

# 도로 침수영역의 탐색을 위한 빅데이터 분석 시스템 연구

송영미<sup>†</sup>, 김창수<sup>\*\*</sup>

## A Study on the Big Data Analysis System for Searching of the Flooded Road Areas

Youngmi Song<sup>†</sup>, Chang Soo Kim<sup>\*\*</sup>

### ABSTRACT

The frequency of natural disasters because of global warming is gradually increasing, risks of flooding due to typhoon and torrential rain have also increased. Among these causes, the roads are flooded by suddenly torrential rain, and then vehicle and personal injury are happening. In this respect, because of the possibility that immersion of a road may occur in a second, it is necessary to study the rapid data collection and quick response system. Our research proposes a big data analysis system based on the collected information and a variety of system information collection methods for searching flooded road areas by torrential rains. The data related flooded roads are utilized the SNS data, meteorological data and the road link data, etc. And the big data analysis system is implemented the distributed processing system based on the Hadoop platform.

**Key words:** Big Data Analysis, Social Media, Flooded Road Areas, Searching System

### 1. 서 론

우리나라는 지구 온난화로 강한 태풍의 발생은 물론 게릴라성 집중호우의 빈도가 증가하고 있어 자연 재해로 인한 피해는 해마다 증가하고 있는 추세이다. 2010년대에 접어들면서 서울은 물론 전국 대도시 지역은 집중호우로 침수 피해를 입었고, 특히 2014년 8월에는 부산지역에서 기록적인 게릴라성 호우가 발생하여 저지대 침수 피해는 물론 일부 도로가 침수되어 차량은 물론 어린이 또는 노약자의 보행에도 위험이 발생하였다. 이러한 관점에서 사람들에게 피해를 줄 수 있는 도로중심의 침수로 인한 재난 예방에 대한 연구가 필요하다. 이러한 관점에서 다양한 원인에 의해 집중호우로 인하여 도로 침수 영역의 탐색을

위한 실시간 재난 정보를 수집할 수 있는 방법이 필요하다. 이를 위해서 기상청에서 제공하는 다양한 정보들은 물론 토목 및 수문학 관점의 오프라인 정보들과 그리고 IT기반의 실시간 정보들을 종합적으로 수집하고 분석할 수 있는 빅(Big) 데이터 기반의 분석 시스템에 대한 연구가 필요하다.

빅 데이터 관점에서 보면 디지털 사회에서 인터넷에 연결될 수 있는 스마트 기기들의 보편화로 인해 네트워크상에 퍼져있는 정보는 일반적으로 처리해 오던 데이터의 양을 초과하였고, 이에 따라 기존의 정보 처리 방식으로 처리할 수 없는 방대한 양의 데이터를 처리할 수 있는 방식이 개발되고 있다. 소셜 네트워크 서비스(Social Network Service : SNS)의 경우 하루 발생 데이터가 7테라 바이트에 이르며, 폐

\* Corresponding Author : Chang Soo Kim, Address: (608-737) Ungbi Bldg., Yongso-Ro 45, Nam-Gu, Busan, Korea, TEL : +82-51-629-6245, FAX : +82-51-629-6230, E-mail : cskim@pknu.ac.kr

Receipt date : July 16, 2015, Approval date : July 22, 2015

<sup>†</sup> Dept. of Information System, Graduate School, Pukyong National University (E-mail : dudal1320@gmail.com)

<sup>\*\*</sup> Dept. of IT Convergence and Application Engineering, Pukyong National University

\* This work was supported by a Research Grant of Pukyong National University(2013Year:C-D-2013-0468).

이스북 또한 10테라 바이트에 접근한[1]. SNS 데이터의 경우에는 실시간으로 데이터의 수집 및 처리가 가능하여 최근의 기법들을 적용할 경우 분석 결과를 바로 제공할 수 있다. 그러나 분석되는 데이터 유형이 관계형 데이터베이스 시스템에서 처리가 가능한 정형적인 데이터 외에 텍스트, 비디오 등과 같은 반정형 또는 비정형 데이터가 90% 이상 존재하면서 처리의 어려움이 있다[2]. 특히 뉴스, 블로그, SNS 등과 같은 서비스에서 포함하고 있는 데이터는 정해진 형식이 없고 불특정 다수가 작성한 텍스트, 오디오 및 비디오 미디어가 포함되어 있고, 텍스트 데이터 영역에서도 축약, 은어 및 비어, 오타자를 포함하는 구어체의 특징을 가지고 있으므로 빅데이터 상에서 실제 사용자가 원하는 정보를 추출해 내는 과정이 신속하고 정확하게 처리되기가 어려운 실정이다. 그리고 이러한 빅데이터는 일반적인 플랫폼 상에서 처리하기에는 속도 상의 문제가 있기 때문에 분산 처리 플랫폼에서의 원활한 병렬 데이터 처리 방식 등이 고려되어야 한다.

Fig. 1은 도로 중심의 침수 영역 정보를 수집하고 위험 여부를 검색하여 사용자 및 정부기관의 관리자에게 실시간 정보를 제공하기 위한 빅데이터 분석 시스템에 대한 개요도이다. 본 연구에서는 호우 발생 시 침수 도로를 검색하기 위해 도로 침수 분석 및 추출이 가능한 다양한 정보를 기반으로 빅 데이터 분산 처리 플랫폼 환경에서 재난 정보를 분석하여 사용자들에게 침수 도로에 대한 정보를 제공할 수 있는 방법을 제안한다. 도로침수 분석 및 추출이 가

능한 데이터의 적용은 다양한 정보들이 이용될 수 있으나, 본 연구에서는 주요 데이터로 소셜 미디어 데이터, 기상 상태 데이터 그리고 도로링크 데이터를 중심으로 연구를 진행하였다.

