

빅데이터 기반의 텍스트를 활용한 개인 행복도 분석 모니터링 시스템 아키텍처 설계

심종성 · 김희철*

인제대학교

Personal Happiness Analysis using Big Data Based Text Design

Monitoring System Architecture Design

Jong-seong Sim · Hee-chul Kim*

Inje University

E-mail : jssin93@naver.com

요 약

전 세계의 수많은 SNS의 텍스트와 일기의 데이터가 업로드 되지만 그 데이터들을 내용을 공유하고 기록하는 것에 미치지 않는다. 일반적으로 소셜빅데이터는 취향, 관심사 파악에 사용되고 있다. 하지만 자신의 상태와 정보를 분석하고 나타내주는 시스템이 필요하다. 따라서 본 논문에서는 행복 다이어리 시스템은 SNS의 데이터와 자체의 일기를 적어 그것들을 빅데이터 시스템에 저장하고 감성분석을 이용하여 자신의 일기와 SNS데이터를 통해 행복도를 나타낼 수 있는 시스템 설계를 다룬다.

ABSTRACT

The text and diary data of many SNSs around the world are uploaded, but it does not go beyond sharing and recording the data. In general, social big data is used to identify taste and interests. However, there is a need for a system that analyzes and displays their status and information. Therefore, in this paper, the happiness diary system deals with the design of the system that can record the data of the SNS and its own diary, store them in the big data system, and express the happiness through their diary and SNS data using emotional analysis.

키워드

빅데이터, 소셜빅.데이터, 감성분석, 행복도

I. 서 론

일기 쓰기는 일상생활 속 개인의 마음속의 숨겨져 있는 감정을 가장 잘 표현 할 수 있는 글쓰기이며 SNS의 텍스트 또한 자신의 상태와 정보를 나타내는 중요한 글이며 정신적 혼란함을 치유하고 상처를 치유하는데 매우 중요한 역할을 한다.[1] 전 세계의 수많은 SNS의 텍스트와 일기의 데이터가 업로드되는 하지만 그 데이터를 활용할 수 있는 방법이 한정되어 있으며 그 데이터들은 내용을 공유하고 기록하는 것에 미치지 않는다. 현재

많은 기업들이 소셜빅데이터를 활용하려는 시스템을 많이 도입하고 있지만 취향, 관심사 등을 파악하는 것에 사용되고 있다.[2] 행복도 다이어리 시스템에서는 개인의 SNS의 데이터를 수집하며, 행복도 다이어리의 어플리케이션을 통하여 자신의 일기를 수집하여 빅데이터에 저장을 한다. 저장한 빅데이터를 가지고서 HBase를 이용하여 데이터를 저장 Spark를 이용하여 데이터 분석 후 행복도 결과 값을 저장한 것을 가지고서 사용자에게 자신의 행복도를 보여준다.

* corresponding author

II. 빅데이터 시스템

행복도 다이어리 시스템은 사용자의 SNS의 계정과 연동시켜 데이터를 가져오거나 Client 자체의 일기 데이터를 기록하는 빅데이터 시스템이다. 그리고 HBase와 Spark를 사용하여 흩어져 있는 데이터를 저장하고 데이터를 감성분석 후 행복도 결과값을 저장한 후 Client 프로그램에서 사용자에게 보여 줄 수 있도록 설계되어있다.

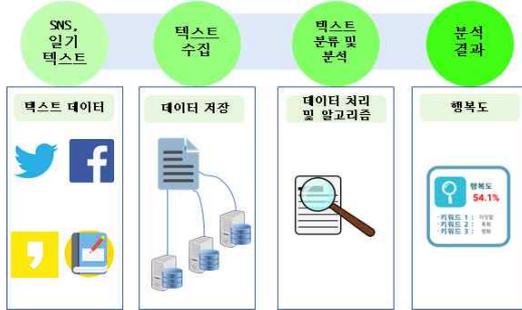


그림 1. 빅데이터 기반의 행복도 분석 모델

III. 감성분석

행복도 다이어리 시스템의 감성 분석을 Stemming Algorithm을 수행하여 동일한 어간을 가진 단어들의 어간을 추출함으로써 같은 형태로 만들 수 있기 때문에 여러 단어들을 동의어로 취급할 수 있다.[3]

