

오피니언 마이닝 기법을 이용한 키워드 분석 시스템 설계

문희준^o, 김동현(교신저자)*

^o동서대학교 소프트웨어학과,

*동서대학교 소프트웨어학과

e-mail: aglide100@gmail.com, pusrover@dongseo.ac.kr

The Design of Keyword Analysis System using a Opinion Mining Scheme

Hee Jun Moon^o, Dong Hyun Kim(Corresponding Author)*

^oDepartment of Software, Dongseo University,

*Department of Software, Dongseo University

● 요약 ●

최근 빅데이터를 통해 여러 가지 분석을 진행하고 있다. 다만 이러한 방식으로는 키워드에 대해 여론에 대한 분석을 거치지 않아 정확한 분석이 힘들다는 문제점을 가지고 있다. 따라서 본 논문에서는 이러한 문제점의 개선을 위해 데이터를 수집하고 이에 대해 감정분석을 수행하는 컨테이너 기반의 시스템을 제안한다. 감정분석 시스템을 적용한다면 키워드에 대해 분석 시에 정확도가 더욱 높아질 것으로 전망된다.

키워드: 오피니언 마이닝(Opinion Mining), 키워드(Keyword), 감정분석(Sentiment Analysis)

I. 서론

최근 다양한 원시 데이터에 대하여 빅데이터 기법을 이용한 데이터 분석 요구가 커지고 있다. 대표적인 예로서 대선 예측[1]이 있다. 이는 전통적인 여론조사를 하던 시절과 비슷한 방식을 이용하지만 가장 큰 차이점은 미디어를 통해 검색되는 관심도를 기반으로 예측도를 산출하는 것이다. 구글(Google)의 구글 트렌드(Google Trends)가 트럼프 대통령의 당선을 예측하면서 이슈가 되었고, 많은 기자가 [1]과 같은 기사를 작성하는 계기가 되었다.

하지만 구글 트렌드의 상황에 해당하는 키워드의 대한 긍/부정을 확인할 수 없다는 단점을 가지고 있다. 이는 구글 트렌드에서 사용한 빅데이터 기법이 해당하는 키워드에 대한 지역별, 빈도별, 시간별 검색량만의 수치를 단순히 비교하였기 때문이다. 이의 경우에 키워드에 대해 여론이 부정적이라도 관심도는 높기에 분석에 대해 오류 비율이 매우 높아질 가능성이 있다.[2].

관련 연구로 오피니언 마이닝(Opinion Mining) 기법[3]이 있다. 이 기법은 오피니언 도출을 위한 방법과 전략을 제안하였다. [4]에서는 오픈소스 라이브러리인 도커를 이용한 시스템 설계를 제안하였다.

본 논문에서는 대용량 데이터인 키워드 검색 결과에 대하여 텍스트가 가지는 감정 중 긍정, 부정 판단을 위한 오피니언 마이닝 기법 기반 키워드 분석 시스템을 제안한다. 제안하는 시스템은 오피니언 마이닝의 방법 중 하나인 감정분석(Sentiment Analysis)을 통해 수집된 데이터에 대해 감정을 분석한다. 그리고 컨테이너를 이용하여

조정가능하고 유연한 시스템을 설계한다.

본 논문의 구성은 다음과 같다. 2장에서 관련 연구를 기술한다. 3장에서 감정분석과 이를 위한 컨테이너 기반의 시스템 설계를 기술한다. 마지막으로 4장에서는 결론에 대하여 기술한다.

II. 관련 연구

[3]에서는 오피니언 마이닝과 이에 오피니언 도출의 판별 기법에 대해 분석 전략에 대한 설계를 제안하고 있다. 문서가 긍정, 부정, 또는 중립 등 어떤 견해를 가졌는지 판별하기 위해서 극성을 계산하게 된다. 극성(Polarity)을 계산하기 위해 미리 구축된 극성 사전을 통해 문서 전체의 극성을 분류 후 전체의 감정을 분류한다.

[4]에서는 오픈소스 라이브러리인 도커 컨테이너를 이용하여 컨테이너 매니저, 베이스 컨테이너, 서비스 컨테이너 그리고 컨테이너 저장소로 나누어진 시스템을 제안한다. 논리/물리적으로 고립된 컨테이너와 SeMC Overlay Network를 통해 컨테이너 간의 통신을 구현하고 컨테이너 매니저를 통해 쉽게 확장이 가능하며 가벼운것이 특징이다. 또한 컨테이너 기반이기에 재사용성과 새로운 환경에 대해 적용이 용이할 것으로 기대된다.