## 2. 관련연구

### 2.1 빅데이터의 분산 처리 환경

빅데이터는 V3라는 용량(Volume), 속도(Velocity), 다양성(Variety)의 세 가지 요소로 구분할 수 있으며 2가지 요소가 충족된다면 빅데이터로 볼 수 있다[3]. 단순히 정보가 크다는 것뿐만 아니라 데이터 생성에서 활용까지의 빠른 속도와 데이터 내용의 형태성에서 특징을 찾을 수 있다. 데이터 유형에는 정형, 반정형 그리고 비정형의 데이터 형식을 모두 포함하고 있으며, 특히 비정형 데이터에는 SNS에서 제공하는 텍스트 또는 미디어 파일을 예로 들 수 있다. 또한 SNS 등의 빅 데이터는 수십 테라, 페타, 제타 바이트에 달하는 용량들을 포함하고 있으며, 급격하게 증가하는 데이터들은 기존 대용량 서버시스템에 저장할 경우 데이터의 장애를 포함한 손실의 문제점을 가지고 있다. 이러한 문제점을 해결하기 위해 최근에는 빅데이터 수집 및 분석을 위해서 아파치하둡과 같은 분산 컴퓨팅 솔루션이 사용되고 있다. 아파치 하둡은 이미 많은 연구자들이 빅 데이터 분석을 위해서 사용하고 있으며, 분산 처리 파일 시스템인 HDFS(Hadoop Distributed File System)와 태스크를 조작할 수 있는 맵리듀스(MapReduce)를 구현한 자바 기반의 플랫폼이다[4].

국내에서는 크고 작은 자연재해가 매년 발생하고 있기 때문에 현 정부에서는 정부 및 지자체가 보유하고 있는 재해 정보에 대해 연구자들과 관련 기관들에게 공유하도록 정부3.0 슬로건을 지원하고 있다. 이러한 정부3.0 정책은 본 연구의 다양한 정보 수집을 가능케 하여 분산 처리와 같은 빅데이터를 처리할 수 있는 새로운 기술을 적용할 수 있는 환경이 가능하게 되었다[5].

### 2.2 기존의 도로 침수 탐색 기법

태풍에 의한 집중호우는 게릴라성 집중 호우든 도로 침수는 도로의 다양한 원인에 의해 침수가 된다. 일반적으로 대부분의 도로는 50년 빈도 이상의 집중

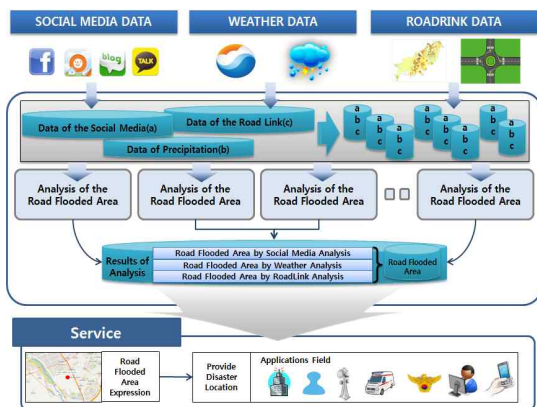


Fig. 1. Big Data Analysis System for Disaster Prevention based Road Link.

오후에도 상하수 처리가 가능하도록 설계되어 있다. 그러나 도로는 집중호우에 의한 높은 지역에서 낮은 지역으로 많은 수량이 이동하면서 나뭇잎, 흙 등 하수처리를 불가능하게 하는 원인들이 존재한다. 이러한 원인들은 대부분 담당 지자체들이 매년 집중호우에 대비한 도로 정비를 진행하고 있지만, 현실적인 측면에서 어려움이 상존한다. 따라서 대도시 주변의 도로의 침수는 완전히 예방할 수 없기 때문에 일어날 수 있다는 가정하에 사람들과 차량에 피해를 예방할 수 있는 방법의 연구가 필요하다.

김은미 등[6]은 도로 침수 탐색을 위해 실제 기상 상태를 활용하여 재난 지역을 추출하는 실시간 도로 침수위험지수 산정 모델을 개발함으로써 게릴라성 호우 등에 의한 도로 침수 구역을 추출하는 연구를 진행하였다. 그리고 실시간 강수량을 이용한 강우 시점의 도로에 대한 침수위험지수를 개발하기 위한 도로별 침수 이력, 도로별 침수위험강우량 지수, 실시간 누적강우량 지수, 도로별 설계 확률 빈도 강우량 등의 데이터를 활용하여 도로 침수위험지수를 산정하였다. 침수 지역의 지형적 특성 반영을 위해 각 도로의 평균고도를 계산하였고 도로 침수가 발생했던 시점의 강수량을 수집하여 기준 값을 설정하였다. 또한 과거의 침수로 인해 교통이 통제된 도로와 날짜를 추출하고 침수흔적도를 수집하여 침수 지역에 포함되어 있는 도로를 추출하였다. 이러한 과정을 통해 실시간으로 강수량이 발생하면 도로가 침수되는 기준 값과 비교하여 도로 침수 영역을 추출하는 연구를 하였다. 이전 연구에서는 도로에 대한 침수 위험 강수량을 자동기상관측장비(Automated Weather System, Automatic Weather Station : AWS)에서 측정된 값에 대해 역거리 가중치 기법으로 부산시를 대상으로 파이썬 스크립트를 사용하여 도로 침수 영역을 계산하였다[7,8].

서태웅 등[9]은 빅데이터 분야인 소셜 네트워크 분석으로 트윗에 제보되는 재해 정보를 추출하기 위해 재해를 분류하고 키워드를 선정한 후, 사회적 여론 형성에 영향력을 발휘하는 트위터 유력자의 리트윗 횟수를 통해 실시간 재해정보 취득을 위해 트위터 리스트 기능을 활용하는 연구를 진행하였다. 이를 위해 재해 정보와 관련된 리트윗을 카운트하고 중요도를 평가하여 침수 지역에 대한 추출 메시지를 설계하고 구현하였다. 여러 SNS 중에서 트위터만을 사용하

여 연구했음에도 수많은 데이터를 사용하여 분석한 연구 결과와 유사한 결과를 제시하고 있다. 이는 SNS 기반의 침수 지역 추출에 대한 메시업의 정보도 실시간 도로 침수 영역을 연구하는데 기반 데이터로 활용될 수 있음을 나타내고 있다.

