

빅데이터 파이널 프로젝트 제안서

Reference

Paper/Report

웹 사이언스 : 소셜 빅데이터 마이닝 기반 이슈 분석보고서 자동 생성 (2014) ([링크/파일](#))

- 본 논문은 지금까지의 소셜미디어 분석과 분석보고서 생성의 세 가지 문제점을 해결하기 위해서 소셜 빅데이터 마이닝에 기반한 이슈분석보고서 자동 생성 시스템을 제안
- 세 가지 문제점은 분석의 고립성, 전문가의 주관성과 고비용에 기인한 정보의 폐쇄성이다. 시스템은 자연언어 질의분석, 이슈분석, 소셜 빅데이터 분석, 소셜 빅데이터 상관성분석과 자동 보고서 생성으로 구성된다.
- 생성된 보고서의 유용성을 평가하기 위해, 본 논문에서는 리커트척도를 사용하였고, 빅데이터 분석 전문가 2명이 평가하였다. 평가결과는 리커트 척도 평가에서 보고서의 품질이 비교적 유용하고 신뢰할 수 있는 것으로 평가
- 보고서 생성의 저비용, 소셜 빅데이터의 상관성 분석과 소셜 빅데이터 분석의 객관성 때문에, 제안된 시스템이 소셜 빅데이터 분석의 대중화를 선도할 것으로 기대

사회적 이슈 리스크 유형 분류를 위한 어휘 자질 선별 (2016) ([링크/파일](#))

- 소셜미디어에 나타난 다양한 이슈 중 여론 형성에 악영향을 끼치는 부정적 사건을 '리스크'로 정의, 그 세부 유형을 자동으로 분류하는 모델을 개발. 이를 위해 소셜미디어에 나타난 다양한 어휘 자질을 선별, 그 효과를 규명
- 특히 리스크 문장의 어휘 구문 특징을 표현하기 위한 자질로 워드 임베딩 학습 결과를 활용.
- 개별 어휘 자질의 특징을 분석하기 위해 언어분석 오류를 보정한 환경에서 수행한 실험 결과, 가장 효과가 큰 자질은 개체명 자질로 분석되었으며, 기본 어휘 자질을 기반으로 주요 술부의 워드 임베딩 결과와 워드 클러스터 결과를 모두 조합한 경우가 최고 성능을 보이는 것으로 파악.
- 실제 소셜빅데이터에 적용하는 환경과 유사하도록 자동 언어분석 결과의 오류를 포함한 조건에서 실험한 결과, 고빈도 평가셋에서는 92.08%의 성능을, 전체 58개 범주 평가셋에서는 85.84%의 성능을 얻었다.

소셜 빅데이터 기반 사회적 이슈 리스크 유형 분류 (2016) ([링크/파일](#))

- 소셜미디어의 정치사회적인 활용도가 높아짐에 따라 소셜빅데이터 기반 온라인 동향분석 및 모니터링 기술에 대한 수요 역시 급증
- 본 논문에서는 이러한 요구에 부합, 특히 여론 형성의 악영향을 끼치는 부정적 이슈 탐지를 위해 사회적으로 파장이 큰 이슈 중 공공여론이 부정적으로 형성될 이슈를 '리스크'로 정의하고 세부 유형을 분류
- 리스크 유형 정의를 위해 뉴스 문서집합을 대상으로 전수조사를 실시하였으며, 이슈 분야 즉 도메인별 특성을 파악하여 세부 유형을 정의.
- 또한 뉴스와 같은 공적미디어를 통해 정의된 리스크 유형이 개인화된 소셜 미디어에 나타난 리스크 유형과 어떤 차이가 있는지를 알아보기 위해 교차분석을 수행.
- 조사 결과에 따라 6개의 도메인별로 58개의 세부 유형을 정의하고 기계학습 방법을 통해 자동 분류 학습 모델을 구축한다. 실험 결과를 통해 소셜 미디어에 나타난 사회적 이슈 리스크를 자동으로 탐지, 분류가 가능함을 보인다.

(보건사회연구원) 2018년 소셜 빅데이터 기반 보건복지 이슈 동향 분석 (2019) ([링크](#))

- 위의 글 외에도 보건사회연구원에서 빅데이터 관련 키워드로 나오는 저널들은 모두 읽어볼 가치가 있음
(방법론 및 학술적인 배경에 대해서 자세히 기술됨)

텍스트 마이닝과 토픽 모델링을 기반으로 한 트위터에 나타난 사회적 이슈의 키워드 및 주제 분석 (2019) ([링크/파일](#))

- 본 연구는 커뮤니케이션이 활발한 SNS 속에서 사회적 이슈가 어떤 주제별로 나뉘어져 있고, 어떤 키워드들이 유기적으로 연결되었는지 그 연결 관계를 알아보고자 함.
- ‘미투’라는 새로운 단어가 생겨남과 동시에 큰 운동으로 번지고 있는 ‘미투운동’을 사회적 이슈로 간주하였고, 여러 SNS 중 특히 실시간 소통이 가장 활발한 트위터를 중심으로 분석을 실시.
- 우선 키워드를 ‘미투’로 하여 관련된 키워드를 각 날짜별로 추출하였고, 주요 키워드를 파악한 후 **토픽 모델링**을 수행하였다. 이를 통해 사회적 이슈를 둘러싼 키워드들이 시간의 흐름에 따라 어떻게 변화하였는지 파악하고, 각 토픽 내의 키워드를 종합하여 토픽별 사회적 이슈의 다양한 관점을 해석.

