



SNS Map : 위치 기반 SNS 데이터 맵핑 시스템

SNS Map : Location-based SNS data mapping system

저자 (Authors)	이승훈, 오대영, 강민혁, 김현철 Seunghun Lee, Daeyoung Oh, Minhyuk Kang, Hyunchul Kim
출처 (Source)	한국정보과학회 학술발표논문집 , 2015.06, 1119-1121 (3 pages)
발행처 (Publisher)	한국정보과학회 KOREA INFORMATION SCIENCE SOCIETY
URL	http://www.dbpia.co.kr/Article/NODE06394329
APA Style	이승훈, 오대영, 강민혁, 김현철 (2015). SNS Map : 위치 기반 SNS 데이터 맵핑 시스템. 한국정보과학회 학술발표논문집, 1119-1121.
이용정보 (Accessed)	상명대학교 천안캠퍼스 203.237.183.*** 2018/05/02 15:46 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

SNS Map : 위치 기반 SNS 데이터 맵핑 시스템

이승훈⁰¹ 오대영² 강민혁³ 김현철¹
 상명대학교 컴퓨터공학과

{mr.leesh90, daeyoungoh365, kkang901231, hyunchulk}@gmail.com

SNS Map : Location-based SNS data mapping system

Seunghun Lee⁰¹ Daeyoung Oh² Minhyuk Kang³ Hyunchul Kim¹
 Dept. of Computer Engineering, Sangmyung University

요 약

매일 트위터, 페이스북, 유튜브와 같은 Online Social Networks(OSN)에서는 하루 수억 건의 많은 메시지가 올라온다. 이 수많은 메시지 사이에서 사용자가 원하는 메시지와 정보를 효율적으로 찾기란 어려운 일이다. 이러한 이유로 좀 더 편리하고 간편하게 원하는 메시지와 정보를 찾기 위한 시스템이 만들어지고 연구가 되고 있다. 특히 맵을 통해서 수집되고 있는 데이터를 보여주는 많은 시스템이 중점적으로 만들어지고 있다. 하지만 이 수집된 데이터를 분석까지 해서 보여주는 시스템은 거의 없다. 본 논문에서는 OSN 중에 하나인 트위터에서 PlanetLab System을 이용한 대규모 데이터를 수집하는 방법과 수집된 데이터를 맵을 통해서 보여주는 시각화, 키워드 분석을 통한 검색, 필터링 기능, Sentiment Analysis & Word Cloud를 이용한 데이터 분석을 이용해서 만든 위치기반 SNS 데이터 맵핑 시스템을 제안하고, 초기 프로토타입인 SNS Map 시스템을 구축한다.

1. 서 론

현재 많은 Online Social Networks(OSN)에서 하루 수억 건의 많은 메시지가 올라온다. 이 메시지들 사이에서 사용자가 원하는 메시지와 정보를 찾아보기에는 너무나도 방대한 양이다. 이러한 이유로 좀 더 편리하고 간편하게 어떠한 사건들을 알아보고 사람들의 생각을 알아보는 시스템이 많이 만들어지고 있다. 지금 당장 올라오고 있는 트윗의 위치를 맵으로 보여주는 Tweeping.net [1], 트윗의 위치와 키워드, 해시태그까지 검색 가능한 The one million tweet map [2], 구글 맵을 이용해서 여러 가지 맵으로 트윗들을 보여주고 검색 가능한 Harvard TweetMap ALPHA [3], 그리고 A WORLD OF TWEETS [4] 등이 있다. 이러한 좋은 시스템들에 더해 본 논문에서는 트윗의 Mapping 작업뿐만 아니라 검색된 트윗을 분석해서 제공까지 해주는 시스템에 관해서 설명한다. 본 논문에서는 2장에서는 트위터와 관련된 사이트 시스템에 관해서 설명하며, 3장에서는 본 논문의 SNS Map 시스템 구조에 대해서, 4장에서는 데이터를 수집하는 방법과 수집하는 데이터의 종류를 설명하고, 5장에서는 SNS Map 시스템에서 제공하는 서비스에 관해서 설명하고 마지막으로 6장에서는 결론에 관해서 기술한다.

2. 관련 시스템

본 논문에서 제안한 시스템인 SNS Map 시스템과 관련된 시스템 총 4가지를 소개한다. 표 1에서 보듯이 총 4가지의 시스템을 비교 분석했다. 기본적으로 시스템 모두 여러 가지 맵을 사용해서 트윗을 맵 상에 출력하는 기능은 모두 가지고 있다. 두 번째 기능은 검색이 가능한 필터링 기능이다. Tweeping.net, A WORLD OF TWEETS을 제외한 나머지 두 시스템은 검색 필터링이 가능하다. 세 번째로는 분석 기능인데 어떤 시스템에서도 데이터에 대해서 분석을 하지는 않았다. 다음으로 데이터에 대한 통계자료 제공인데 이 기능도 The one million tweet map, A WORLD OF TWEETS만 간단한 통계자료를 제공하고 있다. 다섯 번째 기능은 실시간 트윗이다. Harvard의 TweetMap ALPHA를 제외한 나머지는 모두 실시간성을 기준으로 데이터를 보여준다. 이와는 반대로 TweetMap ALPHA는 2013/12/3 ~ 12/17일까지 15일간의 데이터만을 간략하게 사용해서 보여주고 있다. 하지만 TweetMap ALPHA는 다른 시스템과는 다르게