### 3. 도로 침수 영역을 위한 빅데이터 수집 및 분석

앞에서도 설명하였듯이 집중호우에 의한 실시간 도로 침수 영역의 탐색은 도로 침수와 관련된 과거의 정보는 물론 도로 영역의 실시간 상황 정보를 최대한 빨리 수집하여 각 도로 영역의 침수상황을 탐색하는 연구가 필요하며, 이러한 다양한 데이터의 수집은 필수적이다.

이를 위해 본 연구는 도로 침수와 관련된 다양한 데이터 수집을 진행한 후, 분산 처리 플랫폼 환경에서 도로 침수 영역을 탐색하는 방법을 제안한다. 따라서 도로 침수 영역을 탐색하기 위해 사용되는 데이터는 여러 종류가 있지만, 본 연구에서는 주로 SNS 데이터, 기상 상태 데이터 그리고 도로링크 데이터를 중심으로 수집과 분석을 진행한다. 기존의 연구들[6, 7]은 실시간 강수량과 과거 침수가 발생했던 도로의 이력을 활용하여 수식을 통해 침수 도로를 예측하고 있으나, 도로의 정보를 처리하고 광범위한 지역을 동시다발적으로 계산이 발생할 때 속도 측면에서 효율성이 떨어지고 있으며, [9]는 SNS를 통한 침수 지역 추출에 관련된 연구에도 수만 명의 사용자들이 제각기 다른 내용의 정보를 제공하기 때문에 분석된 결과를 얻기까지 실시간적으로 처리되기에 문제점을 가지고 있다.

따라서 본 연구에서는 SNS를 포함하고 뉴스 및 블로그의 미디어까지 포함하는 소셜 미디어 데이터와 각 지역의 강우량을 확인 할 수 있는 기상 상태 데이터, 그리고 도로의 노드 및 고도 등의 정보를 포함하고 있는 도로링크 데이터를 빅데이터로 분산 처리하여 실시간으로 도로 침수 영역을 탐색할 수 있도록 제안한다. Fig. 2는 소셜 미디어, 기상 상태, 도로링크 데이터의 수집 내용과 각 데이터를 데이터베이스에 저장하는 과정을 설계도로 나타낸 그림이다. 소셜 미디어 데이터 수집은 다시 텍스트, 이미지, 그리고 GPS 값 등을 추출하여 침수된 도로 영역과 관련된 정보들로 수집 및 분석되며, 기상 데이터도 AWS

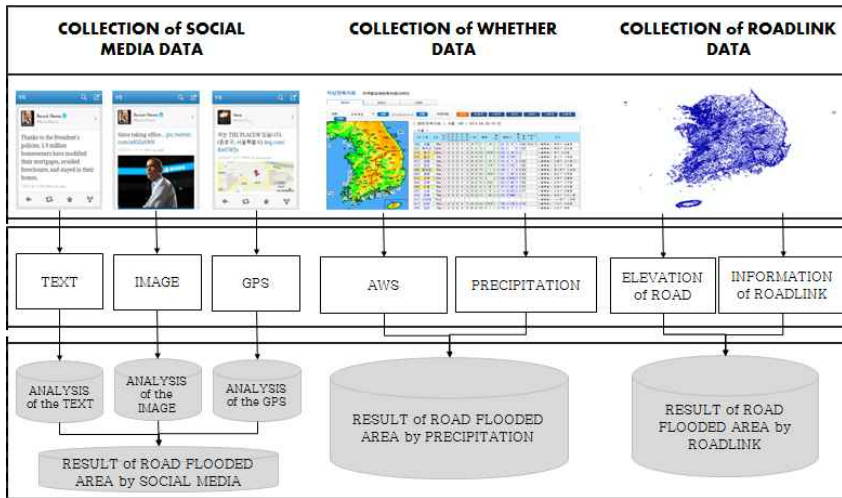


Fig 2. The Data Collection Process of Flooded Roads.

정보와 예측정보들을 수집하여 도로침수 영역에 활용되도록 저장된다. 그리고 도로링크 데이터는 실시간 정보는 아니지만 도로 침수 영역을 추출하기 위해 도로의 고도정보, 도로의 링크 정보들이 저장되어 분산처리 플랫폼 환경에서 적용될 수 있도록 수집 및 분석하는 과정을 나타내고 있다.

3.1 도로침수 정보 수집을 위한 소셜 미디어 데이터

소셜 미디어 데이터는 불특정 다수의 사람들이 다량의 비정형 데이터를 생성하기 때문에 데이터 수집을 위한 주기적인 모니터링이 필요하고, 업데이트 되는 소셜 미디어를 실시간으로 수집하며 데이터가 중복되지 않도록 관리되어야 한다. 이를 위해 본 연구에서는 도로침수와 관련된 1차 데이터 수집을 위해 태풍 및 집중호우는 자연재해와 관련된 것이 대부분이기 때문에 자연재해와 연관된 키워드를 중심으로 SNS를 검색하여 추출된 결과만을 수집한다. 따라서

Fig. 3은 본 연구와 관련된 자연재해 중심의 키워드들을 분류한 것으로, 본 연구에서는 이러한 자연재해 키워드에서 도로 침수와 관련된 정보들을 추출하게 된다.

2차 수집은 자연재해 키워드 외에 사람들이 언급하는 빈도수가 높은 단어를 관리자가 미리 데이터베이스에 재해 키워드로 설정해 놓음으로서 기존의 재해 키워드만으로 수집하지 못한 소셜 미디어까지 수집할 수 있도록 한다. 1차 및 2차 수집 방법을 통해 수집된 결과에서 실제 사용하는 데이터는 텍스트, 이미지 그리고 GPS가 된다.

