

그린뉴딜

정책에 따른

ITS의 확대 추진 및

고도화

2020. 11. 19(목) - 20(금)

제주한라대학교 한라컨벤션센터



• 논문요약본 제출 | 2020년 9월 9일(수) - 2020년 11월 6일(금)



• Full Paper 제출 | 2020년 9월 9일(수) - 2020년 11월 6일(금)



• 사전등록 기간 | 2020년 9월 9일(수) - 2020년 11월 6일(금)

■ 학술발표 Session II : (20일 11:00-12:20)

SESSION 2B – 자율주행 1		발표장 :
좌장 :		
번호	발표 주제	발표 및 제출자
1	자율주행차량 안전 운행을 위한 파발원지 검출 시스템	최윤선, 최동수, 백민석, 우종명
2	자율주행차 학습에 고려되어야 하는 위험환경 조건	Ahmed Khaqan, KIM Jin-Tae
3	가상환경기반 부분자율주행 운전능력평가를 위한 시뮬레이터 개발	김승학, 김연아, 문철
4	자율주행 윤리 가이드라인 및 딜레마 시나리오 개발	김지원, 정하림, 박상민, 이준섭, 송봉섭, 윤일수
5	자율주행차의 라이다 데이터를 이용한 고속도로 밀도 추정 연구	박성호, 이영석, 정창현, 이에쁜, 윤일수
SESSION 2C – 교통 빅데이터 1		발표장 :
좌장 :		
번호	발표 주제	발표 및 제출자
1	랜덤포레스트 기반 지하철 통행 경로 추정	윤현수, 이은학, 김동규
2	도시빅데이터를 활용한 교통 측면의 도시건강 평가지표 개발 연구	박병훈, 홍정열, 박동주
3	교통약자와 일반승객을 고려한 다목적 Skip-Stop 전략 설계	한상욱, 이은학, 김동규
4	광역급행형 시내버스 노선 신설 타당성 평가체계 개발에 관한 연구	고승렬, 박준식
5	코로나19가 시외버스 운송사업에 미치는 영향	문지혜, 박준식
6	GRU를 이용한 서울시 지하철 구간별 혼잡도 예측 모델 연구	권효승, 최창준, 정효석, 송재인, 강민희, 황기연
SESSION 2D – 교통/ITS 정책 및 기타 1		발표장 :
좌장 :		
번호	발표 주제	발표 및 제출자
1	도로자산의 가치평가방법에 관한 연구	도명식, 박성환, 박희성, 송현엽
2	혼잡통행료 징수를 위한 시공간 범위 도출에 관한 연구	김민정, 김아름, 우보람, 정대한, 김회경
3	효율적인 교통 정보 관리를 위한 ELK스택 기반의 행동 패턴 시각화 시스템	김효중, 임해동, 이경민, 인치호
4	정적 및 동적 트래픽 라우팅에서의 네트워크 성능 비교에 관한 연구	Hussain Shahid, KIM Jin-Tae
5	VISSIM을 이용한 터보 회전교차로 효과분석 연구	송민수, 이동민
6	경부고속도로 개통이 국내 사회경제 및 수송에 미친 영향	유다영, 홍정열, 박병훈, 최윤희, 탁지훈, 손의영, 박동주

효율적인 교통량 정보 제공을 위한 ELK스택 기반의 행동 패턴 시각화 시스템

ELK Stack-Based Behavior Pattern Visualization System
for Efficient Traffic Information

김효중, 이경민, 임해동, 인치호
세명대학교 컴퓨터학부

Key Words : ELK스택(ELK Stack), 엘라스틱서치(Elasticsearch), 키바나(Kibana), 시각화(Visualize)