여러 종류의 맵을 사용함으로써 맵의 형태를 다양하게 볼 수 있다는 큰 장점이 있다. 표 1에서 볼 수 있듯이 모든 기능을 제공하고 있는 시스템은 아직 존재하지 않는다. 무엇보다도 분석을 실제로 하고 맵핑 작업까지 같이 해주는 시스템은 존재하지 않는다. 이에 본 논문의 SNS Map 시스템에서는 맵핑은 물론, 키워드 검색, 간단한 톨을 이용한 분석 부분과 실시간 & 과거의 데이터를 이용한 범위까지 서비스를 제공한다.

표 1. 트위터 관련 사이트 시스템

종류 기능	Tweeping [1]	The one million tweetmap [2]	Harvard Tweet Map ALPHA[3]	A WORLD OF TWEETS [4]
맵핑	O	O	O	O
키워드 검색	X	O	O	X
분석	X	X	X	X
대륙별 데이터량 통계	X	O	X	O
실시간 트윗	O	O	X	O
과거 데이터	X	X	O	X

3. 시스템 구조

본 논문에서 사용된 시스템 구조는 그림 1처럼 데이터 수집 부분, 맵핑 부분, 분석 & 결과 부분으로 나누어진다. 데이터 수집은 PlanetLab [5]의 각 노드에서 Twitter Streaming API [11]를 통해서 각 노드에 트위터 데이터를 텍스트 파일로 저장하게 된다. 텍스트 파일로 저장된 데이터를 한 시간마다 서버로 이동시켜서 중복처리 작업을 거쳐서 MongoDB [12]에 저장시킨다. 맵핑은 2가지 부분이 있다. Heat Map을 이용한 실시간 맵핑 부분, MongoDB에 있는 데이터를 이용해서 검색되는 키워드에 따라서 검색 결과를 맵핑 해주는 부분이다. Heat Map 같은 경우에는 Heat Map을 실행할 경우 바로 실행이 되면서 실시

해 볼 수 있다. 또한, 가장 많이 쓰는 단어를 표시함으로써 사람들이 무슨 단어를 가장 많이 언급하는지에 따라서 주요 관심거리에 대해서 예측할 수 있다.

5.2 Mapping and Time-series graph

지금까지 수집된 데이터와 분석 작업까지 마친 데이터를 보여주는 방법에는 다른 시스템에서 볼 수 있듯이 맵을 통한 맵핑 작업이다. 수집되는 트윗 데이터에는 GPS 정보가 다 있어서 GPS 좌표 값에 따라서 맵에다 표시를 할 수 있다. 또한, 보여주는 것뿐만 아니라 사람들이 원하는 트윗이나 특정 단어가 들어가는 트윗들만 따로 키워드 검색을 통해서 맵과 분석을 같이 하는 작업을 통해서 보여 주게 된다. 지도를 통해서 데이터를 보여주는 방법에는 2가지가 있다. 첫 번째로 Twitter Streaming API를 통해서 현재 실시간으로 모이고 있는 데이터를 그림 3과 같이 구글 맵을 통해서 Heat Map을 보여주게 된다. 2장에서 소개한 다른 시스템에도 있는 필수 기능으로 현재 수집되고 있는 트윗 양에 따라서 지도 음영의 정도가 달라진다.

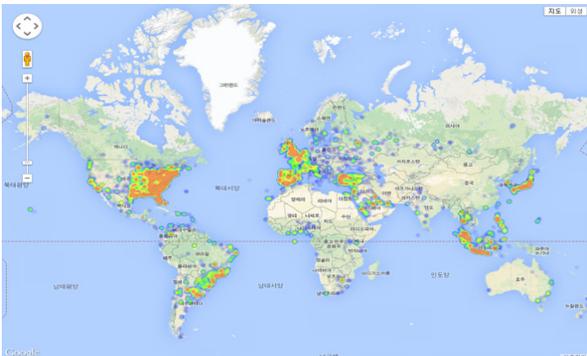


그림 3 실시간으로 수집되는 데이터의 Heat Map

두 번째로는 실시간으로 모여지는 데이터를 보고 실제로 원하는 키워드나, 지정되는 날짜나 시간을 지정해서 검색하게 되면 지정된 트윗들만 따로 검색해서 지도에 맵핑을 하는 방법이다. 맵핑을 하는 작업에는 Leaflet을 사용했다. Leaflet은 자바스크립트 오픈소스로서 Leaflet은 지도에 최대 축소를 통해서 그림 4와 같이 트위터 하나하나의 글을 직접 모두 다 읽어 볼 수 있고 지도의 축척에 따라서 자동으로 영역도 조절해준다. 또한, Leaflet은 JSON 데이터를 사용하기 때문에 트위터에

