

효과적인 평판분석을 위한 수집사이트 프로파일링에 관한 연구

송은지*, 강민식**

*남서울대학교 컴퓨터학과

**남서울대학교 산업경영공학과

A Study on the Profiling of Collect Site for the Effective Reputation Analysis

Eun-Jee Song*, Min-Sik Kang** ,

*Dept. of Computer Science , Namseoul University

**Dept. of Industrial and Management Engineering , Namseoul University

sej@nsu.ac.kr, mskang@nsu.ac.kr

요 약

본 논문에서는 보다 정확하고 효과적인 평판분석을 위하여 서비스 산업별 타겟으로 하는 수집사이트를 프로파일링 하는 방법을 제안한다. 먼저 각 서비스에 특화된 타겟 사이트를 추출하고 등록하고 각 서비스에 관련한 정보 및 의견 공유 게시판과 지식인 추천/질문 등 지식 공유 사이트를 추출한다. 또한 업종별 주요 사이트를 선택하고 등록하여 유효 데이터 수집한다. 이를 통해 실시간 수집 데이터의 활용 기술을 이용하여 수집된 프로파일링을 통한 미디어별 수집 주기 산정하고 수집 엔진의 유연한 확장성을 활용한 실시간 수집 제반 기술 확대할 수 있다. 또한 지속적인 수집원 변경관리를 수행한다. 즉, 신규 생성, 변경, 삭제되는 사이트에 대한 변경관리를 수행하고 지속적인 수집량 모니터링을 통한 수집여부를 점검하며 수집 필터링 규칙에 대한 튜닝으로 데이터 품질 확보하도록 한다.

키워드

Profiling, Collecting site ,Target Collection, Reputation Analysis,

1. 서 론

기업들은 이제 제품의 품질과 가격을 선진하는 것보다는 그 제품과 관련된 서비스로 승부하는 서비스 경제시대가 도래하였다. 이러한 서비스 경제에서는 무엇보다 고객에 대한 데이터가 중요하며 따라서 기업들은 고객과 관련한 보다 많은 양의 데이터를 생성하고 추적하려고 노력하고 있다. 최근 특히, 스마트폰의 출현으로 블로그나 SNS등에 기업이 상품이나 서비스를 팔고자 하는 다양한 소비자들의 목소리가 증가하면서 실시간으로 소비자의 니즈를 파악할 수 있는 방법으로 트위터, 블로그, 카페 등 SNS 상의 빅 데이터를 이용하는 것이 매우 효율적인 방법으로 부상하고 있다. 이렇게 온라인상에서 고객들이 쏟아내는

말을 실시간으로 분석, 조사하는 방법으로 버즈 모니터링이라는 방식이 개발되고 있는데 이것은 웹상의 다양한 빅 데이터 정보를 검색하고 수집하여 분석하는 시스템이다. 이 시스템은 카페와 블로그 등 웹사이트를 비롯해 페이스북 북, 트위터와 같은 소셜 미디어 사이트에 올라온 광범위한 고객의 의견 텍스트 정보를 수집하고 분석한다.

본 논문에서는 보다 효율적이고 정확한 고객만족도 피드백 측정을 위한 빅 데이터를 이용하여 서비스별 품질 평가 방안을 제시하고 각 서비스에 특화된 타겟 수집사이트 프로파일링 수행방법을 제안한다.

II. 타겟 사이트

수집사이트 프로파일링 수행을 위해 먼저 각 서비스에 특화된 타겟 사이트를 추출하고 등록한다.

예를 들어 관광서비스 산업이라 하면 그림1에서와 같이 관광, 여행 등에 관련된 영향력자, 파워블로거 등을 추출하여 최신 트렌드 및 여론동향 분석에 활용할 수 있다. 이를 위해서 먼저 각 서비스에 관련한 정보 및 의견 공유 게시판과 지식인 추천/질문 등 지식 공유 사이트를 추출한다. 또한 업종별 주요 사이트를 선택하고 등록하여 유효 데이터 수집한다.

