

1차년도 진행 결과 개요

사고 데이터 분석을 위한
데이터베이스화 작업

텍스트 분석을 위한
온톨로지 구축

분석된 결과를 기반으로
GPS좌표 도출

세부 연구 목표

보험사고 데이터에 대한
분석 및 정리

비정형 데이터에 대한
형태소 분석 및 색인어 추출

비정형 데이터에 대한
형태소 분석 및 색인어 추출

사고 발생 지역에 따라
분류하여 데이터 베이스화

사고 지점을 정확히
도출할 수 있는 온톨로지 구축

사고 지점을 정확히
도출할 수 있는 온톨로지 구축

연구 개발 수행 내용

분석에 사용된 사고 데이터에 대한
정의를 체계화함

고속도로와 관련된 텍스트에 대해
형태소 분석을 수행하고
중요성이 높은 색인어를 추출함

우편번호 데이터를 매쉬업하여
사고에 대한 범위를 지정함

보험 개발원의 비정형 데이터를 분석하여
사고 발생 지역 별로 데이터를 분류함

사고 지점에 대한 오차율을 줄이기 위해
규칙성 있는 온톨로지를 구축함

온톨로지를 이용하여 텍스트를 분석하고
리턴된 POI키워드를 구글 API에 요청하여
좌표 데이터를 생성함

연구 결과

데이터를 특성에 맞게 분류하여
하나의 컬럼에 저장함

추출된 색인어를 기반으로 POI point가
될 수 있는 키워드를 정리함

대략적인 사고범위가 도출됨

국도와 고속도로 등 데이터가 분류되어
데이터 베이스에 저장됨

정확한 사고 지점을
리턴하는 온톨로지 구축

사고 발생 지점에 대한
GPS좌표 데이터를 생성함

문장 정제작업을 통해 추출되는 과정 및 데이터 결과물

정제 내역 및
필터링 된 문장 개수

필터링 후 추출된 문장의 예시

1
단
계

고속도로 키워드
(237,876개)

- 1) “고속도로” 자차 추돌사고발생
- 2) 경찰차를 피하다가 “순환도로” 근처에서 뒤차와 충돌, 앞문유리파손

2
단
계

고속도로 키워드 + 고속도로 명칭
(17,234개)

- 1) “남부” “순환도로” 주행 중 승용차가 끼어들어 우리차 접촉
- 2) 현출요청, 자차 “남부” “순환도로” 에서 “시흥” “고속도로” 합류지점에서 앞서가던 상대차량이 합류

3
단
계

고속도로 키워드 + 고속도로 명칭
+POI키워드
(11,378개)

- 1) 자차 “국도” 에서 “서해안” “고속도로” “부안” “IC” 진입 후 주행 중 본넷이 잠겨있지 않아 본넷이 꺾임, “서해안” “고속도로” “행담” “휴게소” 근처 현출요청
- 2) 자차 “남부” “순환도로” 에서 “시흥” “대로” 로 “램프구간” “시흥” “대로” “합류지점” 에서 앞서가던 차량이 합류하여 접촉, “1차로” 직진 중 브레이크가 듣지 않아 “가드레일” 2차 접촉

4
단
계

고속도로 키워드 + 고속도로 명칭
+POI키워드 + 방향
(1,344개)

- 1) “서해안” “고속도로” 조난 “분기점” “하행선” “부근” 에서 주행 중 자차가 “차선” 을 밟고 미끄러져
- 2) “중부” “고속도로” “대소” “부근” 에서 공사 중인 앞차와의 충돌 후 “역방향” 차량과 2차 충돌

5
단
계

고속도로 키워드 + 고속도로 명칭
+POI키워드 + 방향 + 거리
(187개)

- 1) “호남” “고속도로” “하행선” “125KM” “5차로” 전방차량정체로 정차 중인 대차의 후미를 자차가 밀리면서 추돌하고 “가드레일” 을 받아버린 사고
- 2) “경부” “고속도로” “상행” “363k” “대전” “방향” “2차로” 주행 중 빙판에 미끄러짐