목 차

- I. 서 론
- II. 교통량 정보 시각화 시스템의 설계
- III. 교통량 정보 시각화 시스템의 구현
- IV. 결 론

I. 서 론

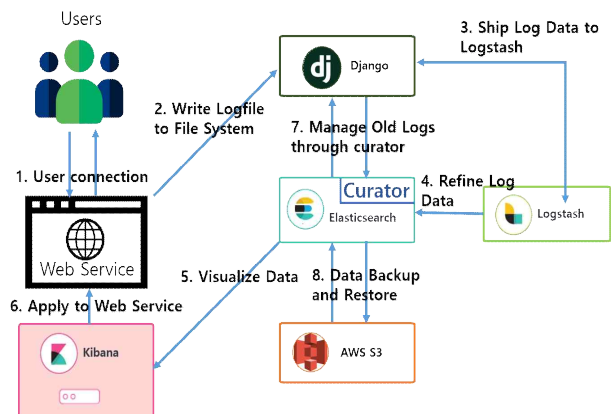
최근 코로나 19(COVID-19) 재확산의 영향으로 전자상거래는 최고의 전성기를 맞이하고 있다. 방역 강화 조치로 외출이 제한된 소비자들의 식자재나 생필품 등 온라인을 통한 물품 구매가 크게 늘었다.[1] 이와 함께 최근 전자상거래 시장도 말 빠른 성장을 보이는데, 최근 여러 쇼핑몰에서 소비자의 메타데이터를 추출하여 사용자의 취향 및 관심사를 분석하고, 메인화면의 상품을 개인별 맞춤형으로 구성하는 등 소비자의 데이터를 적극적으로 활용하는 경우가 증가하고 있다.[2] 그러나 기존 활용 방안들은 관리자의 관점에서 데이터를 관리하는 특성상 소비자 관점에서 상품을 추천해 주는 이유와 그에 따른 소프트웨어 구조, 작동원리 등을 알 수 없는 블랙박스 형태로 이루어지기 때문에 그에 따른 데이터를 시각화할 필요성이 있다. 따라서 본 논문에서는 소비자의 검색패턴과 구매 정보 데이터를 시각화하여 소비자가 본인의 데이터 정보를 더욱더 쉽게 볼 수 있게 UI 형태로 전달하는 것을 목적으로 한다.

ELK스택은 분산검색 엔진인 엘라스틱서치, 데이터 전처리 플랫폼인 로그스태시(Logstash), 이를 시각화하는 키바나의 집합이다. 사용자의 로그를 분석하려면 대량의 로그를 실시간으로 집계해 분석해야 하고 중간 매개자 없이 온라인으로 직접 데이터 접근이 가능해야 한다. 따라서 사용자가 직접 원천 데이터에 접근해 필요한 관점에 따라 데이터 분석이 가능한 Online Analytical Processing(OLAP)[3] 방식을 기반으로 실시간으로 입력되는 로그를 수집하고 분석이 가능한 ELK 스택 이용하여 시각화 시스템을 개발하고자 한다.

II. 교통량 정보 행동 패턴 시각화 시스템의 설계

본 논문에서 제안하는 보행자 교통사고 행동 패턴 시각화 서비스 시스템 구성도는 <그림 1>과 같으며, 기본동작 순서는 다음과 같다.

먼저 서비스 사용자가 웹서비스에 방문하여 서비스를 이용한다. 이때 사용자의 검색 기록과 모든 행동 기록, 구매 기록 등을 포함한 웹서비스의 로그가 장고(Django) 가상환경 서버에 저장된다. 다음으로 가상환경 서버에 저장된 로그들은 로그스태시에 전달하고 로그스태시는 로그를 받아서 로그 레코드를 정제해 엘라스틱서치에 저장한다. 마지막으로 키바나에서 엘라스틱서치에 저장된 로그들을 시각화하고 분석하여 웹 페이지로 전달 후 UI 형태로 나타내도록 하였다.



<그림 2> 보행자 교통사고 행동 패턴 시각화 시스템 구성도

데이터 과부하를 해결하기 위해 큐레이터(Curator)를 사용하여 설정한 기간, 디스크 크기를 기준으로 오래된 로그를 자동으로 삭제한다. 하지만 지워버린 로그를 다시 확인해야 하는 상황을 대비하여 모든 로그를 AWS S3에 백업시켜 큐레이터가 엘라스틱서치에서 로그를 지우더라도 언제나 AWS S3를 통해 복원 및 분석을 할 수 있도록 한다. 마지막으로 키바나를 통해 엘라스틱서치의 데이터를 시각화하고, 시각화한 모습은 변환된 데이터를 바탕으로 웹 서비스에 표출하여 소비자의 검색패턴, 구매패턴 등을 분석한다.

