

# 웹 스크래핑을 이용한 재난 상황 예측에 대한 연구

이동수, 오승민, Ashiquzzman AKM, 김상우, 김진술  
전남대학교 공과대학 전자컴퓨터공학과  
ready1819@gmail.com  
osm5252kr@gmail.com  
zamanashiq3@gmail.com  
swmax88@gmail.com  
jsworld@jnu.ac.kr

## A Research for Disaster Prediction using Web Scraping

Lee Dong Su, Oh Seung Min, Ashiquzzman AKM, Kim Sang Woo, Kim Jin Sul  
School of Electronics and Computer Engineering, Chonnam National University

### 요 약

유무선 통신 기술이 발달하면서 환경에 대한 감지와 제어를 수행하는 센서 네트워크 기술의 확산으로 홍수, 태풍, 지진 등의 재난에 대비하고자 재난 감시 시스템에 대한 연구가 활발히 진행되고 있다. 그러나 최근 우리 사회는 시대가 지남에 따라 자연 재해와 인적 재난의 피해는 그 정도가 여전히 증가하는 추세이다. 재난 상황 발생 시에 신속히 대응하고 방지하는 것보다 재난 상황에 대해 미리 예측하고 대비하는 것은 더욱 중요하다. 본 논문에서는 실내외 재난 상황 발생에 대하여 발생전후의 전반적인 피해를 예방하거나 최소화하기 위하여 사용자의 위치 정보와 서버의 기상 정보를 Node.js와 MySQL, 웹 스크래핑을 통해 구현하고 제어하여 사용자가 재난 상황에 대한 실시간 정보를 제공받을 수 있는 예측 시스템에 대해 구상하였고 크게 예측 시스템과 백업 시스템으로 구성된 재난 상황 예측 시스템을 제안한다.

키워드: Node.js, MySQL, Web Scraping, Disaster Prediction

### 1. 서 론

최근 우리 사회에서 화재, 태풍, 싱크홀, 지진, 가뭄, 폭설, 집중호우 등과 같은 재난 상황 발생에 쉽게 노출 되면서 그 문제점 또한 점점 심각해지고 있는 상황이다. 뿐만 아니라 재해재난이 발생했을 때에 이를 예측하고 예방할 수 있는 수단이 부족하다. 최근에는 폭염, 폭설, 폭우, 가뭄 등 이상기후에 의한 피해가 매해 최고치를 갱신하고 있는 추세이다. 이에 따라 본 논문에서는 Node.js와 MySQL, 웹 스크래핑을 이용한 재난 상황 예측과 피해 예방을 위한 알림 서비스에 대해 연구한 내용을 서술하고자 한다. 이 논문의 2장에서는 재난 상황과 웹 스크래핑에 관한 이론적 논의와 제안하고자 하는 시스템의 개요, 3장에서는 가까운 미래의 재난 상황을 예측하는 시스템의 구조와 관련된 연구 대해 소개하였고, 4장에서는 결론으로 향후 연구 계획에 대해 서술하였다.

## 2. 이론적 논의

### 2.1 재난 상황

재난 상황이란, 지진, 산사태, 화산 폭발, 홍수, 태풍과 같은 자연계의 이상 현상을 원인으로 발생한 재난 상황과 생물 테러, 화재, 유해 물질 유출, 핵 및 방사능 사고, 정전과 같은 인간의 욕심이나 실수로 인한 행동이 원인이 되는 재난 상황을 모두 포함한다. 본 논문에서는 날씨 등

의 자연현상의 변화 또는 인위적인 사고로 인한 인명이나 재산의 피해, 재난 가운데 자연현상과 관련된 천재지변 등의 재해 또는 재앙과 사람의 실수 또는 부주의나 고의로 어난 사고까지의 재난을 그 범위로 둔다.

### 2.2 웹 스크래핑

웹 스크래핑이란, 웹 수집 또는 웹 데이터 추출이라고도 하는데, 쉽게 말해 웹 사이트에서 데이터를 긁어 모으는 작업을 말한다. 기존에 사용되고 있는 오픈 소프트웨어로는, HTTP(Hypertext Transfer Protocol)을 사용하거나 웹 브라우저를 통해 직접 접속할 수 있으며, 사용자가 수동으로 수행할 수 있다. 일반적으로 로컬 데이터베이스나 스프레드시트를 통해 복사되면서 언제든지 검색하거나 분석할 수 있기 때문에 본 연구에서는 웹 스크래핑을 활용한 연구를 진행하였다.

