 결론지을 수는 없다. 하지만 제보자, 관리자 등, 재해 상황에서 상황별 대응 행동에 따라 시간이 크게 지체되는 것을 고려하여 절차를 최소화 한 SNS기반의 재해정보 추출 방법은 진보된 재해 정보 시스템으로 의의 있다고 판단하며, 이 연구를 기초로 추후 연구에서는 유사 시스템과의 계량 가능한 비교 분석, 개선된 시스템의 설계, 실제 재해 상황에서의 유력자, 매개적 유력자 등의 활동으로 키워드 검색 방법을 사용한 실험, 사례 데이터 확보를 통해 재해 정보를 더 신속 정확하게 추출 해낼 수 있을 것이다.

또한, 현재 SNS의 한계점도 찾을 수 있었는데, SNS에 제보되는 글의 신뢰성이다. 모든 SNS는 자신의 글을 삭제하는 기능이 있지만, 누군가 재배포한 자신의 글은 삭제할 수 없다. 그렇기 때문에 실수든, 고의든 잘못된 정보가 한번 타임라인에 노출되면 그 글은 쉽게 퍼져나가 쉽게 혼란을 야기한다. 따라서 사용자 특성, 리트윗 가중치 등의 다양한 방법을 통해 재해정보 취득 절차에서부터 정확한 정보를 추출하는 향후 연구가 요구된다.

참 고 문 헌

[1] 김도령, 소셜 네트워크 서비스기반 어플리케이션의 인터페이스 디자인에 관한 연구, 한양대학교 석사학위논문, 2011.
 [2] 서태웅, “트위터의 재난정보 자동 추출 매시업 연구,” 한국해양정보통신학회 추계학술발표회,

- 제15권, 제2호, pp. 515-517, 2011.
- [3] 강희조, “SNS를 활용한 스마트 재난관리체계에 관한 연구,” 한국향행학회논문지, 제16권, 제4호, pp. 717-722, 2012.
 - [4] 한영미, 방재정보 서비스디자인 연구 : 긴급재난 커뮤니케이션 서비스 개선방안을 중심으로, 이화여자대학교 석사학위논문, 2011.
 - [5] 위기관리경영(前 재난포커스) : 호주 재난상황관리-일본재난정보 전파, <http://www.di-focus.com/news/articleView.html?idxno=2253>, 2010.
 - [6] 김태연, 키워드 그래프를 활용한 트윗 검색 결과 요약, 성균관대학교 학위논문, 2011.
 - [7] 이중기, 김창수, “스마트 폰 애플리케이션 재난정보 서비스 및 검색기능 구현,” 멀티미디어학회논문지, 제16권, 제2호, pp. 273-280, 2012.
 - [8] 이원태, 차미영, 양해륜, “소셜미디어 유력자의 네트워크 특성 - 한국의 트위터 공동체를 중심으로,” 서울대학교 언론정보연구소 언론정보연구, 제48권, 제2호, pp. 44-79, 2011.
 - [9] 이범석, 김석중, 황병연, “트위터 사용자가 제공한 위치정보의 신뢰성 분석,” 멀티미디어학회논문지, 제15권, 제7호, pp. 910-919, 2012.
 - [10] 황현숙, 최은혜, “지리정보시스템을 활용한 재난피해자 관리시스템,” 한국지리정보학회지, 제14권, 제1호, pp. 59-72, 2011.



서 태 응

2003년~2010년 부경대학교 컴퓨터멀티미디어공학과 공학사
 2010년 3월~현재 부경대학교 정보공학과 석사
 관심분야 : 소셜네트워크, 재난정보관리, 지리정보시스템



박 만 곤

경북대학교 수학교육(이학사)
 경북대학교 전산통계학(이학박사)
 Philippine Women's University (국제행정학석사)
 University of Rizal System, Philippines (명예 기술학박사)

Dept. of Electrical & Computer Engineering, University of Kansas (Post Doc.)
 1981년~현재 부경대학교 IT융합응용공학과 교수
 2008년~현재 한국멀티미디어학회(KMMS) 회장 및 명예회장
 2002년~2007년 정부간 국제기구 CPSC (콜롬보플랜기 술교육대학) 총재 (Director General and CEO)
 2004년~2007년 Asia Pacific Accreditation and Certification Commission 아태지역 인증 및 검증위원회 위원장
 2005년~2007년 유네스코 (UNESCO-UNEVOC) 자문위원, 아시아개발은행(ADB) 자문관
 관심분야 : 소프트웨어신뢰성공학, 비즈니스 프로세스 재공학 (BPR), 소프트웨어 공학 및 재공학, 멀티미디어정보처리기술, 정보시스템성능평가, ICT-based HRD System



김 창 수

1991년 중앙대학교 컴퓨터공학과 박사
 2006년~현재 유비쿼터스 부산도시협회 방재분과위원장
 2012년~현재 부경대학교 산업과학기술연구소 소장

2013년~현재 한국멀티미디어학회 정책자문위원
 1992년~현재 부경대학교 IT융합응용공학과 교수
 관심분야 : 방재IT, UIS/GIS, 운영체제, 재난관리, 공간 검색, 도시방재 등