GPS 데이터는 글을 게시한 사람이 재해 위치를 알리고자 GPS 정보를 적제하였다고 판단할 수 있으나 실제 도로 침수 영역의 위치와 다를 수 있기 때문에 게시된 텍스트 영역을 참고하여 텍스트 상에서 분석된 재해 위치와 GPS 영역의 위치가 유사할 경우만 사용하도록 제안하고 있다. 그리고 이미지 콘텐츠

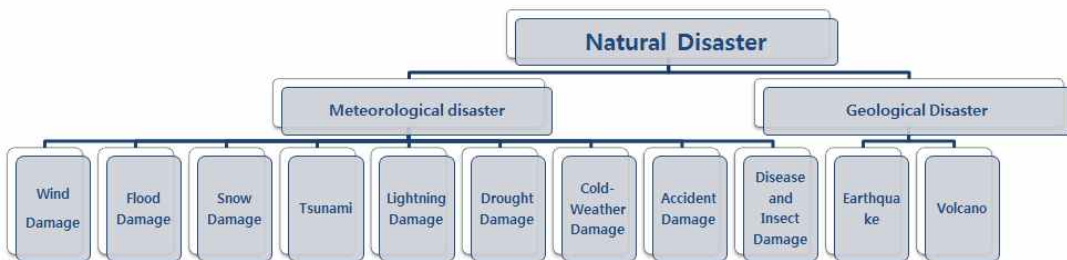


Fig. 3. Keyword of the Natural Disasters related Road Flooding.

의 경우는 EXiF(EXchangable Image File format) 태그를 포함하고 있는 디지털 사진일 경우 GPSInfo 태그가 존재하고, 하위 단계에 GPSTagDirectionRef와 GPSTagDirection 태그 값이 존재하는데 우리는 이를 활용한다. GPSTagDirectionRef 태그는 피사체의 방향 부호를 그리고 GPSTagDirection 태그는 피사체의 방향을 나타내는데, 디지털 사진이 촬영되었을 때의 방향과 각도를 계산하여 위치를 추출하고 근접 도로의 침수가 발생 여부를 판단한다[10].

마지막으로 텍스트 데이터 분석은 기존 소셜 미디어 분석 기법 또는 사회 분야에서 이슈 탐색을 위해 개발한 OpenAPI를 활용해서 침수 도로를 탐색한다. 본 논문에서 사용하고자 하는 OpenAPI는 Saltlux사에서 개발한 O2 Service OpenAPI의 주제어 분석 API를 활용한다. 이는 주어진 질의어에 대해 연관 주제어들을 추출하고 단어를 의미단위로 분리하여 단어의 출현 빈도를 기준으로 자주 등장하는 단어에 대한 순위를 판별한다. 과거 침수가 일어났던 기간의 SNS, 뉴스, 블로그를 수집하고 수집된 데이터를 주제어 분석 API로 분석하여 사람들이 많이 언급한 순서대로 추출된 단어에 대한 순위를 살펴보고, 시스템은 재해 키워드로 분류 가능한 단어들을 재해 키워드 테이블에 추가한다. 출현 빈도 순위의 단어들을 주소화 할 수 있는 지역에 한하여 좌표로 변환하는 지오코딩을 수행하여 도로 침수 영역을 탐색하고 도로의 위험도를 높이는 과정으로 진행한다.

### 3.2 강수량 정보를 위한 기상 데이터

본 연구에서는 도로 침수에 대한 전체 도로를 분석하는 것은 현실적으로 한계가 있기 때문에 각 도로별로 과거에 침수되었던 도로를 중심으로 현재 실시간 강수량 정보와 과거 침수되었던 시기의 강수량 정보를 활용한다. 과거 침수되었던 도로의 정보는 기상청 및 기사를 통해 수집한 지역, 도로 침수 영역, 일시 그리고 강수량 등이 과거 침수이력 데이터에 기본적으로 수집되어 있어야 하며, 이를 기반으로 연구를 진행한다. 따라서 본 연구에서는 도로 침수가 발생했던 지역의 강수량을 모두 합산한 후 평균을 내는 과정을 통해 도로의 침수가 발생하는 강수량의 임계치를 설정한 수, 실시간으로 수집되는 15분 강수량이 도로 침수가 발생할 가능성이 있는 임계치 값 이상이 되었을 때 도로 침수가 발생할 수 있는 도로

로서 위험도를 높이는 데이터로 활용된다.

기상 상태 데이터는 실시간 기상 측정 및 연산, 저장, 표출 등 모든 과정을 자동으로 처리하는 데이터로써 기상청에서 제공하는 AWS 데이터를 활용한다[8]. 강수량은 15분 강수량, 60분 강수량, 6시간 강수량, 12시간 강수량 그리고 일일 강수량 정보가 활용될 수 있고, 이 외에 풍속, 지역 코드 등의 정보가 수집된다. 본 연구에서는 짧은 시간에 좁은 지역에 집중호우가 발생하는 국지성 호우와 같은 재해 상황을 판단하기 위해 가장 최신의 강우 정보를 확인할 수 있는 15분 강수량의 데이터를 주로 활용하고 있으며, 15분 강수량을 AWS 지역별로 데이터를 수집하여 데이터베이스에 저장하도록 한다. 그리고 비가 오지 않는 날씨는 강우량 정보를 수집 할 필요는 없기 때문에 30분 간격으로 강우 정보를 확인하도록 하고, 직전 수집한 강우 데이터와 현재 수집한 강우 데이터를 비교하여 강수량의 차이가 생기는 순간을 비가 내리기 시작하는 시점으로 판단하여 15분 강수량을 AWS 지역별로 수집하게 된다. 그리고 본 연구에서도 김은미 등[6,7]이 연구한 기법과 유사하게 역거리가중치 기법을 적용하여 확률 강수량 예측 모델로 이용하고 있다. 이렇게 계산된 예측 강수량은 도로가 침수되었던 과거 이력 정보에서 추출한 강수량 기준치와 비교하여 기준치 이상일 때 본 연구에서도 도로 침수 영역으로 분석되도록 하였다.