웹 스크래핑은 주로 파이썬과 라이브러리를 이용하여 사용되는데 주로 Selenium, BeautifulSoup, Pandas를 이용한다. Selenium은 테스트 자동화 도구로서 정적, 동적 스크래핑이 모두 가능하다. 로그인이나 자바스크립트에 의한 동적 자료들을 불러오기 위해 사용한다. BeautifulSoup는 파싱(많은 데이터에서 사용자에게 필요한 특정한 데이터만 빼내는 작업)을 하면서 시간이 많이 걸리는 작업을 해당 라이브러리를 사용하여 자동화함으로써 시간을 단축하고 데이터가 정리되는 이중 효과가 있으며 Selenium으로

가져 온 데이터들 중에서 예측을 위한 데이터 값만 가져온다. Pandas는 Python Data Analysis Library의 약자로서, 데이터를 정렬할 수 있는 자료 구조이고 엑셀의 스프레드시트 형태라고 보면 된다. 가져온 데이터들을 CSV파일로 저장하여 관리, 재난 예측을 위한 기계학습 데이터로 사용하거나 대피시설 및 급수시설에 대한 정보를 저장한다.

```

In [8]: soup = BeautifulSoup(dreq,"html.parser")
        safe_shelter_list = soup.select("table > tbody > tr > td > a")
        driver.implicitly_wait(10)

In [10]: dataframe = pd.DataFrame(safe_shelter.columns=["new_address", "old_address", "indexrange(1, len(safe_shelter_list)+1)"])
        dataframe.to_csv(r"C:\Users\kdm\OneDrive\Documents\data_shelter.csv", encoding="utf8")

element = driver.find_element_by_id("sbLawArea1").send_keys("광주광역시")
element = driver.find_element_by_id("sbLawArea2").send_keys("북구")
element = driver.find_element_by_id("sbLawArea3").send_keys("용봉동")

search = driver.find_element_by_xpath("//a[@id='btnSearchOk']")
driver.execute_script("arguments[0].click();",search)
time.sleep(3)

for j in safe_shelter:
    print(j)

['광주광역시 북구 우치로 77-0 (용봉동)', '광주광역시 북구 용봉동 239번지 2호']
['광주광역시 북구 용주로30번길 16 (용봉동, 삼원맨션)', '광주광역시 북구 용봉동 1225번지 1호']
['광주광역시 북구 용봉역지로 16 (용봉동)', '광주광역시 북구 용봉동 1074번지 6호']
['광주광역시 북구 비엔날레로 112-5 (용봉동, 신동아아파트)', '광주광역시 북구 용봉동 850번지']
['광주광역시 북구 저불로 80 (용봉동)', '광주광역시 북구 용봉동 1355번지 1호']
['광주광역시 북구 저불로 21 (용봉동)', '광주광역시 북구 용봉동 1406번지 1호']
['광주광역시 북구 용주로30번길 60 (용봉동)', '광주광역시 북구 용봉동 1468번지']

new_address                                old_address
0 광주광역시 북구 우치로 77-0 (용봉동)      광주광역시 북구 용봉동 239번지 2호
1 광주광역시 북구 용주로30번길 16 (용봉동, 삼원맨션)  광주광역시 북구 용봉동 1225번지 1호
2 광주광역시 북구 용봉역지로 16 (용봉동)          광주광역시 북구 용봉동 1074번지 6호
3 광주광역시 북구 비엔날레로 112-5 (용봉동, 신동아아파트)  광주광역시 북구 용봉동 850번지
4 광주광역시 북구 저불로 80 (용봉동)              광주광역시 북구 용봉동 1355번지 1호
5 광주광역시 북구 저불로 21 (용봉동)              광주광역시 북구 용봉동 1406번지 1호
6 광주광역시 북구 용주로30번길 60 (용봉동)        광주광역시 북구 용봉동 1468번지

```

(그림 1) 재난발생시 근처 대피시설의 위치(국민안전포털)

### 2.3 제안된 시스템 개요

클라이언트는 항상 서버와의 연결로 시작하고 서버는 항상 클라이언트의 요청을 기다린다. 예를 들면, 클라이언트가 서버의 변경에 영향을 받지 않고 인식되지 않는 상태에서 서버를 교체, 복구, 업그레이드, 재배포하는 것이 가능하다. 모든 데이터는 서버에 저장되며, 서버는 접근과 자원을 효과적으로 제어하여 적절한 권한을 가진 클라이언트만이 데이터에 접근하고 변경할 수 있도록 보장한다.