III. 보행자 교통사고 행동 패턴 시각화 시스템의 구현

본 논문에서는 <그림 1>의 구조를 바탕으로 웹 페이지를 직접 구현하여 사용자가 나의 소비 패턴을 볼 수 있는 UI를 제작하였으며 테스트 시나리오 제원은 <표 1>과 같다.

<표 1> 테스트 시나리오 제원

구 조	기 능	기술명
Web Framework	MVT Pattern	Django
Module Package	Data Analysis & Visualization	Django-elasticsearch-dsl
	Korean Morpheme Analyze	Nori
	Run Localhost IP Server	virtualenv(http://127.0.0.1:8000/)
	Log Version Control	Curator
Web Service	Data Storage&Backup	AWS S3

1. 기본 설정 및 데이터 시각화 구현

제안하는 사용자 행동 패턴 시각화 UI 시스템 기본 설정 순서는 다음과 같다. 첫 번째 장고 패키지 관리자 시스템에서 엘라스틱서치를 사용하기 위한 Django-elasticsearch-dsl, 가상환경 웹페이지를 위한 virtualenv, 한글 형태소 분석을 하기 위한 Nori 모듈을 설치한다. 두 번째, 장고에 로그가 기록되게 하기 위해서 settings.py 파일에서 TCP 네트워크 소켓을 통해 로그스태시로 로그 레코드를 전달하도록 설정한다. 세 번째, 로그 레코드를 받은 후 엘라스틱에서 제공하는 grok 필터를 사용하여 메소드와 검색어를 추출하고 퍼센트어로 되어 있는 검색어를 디코딩 한다. 그리고 가공된 데이터는 엘라스틱서치로 전달하도록 설정하고 전달된 로그들은 'logstash-yyyy.mm.dd'로 저장된다. 네 번째, 엘라스틱서치에 전달된 데이터로 키바나에서 설정한 시간인 지난 30일간 어떤 단어가 많이 검색되었는지, 어떤 단어를 가진 물품을 구매

했는지 분석하고, 한글 형태소 분석기인 Nori로 Json파일을 읽은 후 한글로 변환한다. 마지막으로, 변환된 데이터를 MVT 패턴 분석 기능을 적용한 웹프레임워크를 기반으로 그래프, 대시보드, 캔버스, 태그 클라우드 등 다양한 형태로 UI를 설계하여 표현한다.

2. 오래된 로그 자동관리 및 복원

서비스를 장기간 운영하면 많은 과거의 로그가 디스크 공간을 차지하게 된다. 설정한 기간, 디스크 크기 기준으로 오래된 로그를 삭제하는 프로그램인 큐레이터를 사용하여 이용률이 떨어지는 과거의 로그들을 자동으로 관리할 수 있도록 하였다. 먼저 로컬의 엘라스틱서치에서 'logstash'로 시작하는 인덱스가 30일이 지났으면 삭제하도록 두 파일을 설정하고 큐레이터를 실행한다. 그러면 큐레이터는 엘라스틱서치에 30일이 지나간 로그가 있는지 확인한 후 로그가 존재한다면 삭제한다.

저장소의 효율성을 위해 데이터를 삭제했다고 하더라도 기본 데이터에 문제가 발생하였을 경우 복구할 수 있어야 한다. 우선 엘라스틱서치에서 로그 데이터를 AWS S3에 간편히 백업할 수 있도록 repository-s3 플러그인을 설치한다. 그 다음 로그를 백업할 저장소를 지정한 후 엘라스틱서치에서 제공하는 데이터를 버전별로 저장하여 이전 상태를 유지해 주는 snapshot 기능을 통해 S3에 로그를 백업한다.