### 3.3 도로링크 데이터

다음은 침수도로의 영역을 검색하기 위한 도로링크 데이터의 수집방법이다. 도로 침수영역을 표출하기 위해 기존의 지도 위에 도로링크 데이터를 포함하여 도로 침수 영역의 라인과 영역을 표시할 수 있도록 정보를 분석하는 과정이 필요하다. 도로링크 데이터는 도로의 고도, 도로와 도로들로 나누어지는 영역의 노드 그리고 도로 라인뿐만 아니라 도로의 이동 방향 등과 같은 도로의 기본 적인 모든 정보를 담고 있는 데이터들이 수집되게 된다. 이러한 도로 데이터는 방대한 양의 정보이기는 하지만 정적인 데이터이기 때문에 한 번의 정보 수집과 분석 방법에 대한 기법만 적용하면 되기 때문에 어려운 문제는 아니다. 도로링크 데이터 자체는 실시간성을 가지고 있지 않으나 침수 영역에 대한 도로 정보를 실시간으로 표출시켜 주어야 하기 때문에 기존의 서버 플랫폼 보다는

분산처리 시스템 환경에서 링크 도로의 영역을 실시간으로 지도에 표출하는 것이 효율적이다. 이러한 도로링크 데이터는 상대적인 저지대의 도로 링크들을 검색하기 위해 도로의 각 고도의 값을 비교 계산하여 고도 값의 차가 음수 값이 나오는 도로를 주변의 고도보다 낮은 지역으로 판단하게 된다. 이러한 도로들 중에서 절대 값이 큰 값일수록 도로링크 중에서 침수 위험이 높은 도로로 판단하여 위험도를 상위에 오도록 설계하고 있다. 또 다른 고려는 도로의 경사도를 확인하여 경사가 급한 도로링크 일수록 침수 위험 우선순위를 높여주도록 설계한다. 한 개의 도로링크는 두 갈래로 나뉘는 길마다 노드로 구분되어 있고 각 노드는 좌표 값을 가지면서 세밀하게 나누어져 있다. 따라서 침수된 도로의 시작 부분과 끝 부분에 대해 설정하는 조건을 주는 것도 침수된 도로 영역의 중요한 데이터가 된다. 침수 도로로 위험도가 높게 판단된 도로링크는 전체 링크가 침수되었다고 가정하고, 수집된 정보들을 누적하여 계산하면서 위험도가 높게 되면 침수된 도로링크로 판단하는 기준에 도달한다면 침수된 도로 좌표 값을 총 계산하여 도로링크의 범위를 설정하도록 한다[6,7].

#### 4. 빅데이터 분석을 활용한 도로 침수 영역 탐색 시스템 개발

##### 4.1 빅데이터 분산 처리 플랫폼 설계

도로 침수 영역을 탐색하기 위한 빅데이터로서 소셜 미디어 데이터, 기상 상태 데이터 그리고 도로 링크 데이터를 수집 및 분석하는 과정을 3절에서 설명하였다. 데이터를 수집하고 분석하는 과정을 분산처리하기 위해 본 연구에서는 하둡이라는 플랫폼을 사용한다. 아파치 하둡은 분산 파일 시스템인 HDFS 플랫폼이 기본이며 데이터 파일을 HDFS 블록 단위(기본 64MB)로 각 노드에 분산하여 맵리듀스 태스크를 할당 처리하도록 이루어져 있다. 맵 태스크에서 할당하는 스케줄링, 합병 기능은 아파치 하둡에서 자동으로 처리된다[2,3,11]. 이러한 분산 처리 플랫폼에서 소셜 미디어 데이터는 다시 세 가지 영역으로 구분하여 분석하고, 기상 상태 데이터는 실시간으로 AWS 데이터를 제공받아 분석하여 실제 도로 링크 데이터에 부합하는 도로 침수 영역에 대한 결과를 표출하게 된다.

소셜 미디어 데이터는 Fig. 2에서 설명한 것과 같이 텍스트 영역, 이미지 영역, GPS 영역으로 구분한다. GPS 영역의 처리를 먼저 처리하여 침수 영역에 우선순위를 두고, 다음으로 이미지 영역의 EXiF 태그를 사용하여 이미지의 피사체가 찍힌 각도와 방향을 고려한 위치를 분석하게 된다. 그리고 텍스트 영역의 소셜 분석을 통해 침수 위치를 탐색하게 한다. 기상상태 데이터는 15분 강우량을 중심으로 데이터를 수집하고 과거에 침수한 이력과 비교하여 도로 침수가 일어났던 강우량보다 초과하였을 경우 도로 침수가 발생할 수 있다고 판단하여 도로 침수가 발생 가능한 영역으로 가중치를 두도록 하였다. 이는 기본적으로 과거에 도로침수가 발생한 경우 대부분의 지자체 등은 하수도 확장 등 적어도 침수가 발생했던 강우량으로는 침수가 발생하지 않도록 준비를 하고 있다는 관점에서 과거 정도의 강우량이 발생할 경우 확률빈도로 해당 도로 영역의 침수 발생 가능성을 높게 보고 가중치를 설정하도록 한 것이다. 도로 링크 데이터에는 과거에 침수가 된 도로의 정보를 포함하고 있으므로 과거 호우 시에 침수가 발생했던 도로는 침수가 상습적으로 일어나는 도로에 대한 조건을 확인할 수 있다. 상습 침수 도로로 분석된 도로는 강한 호우 시 집중 모니터링이 계속 되어야 하며 침수 도로로 확인되는 도로는 도로링크 데이터에서 위치가 확인되어 도로링크 부분이 표시된다.