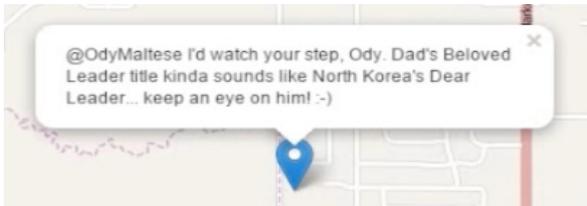


그림 4 검색 필터링을 통해 검색된 트윗을 Leaflet을 이용해 확대한 맵

서 제공해주는 데이터와 형태가 맞아 사용기가 매우 편리하다. 검색 키워드와 날짜 등에 따라서 제한된 검색을 통해 추출된 데이터를 Leaflet을 통해 맵에 맵핑 시켜준 후에는 Sentiment Analysis를 통해서 추출된 결과를 통계적 그래프를 통해서 요약해 출력한다. 그림 5는 2014년 4/10~4/30일까지 날짜를 제한하고 Korea라는 단어가 들어간 트윗을 미국지역에서만 선택해서 감정 분석을 실행시켜서 정리한 결과이다. 연두색은 Neutral로 감정을 나타내는 글이 아닌 어떠한 사실을 나타내는 글이거나 어느 쪽으로도 기울지 않은 감정을 나타내고 있는 글에 해당한다. 빨간색은 Negative로 부정을 뜻한다. 트윗에 부정적인 어휘가 많이 들어간 경우를 나타낸다. 파란색은

Positive로 긍정을 뜻하는 색깔로 긍정적인 어휘를 많이 포함한 트윗의 숫자를 나타낸다. 그림 5를 통해서 알 수 있는 사실은 4/16일에 우리나라에 대한 글을 적던 미국인들의 숫자가 갑자기 증가하고 무슨 일이 일어났고 부정적인 어휘가 많이 들어가는 트윗들로 보아서 안 좋은 일이 있었던 것을 예측할 수 있다. 바로 세월호 사건이 일어났던 날이다.

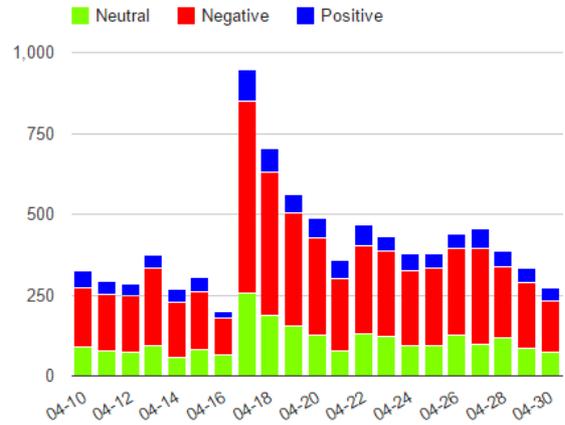


그림 5. 2014년 4/10~4/30 미국지역, Korea라는 키워드 검색의 Sentiment Analysis Graph

6. 결론

본 논문에서는 지속적인 시간 동안 효율적이고도 많은 양의 트위터 데이터를 수집하는 방법과 그렇게 수집된 데이터의 좌표를 통해서 맵에 표시해주고, 특정 키워드 검색, 분석 등을 통해 나온 결과를 그래프를 통해서 보여주는 위치기반 SNS 데이터 맵핑 시스템을 제안하고, 그 초기 프로토타입으로 SNS Map을 구현했다. 중점적인 향후 과제로는, 대규모의 SNS 데이터를 처리하고 지도에 맵핑 시켜 줄 수 있도록 시스템을 확장하는 것이다. 또한, 본 시스템에서는 우선 ALPHA 버전 구현물로 우선 트위터 데이터를 이용한 시스템을 구축하였지만, 인스타그램, Pinterest, Tumblr, Foursquare 등 GPS 정보, 사진, 글 등을 이용하는 다른 OSN의 데이터도 사용할 수 있게 만들어서 더 많은 데이터를 이용하도록 확장할 것이다.

참고문헌

- [1] tweetping.net <http://tweetping.net>
- [2] The one million tweetmap <http://onemilliontweetmap.com>
- [3] TweetMap ALPHA <http://worldmap.harvard.edu/tweetmap>
- [4] A WORLD OF TWEETS <http://aworldoftweets.frogdesign.com>
- [5] PlanetLab <https://www.planet-lab.org>
- [6] Word_cloud https://github.com/amueller/word_cloud
- [7] C. Ding et al., "Crowd crawling: towards collaborative data collection for large-scale online social networks." ACM COSN, 2013.
- [8] S. Baccianella et al., "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining." LREC. Vol. 10. 2010.
- [9] T. Sakaki et al., "Earthquake shakes Twitter users: real-time event detection by social sensors." WWW, 2010.
- [10] Leaflet <http://leafletjs.com>
- [11] Twitter Streaming API <https://dev.twitter.com/streaming/overview>
- [12] MongoDB <http://www.mongodb.com>