<그림 2> AWS S3를 통한 데이터 복원 테스트

데이터 복원 실험을 위해 7일간 로그 데이터에서 12시간의 로그 데이터를 삭제 후 복원할 snapshot을 지정하여 복원했으며, <그림 2>와 같이 정상적으로 데이터 복원이 된 것을 확인할 수 있다.

3. 시각화 UI 표현

엘라스틱서치에서 최종적으로 키바나에게 시각화 데이터를 전달한 후 소비자의 개인 페이지에 소비자 개인페이지의 행동 패턴 데이터를 UI 형태로 나타내도록 하였다. UI의 자세한 설명은 해당 그림의 ‘+’단추를 클릭 시 확대하여 볼 수 있고, 오른쪽 사슬 모양 단추를 클릭 시 해당 자료 URL로 접속되게 설계하였다. 구현된 시스템의 데이터 분석 시스템 UI는 다음 <그림 3>과 같이 구성하였다.



<그림 3> 구현 시스템의 데이터 분석 시각화 UI

또한, 이를 다양한 형태로 시각화 하였는데, 막대 그래프 형태(왼쪽 위)는 검색 카테고리의 키워드와 그에 따른 가격을 나타내도록 하였고, 도넛형 그래프(가운데 위)는 소비자의 구매 카테고리 그 비율을 나타내도록 하였으며, 캔버스(오른쪽)는 소비자의 구매 항목을 한눈에 보기 쉽게 캔버스 형태로 나타내도록 하였다. 또한, 태그 클라우드(가운데 아래)는 텍스트데이터로 표출하며, 매트릭 집계(수, 평균, 합 등)를 기반으로 글자 크기나 색상을 통하여 강조된다. 마지막으로 연결그래프(왼쪽 아래)는 소비자와 각종 카테고리와의 연결성을 나타낸다.

IV. 결 론

본 논문에서는 ELK스택을 이용한 보행자 교통사고 분석 웹서비스 시각화 시스템설계를 제안하였고, 설계한 시스템을 실제 웹서비스에 테스트하였으며 시각화 기능을 통해 본안과 소비습관이나 관심 카테고리를 매우 편리하게 확인할 수 있는 시스템을 구현하였다.

웹 구현을 위해 MVT 패턴을 기반으로 제작된 장고 프레임워크를 사용하였고, 패키지모듈로는 Django-elasticsearch-dsl을 사용하여 로그 분석 및 시각화를 하였으며, 시각화 데이터를 웹 UI로 나타내도록 하였다. 또한, 한글 지원을 위한 한글 형태소 분석기 Nori 모듈, 웹 구동 테스트를 위한 가상환경 웹 모듈 virtualenv, 오래된 로그 자동관리를 위한 큐레이터를 이용하였으며, 삭제된 로그 저장과 백업 기능을 위해 AWS S3 웹 서비스를 기반으로 개발하였다.

※ 이 논문은 2020학년도 세명대학교 대학혁신지원 사업에 의한 연구임

참고문헌

1. Jain, Neeraj Kumar, R. K. Saini, and Preeti Mittal. "A review on traffic monitoring system techniques." Soft Computing: Theories and Applications. Springer, Singapore, 2019. 569-577.
2. Mujaddid, Zahid, and S. T. Husni Thamrin. "Performance Comparison of Search Feature With Django Querysets and Elasticsearch in a Web Application." Diss. Universitas Muhammadiyah Surakarta, (2020): 1-14.
3. Zhan, Chaoqun, et al. "Analyticdb: Real-time olap database system at alibaba cloud." Proceedings of the VLDB Endowment 12.12 (2019): 2059-2070.
4. Bajer, Marcin. "Building an IoT data hub with Elasticsearch, Logstash and Kibana." 2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW). IEEE, (2017): 63-68.
5. Kamuthurai, Ahilaa Rani. "A Comparative Study of Self Hosted Elasticsearch vs AWS Elasticsearch Service." California State University, Northridge, (2020):.1-44.