##### 4.2 도로침수 영역 탐색 시스템 개발

본 절은 앞에서 설계한 다양한 빅데이터 수집과 분석 기법에 대한 도로침수 영역 탐색 시스템의 개발 내용을 설명한다. Table 1은 앞의 설계 기법을 기반으로 아파치 하둡 기반의 분산처리 플랫폼을 구현하기 위한 시스템 환경을 나타내고 있다. 하둡은 오픈소스로 제공되는 플랫폼으로써 우분투, 페도라, 센트OS 등과 같은 여러 운영체제를 통해 쉽게 적용할 수 있다. 본 시스템에서는 14.04.1 LTS 64bit를 사용하고 아파치 하둡 플랫폼은 안정화 문제 등을 이유로 1.0.0의 초기 버전을 사용하였다. 그리고 아파치 하둡은 자바를 기반으로 설계되어 맵리듀스 기능을 포함한 전반적인 프로그램은 자바 코드를 구성하였고, 익스포트 후 jar 파일을 아파치 하둡에서 컴파일하였다.

다음은 네트워크상에서 기상 상태 데이터를 데이터베이스로 받아올 때는 PHP 서버와 자바스크립트

Table 1. Environment of the System Implementation

Category	Environment
OS & Platform	<ul style="list-style-type: none"> <li>• Ubuntu 14.04.1 LTS 64bit</li> <li>• Hadoop 1.0.0</li> </ul>
DataBase, Server, Language	<ul style="list-style-type: none"> <li>• Oracle 11g, MongoDB</li> <li>• PHP</li> <li>• Javascript, HTML</li> </ul>
API	<ul style="list-style-type: none"> <li>• Twitter Open API</li> <li>• O2 Service Open API</li> <li>• Google Map API</li> </ul>
Library & Tool	<ul style="list-style-type: none"> <li>• EXIF Library</li> <li>• Eclipse Java EE IDE</li> </ul>

및 HTML을 활용하여 인터넷상에서 데이터를 획득한 후 데이터베이스에 저장하도록 하였다[12]. Fig 4는 기상청 페이지에서 AWS를 획득하는 과정을 나타내고 있다. Fig. 4의 (a)는 AWS의 각 지역별 분/시간별 강수량과 풍향, 풍속, 위치 정보 등이 수집되며, (b)와 (c)는 이러한 데이터들의 정보를 본 연구에서 활용하기 위한 PHP 서버와 Mongo DB의 수행 코드를 나타내고 있다.

데이터베이스는 도로링크 데이터의 정형 데이터를 처리하기 위해 오라클 데이터베이스를 사용하였고 반정형, 비정형 데이터를 관리하기 위해 NoSQL (Not only SQL)인 MongoDB를 사용하였다. 소셜 미디어 분석 중에서는 SNS에 대한 수집을 위해 트위터 API를 활용하였고, SNS 이외의 뉴스나 블로그와 같은 소셜 미디어를 수집 및 분석하기 위해 O2 Service OpenAPI를 사용하여 주제어 분석을 수행하였다. 실시간 소셜 미디어 서비스의 분석 결과는 도로 침수가 의심되는 지역의 주제어에 대해 순위를 결정하는 프로그램을 거쳐 우선순위가 높은 도로 침수 영역을 탐색한다.

Fig. 5는 본 연구의 설계 기준에 따라 도로 침수영역 탐색 시스템의 구현 예를 나타내고 있다. 구현된 시스템의 분석 결과로 도로 침수 영역에 대한 정보를 시각화하기 위해서는 지도에 표출 할 수 있어야 하기 때문에 본 연구에서는 구글 맵 API의 구글 지도를 활용하여 도로 링크 영역에 대한 데이터를 XML 형태로 변환하여 자바스크립트로 분석하고, 일치하는 도로 침수 영역을 탐색하여 표출하도록 구현하였다. 소셜 미디어 메뉴에서는 관리자 및 사용자가 실제로 블로그나 뉴스, SNS에 올라오는 정보를 가시적으로

판단할 수 있고, 관리자가 버튼을 클릭함으로써 간단하게 데이터를 수집 및 저장할 수 있도록 하였다. 사용자는 데이터를 임의로 컨트롤 할 수 없고 데이터를 읽을 수만 있도록 하였다. 따라서 관리자 메뉴에서 관리자로 로그인 할 경우에만 데이터 조작이 가능하도록 하였다.

구현된 시스템은 년도 또는 검색 기간을 선택하여 과거와 현재의 도로 침수 영역을 탐색할 수 있고, 재난 키워드를 선택함으로써 소셜 미디어에서 검색되는 결과를 도출할 수 있으며, 재난 키워드를 선택하지 않을 경우 모든 재난 키워드에 대한 검색이 실시되어 결과가 나타나도록 하였다. 도로 침수 영역 결과는 표와 함께 지도로 도출되어 사용자가 서비스 받기에 용이하도록 구성하였으며, 분석 결과에는 소셜 미디어 검색 결과와 현재 강우 상태를 그래프 형태로 나타내고 있다.

### 5. 결 론

최근의 주요 관심 분야는 재해 예방에 대한 연구와 재해 예방과 관련된 방대한 데이터의 분석을 위한 빅데이터 분석 기법들이 연구되고 있다. 본 연구에서는 자연재해에서도 태풍이나 게릴라성 집중 호우에 의한 도시 지역의 도로 침수 영역에 대한 탐색 기법에 대해 연구를 수행하였다. 이를 위해 본 연구에서는 두 단계의 과정으로 시스템을 구현하고 있다. 첫째는 도로 침수영역의 정보를 수집하기 위한 방법으로 SNS 데이터, 기상 상태 데이터, 도로링크 데이터를 수집하여 도로침수 영역의 탐색 기반 데이터로 활용할 수 있도록 설계하였다. 그리고 DSNS 데이터의 경우는 다시 텍스트 영역, 이미지 영역, GPS 영역으로 구분하여 향후 구현 시스템에서 도로 침수 영역의 위치와 실시간 정보를 제공할 수 있도록 설계하였다. 두 번째는 수집된 정보를 기반으로 하둡이라는 분산처리 플랫폼을 구현하여 사용자들에게 실시간 도로 침수영역을 화면에 표출하도록 구현하였다. 구현된 시스템은 여러 가지 기준과 재난 키워드로 실제 침수된 도로 영역을 탐색할 수 있는 기능과 분석 결과에 따른 현재 강우 상태 등을 그래프 형태로 나타내고 있다.