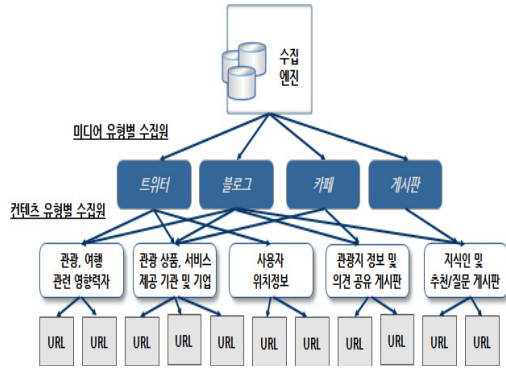


그림1. 관광 산업군에 특화된 타겟 사이트 추출

이를 통해 실시간 수집 데이터의 활용 기술을 이용하여 수집원 프로파일링을 통한 미디어별 수집 주기 산정하고 수집 엔진의 유연한 확장성을 활용한 실시간 수집 제반 기술 확대할 수 있다. 또한 지속적인 수집원 변경관리 수행한다. 즉, 신규 생성, 변경, 삭제되는 사이트에 대한 변경관리 수행하고 지속적인 수집량 모니터링을 통한 수집 여부 점검하며 수집 필터링 규칙에 대한 튜닝으로 데이터 품질 확보하도록 한다.

III. 수집 사이트 프로파일링

본 연구에서는 각종 서비스 모델 구축을 위한 사전 온라인 조사를 통하여 수집한 모니터링 대상 사이트 및 사이트 선별 기준에 의하여 발굴된 사이트를 대상으로 산업군 별 수집사이트를 추출한다. 수집 사이트 조사 및 설정 절차를 통하여 각 서비스 산업군에 해당되는 주요 수집원 및 타겟 모니터링 대상 추출을 통한 수집 정확도 향상 및 수집 인프라(H/W)를 최소화하여 보다 효율적인 운용이 가능하다. 수집 사이트 선별 기준은 표1과 같다. 제안한 수집 사이트 프로파일링과정을 실제 관광 서비스산업에 적용한 결과 매스미디어 상의 다양한 카테고리별 수집 사이트의 특성, 수집 근거 등 사이트 활성화에 대한 분석을 통한 수집 주기를 선정하고, 수집원 특성별 영향도를 분석하여 미디어별 가중치를 적용하고 관광 관련

파워블로거 (여행, 맛집 등) 및 우수 블로거를 추출하여 수집에 추가하였다. 그 결과 수집원별 가중치 적용을 위한 임계치 설정의 근거자료로 활용하며 약 3천개의 타겟 수집원을 추가하였고 타겟 사이트가 추출되었다.

표1. 수집사이트 선별 기준

구분	내용	비고
게시글	사이트명	산업군별 대표사이트 개별 추출
	게시판명	
작성자정보	작성자 분류	
사이트 활성화	섹션(기사)	
	회원수,가입자수 당일새게시글수	사이트 활성화 수준 파악
	조회수	
사이트 수집	댓글	
	최근 게시글수	사이트 활성화 수준 파악
유효성	등업 여부	회원등급제 운영시 등업 필요
	로그인 여부	로그인 시 계정정보 필요
유�효성	기계적접근가능여부	사이트의 복잡도로 인한수집 불가여부 확인

IV.결 론

최근 IT업체들은 수많은 소셜 미디어를 통해 폭발적으로 쏟아지는 빅 데이터를 분석하기 위한 시스템을 경쟁적으로 개발하고 있는데 대표적인 것이 버즈 모니터링시스템이다. 본 논문에서는 소셜미디어 상의 빅 데이터를 수집하는데 있어 보다 효율적이고 정확한 고객의 피드백을 파악하기 위해 타겟 수집사이트를 프로파일링하는 방법에 대해 제안하고 제안한 방법에 의해 관광 서비스 산업에 적용한 결과 약 3천개의 타겟 수집원을 추가하였다.

참고문헌

[1] K.P. Nam, "System Implementation of the Customer Satisfaction Survey Using Internet ", *TheKoreanJournalofAppliedStatistics*vol.18,no.3,pp.713-727, 2005.

[2] C.H. Lee, "Technology Trends of Issue Detection and Predictive Analysis on Social Big Data", *ETRI Electronics and Telecommunications Trends*, pp62-71, 2013.

[3] 김철원, "관광산업 경쟁력 평가모델 개발", 한국관광연구원, 2000.

[4] E.J. Song, S.J. Hong and M.S. Kang, "A Study on the System for Data Collection and Analysis on Social Network" *Proceedings of International Workshop on Networking and Communication*, 2013.