III. 오피니언 마이닝 기법을 이용한 키워드 분석 시스템 설계

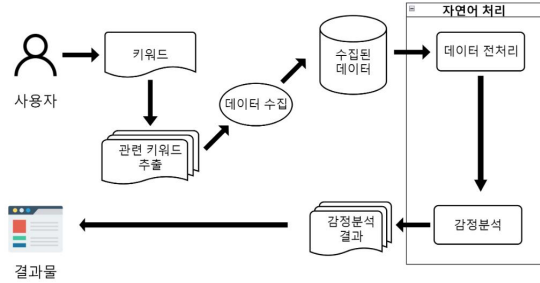


Fig. 1. 시스템 개념도

그림 1은 본 논문에서 제안하는 시스템의 개념도이다. 본 개념도에서는 사용자가 키워드를 입력 후 해당되는 키워드의 관련된 키워드를 추출한다. 그리고 추출된 정보를 통해 웹으로부터 데이터를 수집 한다. 수집된 데이터를 감정분석을 통해 결과물을 사용자에게 보여준다.

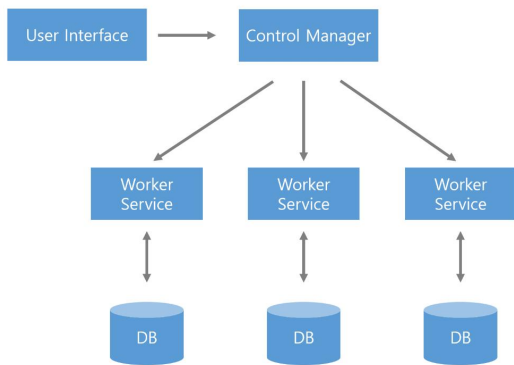


Fig. 2. 시스템 구성도

그림 2는 이러한 개념도를 구현을 위한 시스템 구성도이다. 사용자는 유저 인터페이스인 웹 UI를 통해 서버와 통신하며 결과와 키워드를 주고받는다.

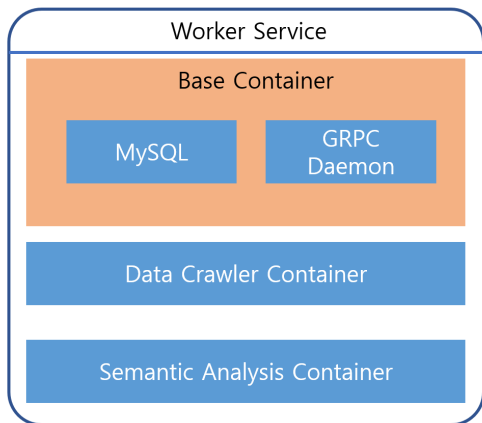


Fig. 3. 워커 서비스

이때 제어 매니저를 통해 그림 3의 워커리는 컨테이너화된 서비스를 작업에 맞게 프로비저닝한다. 각각의 워커리는 서비스는 DB와 내부 컨테이너로 구성되어 있으며 GRPC 프로토콜을 통해 서로 통신한다. 그리고 데이터 수집을 위한 컨테이너와 감정분석을 실행하는 컨테이너를 순번대로 실행하며 감정분석된 결과를 외부에서 관리되는 DB에 저장한다.

IV. 결론

구글 트렌드를 통한 분석은 결과에 대해 감정 판독이 불가능하여 키워드에 대해서 상이한 결과를 낼 수 있다는 문제점을 가지고 있다. 본 논문에서는 키워드에 대해 검색 결과에 대하여 극성을 추가하여 분석하는 시스템을 제안하였다. 향후 연구로는 컨테이너를 기반으로 데이터 수집과 감정분석을 하는 시스템을 구현하는 것이다.

ACKNOWLEDGEMENT

본 연구는 2022년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음(2019-0-01817)

REFERENCES

- [1] 이한빛, "빅데이터로 본 20대 대선", <http://www.joongboo.com/news/articleView.html?idxno=363529953>, 중부일보, 2022.03.10
- [2] 신현보, "표 뒤의 감정 읽지 못한 구글 대선 예측 처음으로 틀렸다", <https://www.hankyung.com/international/article/2020110992297>, 한경국제, 2020.11.09.
- [3] 정은희, 이병관, "오피니언 마이닝 기반 SNS 감성 정보 분석 전략 설계", 한국정보전자통신기술학회논문지, Vol. 8, No. 6, 2015.08.06
- [4] 정근훈, 박준석, 이극, "리눅스 컨테이너를 이용한 웹기반의 DevOps 플랫폼 연구", Journal of the Korea Convergence Societ, Vol. 10, No. 12, pp. 71-80, 2019