본 연구에서 제안한 빅데이터 기반의 도로침수 영역 탐색 시스템은 강수량을 중심으로 도로 침수 영역

AWS 지명 / 코드	강수량	강수량15	강수량30	강수량45	강수량60	강수량75	강수량90	기온	풍향1	풍속1	풍향10	풍속10	습도	현열기온	위치		
90 속초	18㎜	0.5	3.0	14.0	16.5	16.5	12.3	70.1	ENE	3.5	79.8	E	3.6	100.1	1021.9	강릉도 고성군 고성읍 풍동리	
92 양양(봉)	75㎜	-	1.0	4.0	21.5	25.5	25.5	11.1	80.0	E	3.0	80.0	E	3.1	1022.6	강릉도 양양군 속칭면 풍동리	
95 철원	154㎜	0.5	2.0	8.5	9.0	9.0	12.0	22.7	NNE	4.1	33.2	NNE	3.2	88	1018.2	강원도 철원군 철원읍 풍동리	
96 옥도	96㎜	0	0	0	0	0	0	15.5	222.0	SW	3.2	217.8	SW	3.5	85	1016.6	강원도 철원군 철원읍 풍동리
98 동두천	109㎜	0.3	1.6	5.7	6.0	6.0	12.6	19.4	NNE	2.4	32.9	NNE	1.9	83	1017.2	경기도 용두천시 영정동	
99 파주	29㎜	0.1	0.9	4.6	4.9	4.9	13.1	73.9	ENE	3.2	50.7	NE	4.0	85	1017.2	경기도 파주시 문성읍 풍동리	
100 대관령	773㎜	2.0	8.0	38.5	60.0	60.0	7.2	90.4	E	5.9	86.0	E	5.5	97	1020.5	(산간)강원도 영월군 대관령면 풍동리	
101 춘천	78㎜	0.6	2.7	12.9	14.5	14.5	12.7	32.7	NNE	0.9	12.2	NNE	1.3	95	1019.6	강원도 춘천시 풍동동	
102 백령	145㎜	0	0	0	0	0	0	16.7	39.8	NE	4.5	48.5	NE	4.6	63	1016.6	인천광역시 울진군 죽변면 풍동리
104 박간동	79㎜	1.5	5.0	24.5	44.5	44.5	11.2	53.9	NE	1.8	63.4	ENE	2.9	86	1021.5	강원도 강릉시 사천면 풍동리	
105 간동	26㎜	1.5	5.5	25.0	48.5	48.5	12.2	92.1	E	2.9	110.7	ESE	2.3	90	1021.5	강원도 강릉시 풍동동	
106 동해	40㎜	1.0	2.5	19.0	47.5	47.5	13.0	17.7	NNE	6.5	18.1	NNE	6.0	85	1020.1	강원도 동해시 풍동동	
108 서울	86㎜	0.5	2.0	9.5	15.5	16.0	13.3	61.0	ENE	4.6	64.0	ENE	4.3	97	1016.9	서울특별시 용두동 풍동동	

(a)

```

<!-- ajax로 ajaxFF -->
<script>
    $(function() {
        $.ajax({
            url: 'index.php',
            type: 'POST',
            data: {
                'aws_no': $('#aws_no').val(),
                'aws_title': $('#aws_title').val(),
                'altitude': $('#altitude').val(),
                'precipitation15': $('#precipitation15').val(),
                'precipitation60': $('#precipitation60').val(),
                'precipitation90': $('#precipitation90').val(),
                'precipitation12H': $('#precipitation12H').val(),
                'dayPrecipitation': $('#dayPrecipitation').val(),
                'address': $('#address').val()
            },
            success: function(data) {
                console.log(data);
            }
        });
    });

```

(b)

```

public function index()
{
    include("simple_html_dom.php");
    $m = new MongoClient();
    $db = $m->upsil;
    $collection = $db->data;
    $collection->remove();
    $collection = $db->createCollection("data");

    ...

    if($i==34){
        echo (int)substr($a[0],strpos($a[0],">")+1, strpos($a[0], "</a>")-strpos($a[0], ">")-1). "<br>";
        echo substr($a[1], strpos($a[1], ">")+1, strpos($a[1], "</a>")-strpos($a[1], ">")-1). "<br>";
        echo (int)substr($a[2], 0, strpos($a[2], "m")). "<br>";
        echo (float)$a[4]. " " . (float)$a[5]. " " . (float)$a[6]. " " . (float)$a[7]. " " . (float)$a[8]. " <br>";
        echo $a[18]. "<br>";

        $document = array(
            "aws_no" => (int)substr($a[0], strpos($a[0], ">")+1, strpos($a[0], "</a>")-strpos($a[0], ">")-1),
            //aws_title => substr($a[1], strpos($a[1], ">")+1, strpos($a[1], "</a>")-strpos($a[1], ">")-1),

            "altitude" => (int)substr($a[1], strpos($a[1], ">")+1, strpos($a[1], ">")-strpos($a[1], "<")),
            "precipitation15" => (float)$a[4],
            "precipitation60" => (float)$a[5],
            "precipitation90" => (float)$a[6],
            "precipitation12H" => (float)$a[7],
            "dayPrecipitation" => (float)$a[9]
        );
        //address" => $a[18]

        $collection->insert($document);
    }
}

```

(c)

Fig. 4. Collection of the AWS Data. (a) Page of the AWS Data (b) Page of the AWS Data Source Code, (c) Code using PHP Server and MongoDB.