Website/Solution

펄스케이

<https://www.pulsek.com/portal/main/main>



리포트래시

<https://www.reportrash.com/>

매체	기자 수	신고 기사	기사 좋아요	기사 싫어요
조선일보	363	1640	116	20894
중앙일보	215	1107	76	13686
동아일보	167	636	26	6804
뉴데일리	60	545	29	6097
한국경제	129	438	38	4373
한국일보	97	290	35	3202
연합뉴스	112	280	35	2318
국민일보	82	255	32	2729
내셔널리뷰	77	272	25	2217

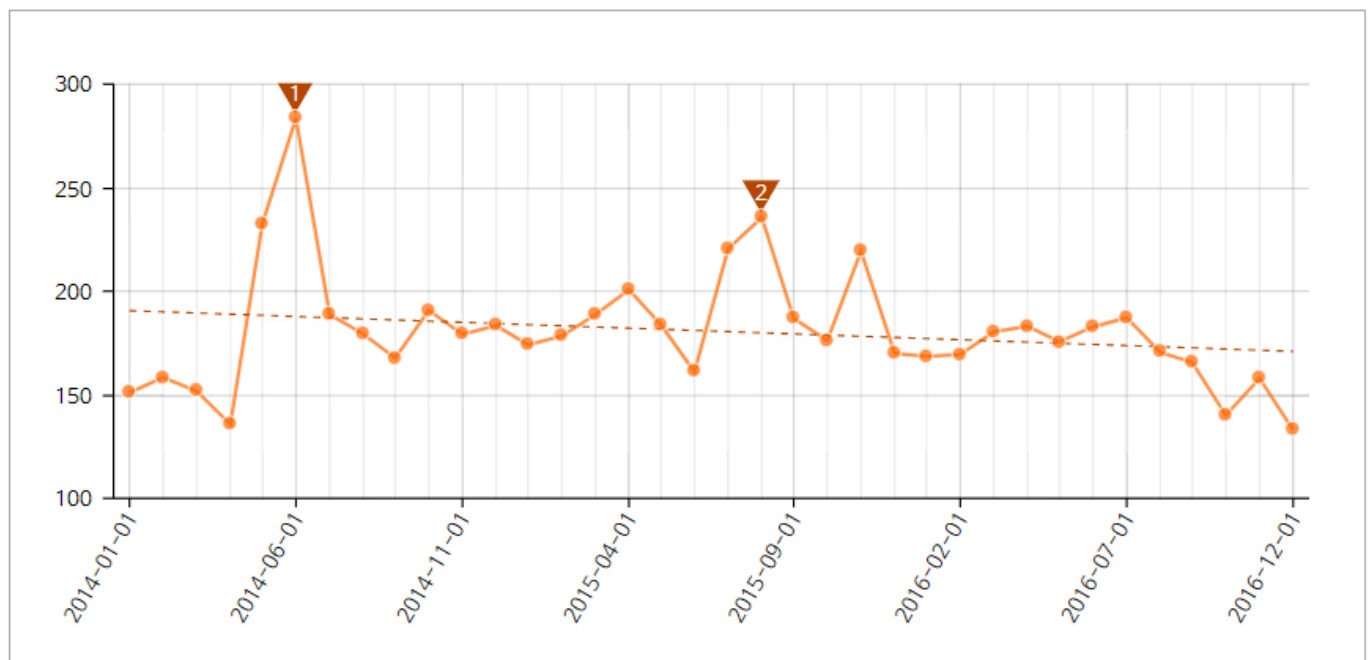
다음소프트 - 인공지능 보고서

http://www.daumsoft.com/resources/report_McDonald.html

맥도날드 버즈 추이

- '맥도날드'는 3년동안 블로그에서는 총 438,786,255건의 문서 가운데 **389,652**건, 트위터에서는 총 7,374,986,536건의 문서 가운데 **1,301,415**건, 게시판에서는 총 37,530,579건의 문서 가운데 **27,618**건의 문서에서 출현하였으며, '**하락세**'를 보이고 있습니다..
- '맥도날드'는 '**2014년 06월**', '**2015년 08월**'에 이슈가 발생하였습니다.
- '2014년 06월'의 이슈 키워드에는 '**맥피즈**', '**기말고사**', '**공주**'가 있습니다.
- '2015년 08월'의 이슈 키워드에는 '**시그니처**', '**맥와퍼**', '**세계평화 날**'이 있습니다.

다음은 '맥도날드'의 10만 건당 문서 빈도의 추이입니다.



SOMETREND

<http://www.some.co.kr/>

실시간 트렌드

- 1 진술거부권
- 2 좋은결과
- 3 케이팝
- 4 민주주의
- 5 초롱한그
- 6 프로듀스

연관어 맵

(기간 : 2019.10.15 ~ 2019.11.14)



7	레드벨벳
8	하이터치
9	멜론뮤직
10	부문후보엑스원
11	시카고
12	검찰개혁