본 시스템의 구성 요소는 크게 예측시스템과 백업시스템으로 구성하였고 가까운 미래에 다가 올 재난 상황에 대한 알림 및 경고 서비스를 서버에서 예측하여 사용자에게 실시간으로 전달하는 역할을 주로 담당한다.

## 3. 관련 연구

### 3.1 예측시스템

예측시스템은 서버와 클라이언트로 나뉘는데, 클라이언트와 서버는 항상 연결이 되어 있어야 하며 서버는 클라이언트에게 메타데이터를 제공하는 방식이다. 서버는 기상청과 센서로부터 받은 데이터를 서버의 데이터베이스에 저장하고 Socket.IO 통신을 통해 사용자의 요청을 받는다. XML파일로 기상청 날씨 데이터가 있는 RSS파일에 대한 정보를 가져와 데이터베이스 저장하고, HTTP GET

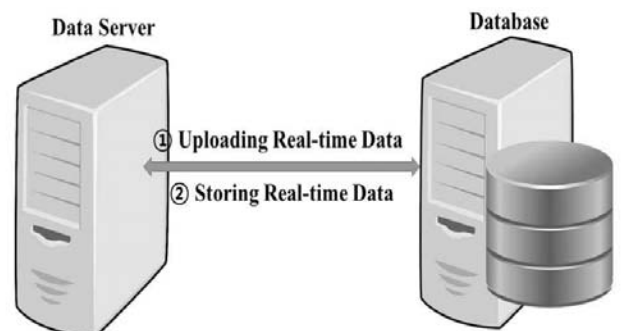
method를 이용하여 서버에서 정보를 검색한 후, MySQL 데이터베이스 서버는 유니코드를 지원하지 않기 때문에 모든 문자를 Hex를 이용하여 인코딩한다. 인코딩한 문자열을 Socket.IO 통신으로 도시와 구역, 날씨에 대한 정보를 사용자로부터 요청받아 날씨와 지역에 대한 데이터 분류화를 모두 처리한다. 한편, 서버가 필요한 정보가 있는 웹 사이트의 HTML 데이터에서 사용자에게 필요로 하는 데이터를 추출, 가공하여 저장하고 실시간으로 사용자의 위치 정보와 위치한 장소의 날씨 데이터(온도, 습도, 풍향, 풍속)에 대한 양적 자료가 담긴 기상 정보를 추출하여 가까운 미래에 다가 올 재난 상황에 대한 경고메시지를 전달해 줌과 동시에 사용자가 재난 발생 지역 근방에 도달하기만 하여도 경고 메시지로 알려준다. 또한, 클라이언트에서도 마찬가지로 서버에게 재난 발생 시 해당 지역의 상황을 알려줄 수 있다.



(그림 2) 실시간 예측 정보를 받는 클라이언트

### 3.2 백업시스템

재난 상황에서의 백업시스템은 재난 상황 발생 시 백업 솔루션으로 정보를 일괄적으로 데이터화 및 저장한다. 백업시스템에서 위치 정보에 의한 기상 정보는 기온과 강수량에 근거한 예측 데이터로 재난 상황 발생이나 재난 상황 예측을 위한 백업 솔루션뿐만 아니라 다양한 분야에서도 유용한 자료로 쓰인다. 구체적으로는, 사용자가 서버에게 보내는 데이터에 대해 스토리지 서버에 업로드를 하면서 EJS(Embedded Javascript)와 Node.js를 사용한다. 모든 데이터를 EJS로 HTML을 채운 다음, HTTP GET/POST 방식을 통해 서버로 전송 후 서버는 파일의 정보를 데이터베이스 및 파일 시스템 서버에 저장한다. MySQL 데이터베이스 서버에서도 역시, 유니코드를 지원하지 않기 때문에 모든 문자를 HEX로 인코딩한다.



(그림 3) 재난 상황 예측 데이터 서버와 데이터베이스

#### 4. 결론

본 논문에서는 예측시스템과 백업시스템에서 실내외 재난 상황 발생에 대하여 발생전후의 전반적인 피해를 예방하거나 최소화하기 위하여 사용자의 위치 정보와 서버의 기상 정보를 Node.js와 MySQL, 웹 스크래핑을 통해 구축 및 제어하여 사용자가 재난 상황에 대한 정보를 실시간으로 제공받아 재난 상황에 대해 대피하고 예방할 수 있도록 도와주는 재난 상황 예측 시스템에 대해 제안하였다.