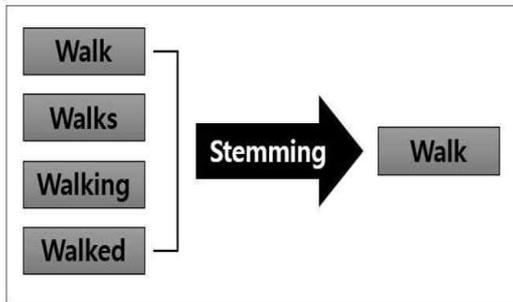


그림 2. Stemming의 적용 사례

행복도 분석은 구축된 사전을 통해 텍스트가 입력되면 해당 텍스트를 사전의 단어들과 같은 형태로 만들기 위해 전처리가 수행되고 만약 사전에 동일한 단어가 있다면 해당 단어의 회귀계수를 합산하여 행복 또는 불행의 점수를 얻고 얻어진 행복 점수와 불행 점수의 절대 값의 합을 100으로 보고 이 중 행복 점수의 비율을 따져서 행복도를 추정한다.[3]

Contents	Sentiment
사람/NNNG 사회/NNNG 대화/NNNG 능력/NNNG 포기/NNNG 때/NNNG 이	0
않/NNNG 순수/NNNG 도/NNNG	0
하루/NNNG 참/NNNG 주/NNNG	0
일/NNNG 예기/NNNG 기분/NNNG 때/NNNG 예기/NNNG	0
소년/NNNG 슬픔/NNNG 역사/NNNG 고향/NNNG 파도/NNNG 소년/NNNG 소	0
노골적/NNNG 감감/NNNG 후한/NNNG 생각/NNNG 처음/NNNG 간도/NNNG	1
친구/NNNG 기록사/NNNG 친구/NNNG 기록사/NNNG 친구/NNNG 오스만/	1
사람/NNNG 거짓말/NNNG 거짓말/NNNG	1
죽일/NNNG	1
아노/NNNG 사건/NNNG	1

1 Word	Coefficient	1 Word	Coefficient
2 갈사/NNNG	1.425297691	2 사기/NNNG	-0.80543623
3 절친/NNNG	0.923666193	3 게이/NNNG	-0.60613089
4 내입/NNNG	0.892668588	4 갑박/NNNG	-0.5614702
5 기분/NNNG	0.876681566	5 갈국/NNNG	-0.52001946
6 속하/NNNG	0.835907182	6 배허/NNNG	-0.51655309
7 본격적/NNNG	0.670106887	7 여성/NNNG	-0.49076357
8 선물/NNNG	0.603535444	8 가이/NNNG	-0.44658601
9 미소/NNNG	0.596299148	9 주희/NNNG	-0.44190333
10 타가/NNNG	0.592273911	10 주타/NNNG	-0.42580739
11 아이온/NNNG	0.559816824	11 수탈/NNNG	-0.41895744

그림 3. 행복과 불행의 회귀 계수 예

IV. 행복도 분석 Client

개인 소셜 미디어 이외에도 자신의 상태와 감정을 기록할 수 있는 행복도 다이어리 Client를 개발 하고 개인 소셜 미디어와 Client와의 연동을 통해서 SNS이 텍스트를 분석해 개인의 행복도를 모니터링 할 수 있도록 설계한다.



그림 4. 개인 행복도 모니터링 분석 예

V. 결 론

개인의 감정사향이나 정신적인 문제에 더 접근하기 어려울 때 행복도 다이어리 시스템은 자신이 알아채지 못한 부분에서 많은 도움을 줄 수 있다. 또한 행복도 다이어리 시스템을 활용하여 특정 상황이 발생했을 때를 각 개인의 행복도를 확인 할 수 있으며 어느 상황, 상태, 키워드에 따른 행복에 미치는 영향을 파악하는 시스템의 개발이 필요하다.

Acknowledgement

이 논문은 2019년도 산업통상자원부의 '창의산업융합 특성화 인재 양성사업'의 지원을 받아 연구되었음(과제번호 N0000717)

References

- [1] Tae-jin Park, “Study of journal writing relating to self-therapy and resuming self-identity for teenagers - focusing on case study through 'dialogue' writing, a journalistic therapy tool”, *작문연구*, Vol 11, No 0, pp. 9-44, Nov. 2010
- [2] Seoung-Hyun Seo, “Design and Implementation of Social BigData Analyzing System based on Hive”, *건국대학교 대학원 : 인터넷·미디어공학과*, Aug. 2015
- [3] Ki-Won Choi, “A Study on the Sentiment Analysis Using Text Mining and the Development of Regional-Based Happiness Analysis System”, *인제대학교 일반대학원 : 디지털향노화헬스케어학과 헬스케어 소프트웨어*, Aug. 2018