을 탐색한 결과를 제시하고 있지만, 강수량이 아닌 인위적인 도로 붕괴 등의 원인으로 도로를 사용할 수 없는 경우에도 본 연구의 확장 기법으로 적용할 수 있을 것으로 판단된다. 다만, 본 연구는 제한적인

도시와 도로 중심으로 연구된 결과를 제공하고 있기 때문에 보다 확장된 정보수집과 결과를 표출하는 연구가 필요하며, 실시간 정보의 정확성을 높이는 연구도 필요하다.



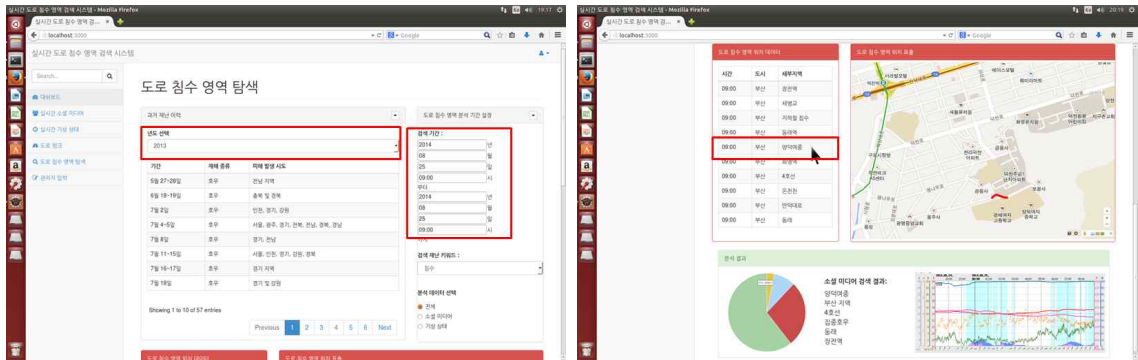


Fig. 5. Implementation Display of the Road Flooded Area Search.

REFERENCE

[ 1 ] M. Kim, S. Han, Y. Cui, H. Lee, and C.S. Jeong, "A Hadoop-based Multimedia Transcoding System for Processing Social Media in the PaaS Platform of SMCCSE," *Journal of Korean Society for Internet Information Transaction on Internet and Information Systems*, Vol. 6, No. 11, pp. 2827-2848, 2012.

[ 2 ] Y.I. Cho, "Understanding and major issues of the Big Data," *Journal of Korean Association for Regional Information Society*, Vol. 16, No. 3, pp. 43-65, 2013.

[ 3 ] H.J. Lee, "Application of Big Data Hadoop platform," *The Korean Institute of Communications and Information Sciences*, Vol. 29, No. 11, pp. 49-56, 2013.

[ 4 ] S.W. Seo, J.H. Kim, Y.S. Park, J.S. Lee, and J.S. Myeong, *The Hadoop&NoSQL for Mass Data Analysis and Processing*, GilBut Publishers, Korea, 2013.

[ 5 ] Ministry of Security and Public Administration, *2013 Safety Administration White Paper*, 11-1312000-000095-10, 2014.

[ 6 ] E. Kim, H.S. Hwang, and C.S. Kim, "A Study on the Development of the Flood Risk Index for Roads Considering Real-time Rainfall," *Journal of Korea Multimedia Society*, Vol. 16, No. 5, pp. 610-618, 2013.

[ 7 ] E. Kim, K.H. Rhee, and C.S. Kim, "A Study

on the Method of Development of Road Flood Risk Index by Estimation of Real-time Rainfall using the Coefficient of Correlation Weighting Method," *Journal of Korea Multimedia Society*, Vol. 17, No. 4, pp. 478-489, 2014.

[ 8 ] The National Weather Service Blog, [http://kma\\_131.blog.me/220044187761](http://kma_131.blog.me/220044187761) (accessed July 20, 2014)

[ 9 ] T.W. Seo, M.G. Park, and C.S. Kim, "Design and Implementation of the Extraction Mashup for Reported Disaster Information on SNSs," *Journal of Korea Multimedia Society*, Vol. 16, No. 11, pp. 1297-1304, 2013. 11.

[ 10 ] Y.H. Chun, *Development of a Decision Support System form Extraction Information and Determination Attributes of Disasters based on Social NetWork*, Master's Thesis of Pukyong National University, 2013.

[ 11 ] Dalseong Gun Disaster and Safety Countermeasures Headquarters Homepage, [http://www.dalseong.daegu.kr/jaehe/sub\\_menu/accident\\_public/accident\\_category.html](http://www.dalseong.daegu.kr/jaehe/sub_menu/accident_public/accident_category.html) (accessed Aug. 10, 2014)

[ 12 ] B.G. Gu, "Hadoop Map/Reduce Implementation Way on GPU Cluster," *Proceedings of KIIT Summer Conference*, pp. 49-52, 2012.

[ 13 ] S. Hur and Y. Kim, "A Design of Analysis System on TV Advertising Effect of Social Networking Using Hadoop," *Journal of*

*Internet Computing and Services*, Vol. 14,  
No. 6, pp. 49-57, 2013.



송 영 미

2013년 부경대학교 IT융합응용  
공학과 학사 졸업.

2015년 부경대학교 정보시스템  
협동과정 석사 졸업.

관심분야 : 방재IT, 재난 데이터  
빅데이터 분석, Hadoop  
을 활용한 빅데이터 분석,

소셜 네트워크 서비스(SNS) 수집 및 분석 등



김 창 수

1991년 중앙대학교 컴퓨터공학과  
박사

2006년~현재 유비쿼터스 부산  
도시협회 부회장

2006년~현재 (사)그레고리장학  
회 이사

2015년~현재 한국멀티미디어학회 부회장

1992년~현재 부경대학교 IT융합응용공학과 교수

관심분야 : 방재IT, UIS/GIS, 운영체제, 시멘틱 웹, 재난  
관리, 공간검색, 도시방재 등