웹 스크래핑을 이용한 재난 상황 예측에 대한 연구를 통해 사용자가 자신의 위치 정보와 해당 시간에 따른 기상 정보를 기반으로 가깝게 다가 올 재난 상황을 예측하여 재난 지역에 접근 했을 시 서버로부터 경고 알림과 동시에 가까운 대피소와 그곳을 향하는 대피로에 대한 정보를 사용자가 받을 수 있도록 개발하였다. 향후에는 재난 상황 발생 지역 근처에 위치한 사용자들에게 다가올 재난 상황에 대한 경고 알림 서비스뿐만 아니라 웹 스크래핑을 통한 지난 몇 년 간의 기상 및 재난 상황들에 대한 히스토리를 모아다가 다가올 재난 상황 발생 시에 습도, 온도, 일조량, 풍향, 풍속, 강수량, 홍수, 물난리 재난에 대한 데이터 기록들을 기계학습에 적용 후, 웹이나 앱을 통한 피해 예상 반경 1km 내에 사용자들이 경고 메시지가 담긴 푸시 알림 서비스 및 실시간 재난 상황 현장의 사진이나 영상을 제공하는 서비스를 연구할 예정이다. 시간 정보에 의한 공간 정보 기반의 기후변화에 의한 다양한 피해를 빅 데이터 기법과 융합 데이터베이스를 활용하여 산업계의 사전적 피해 최소화 및 예방을 위한 재난 상황 예측 및 대응 체계 구축 방안에 대한 연구가 활발히 이루어지길 기대한다.

#### 감사의 글

본 논문은 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음. (IITP-2019-2016-0-00314).

#### 참 고 문 헌

- [1] 정영진, 류근호, 김학철. “실외 센서네트워크 기반 재해방지 시스템을 위한 위험지역 예측기법”, 『정보처리학회논문지D』 제13권 제6호, 2006. pp.775-788.
- [2] 허준영, 이주호. “미래 초대형 중대재난 예측 시나리오 탐색 연구 - 풍수해 재난 사례를 중심으로 -”, 『Crisisonomy』 제14권 제6호, 2018. pp.29-40.
- [3] 이병훈, 이병진, 오승희, 이용태, 김경석. “복합재난 예측 모형 설계를 위한 공통 입출력 파라미터 도출 연구”, 『한국위성정보통신학회논문지』 제12권 제4호, 2017. pp.34-41.
- [4] 박해경, 이동근. “기계학습과 시스템 다이내믹스를 이용한 재난예측 및 대응정책효과 시뮬레이션 경기도 단기

가뭄을 사례로”, 『한국방재학회논문집』 제19권 제1호, 2019. pp.45-53.

[5] 이동규. “재난관리 예측적 거버넌스 시스템 구축을 위한 시론적 검토 - 미래예측적 이상신호 감지를 위한 협력적 재난관리 의사결정 시스템 제언을 중심으로 -”, 『Crisisonomy』 제12권 제2호, 2016. pp.35-52.

[6] 정운철, 윤홍식, 정명훈, 권정환. “재난별 피해예측결과와 연계하기 위한 건축물 인벤토리 구축 방안에 대한 연구”, 『한국방재학회논문집』 제18권 제3호, 2018. pp.117-123.

[7] 정영철, 배용근 “u-Gov의 재난 대응체계를 위해 UT를 응용한 예측 프로토타입과 정책적과제 연구”, 『한국컴퓨터정보학회논문지』 제15권 제10호, 2010. pp.173-182.

[8] 이병진, 이병훈, 오승희, 이용태, 김경석 “주요 시설물 피해를 고려한 복합재난 피해 예측 시스템 방안 연구”, 『한국위성정보통신학회논문지』 제12권 제4호, 2017. pp.18-25.

[9] 오영록, 정건희. “재난통계를 활용한 대설피해 예측 및 대설 피해 적실심 기준 결정 방안”, 『대학토목학회논문집(국문)』 제37권 제2호, 2017. pp.325-331.

[10] 구남경, 이강환. “재난 재해 지역의 산불 확산경로와 이동속도 예측 알고리즘”, 『한국정보통신학회논문지』 제20권 제8호, 2016. pp.1581-1586.