

인터넷 커뮤니티 콘텐츠 분석을 위한 사례베이스 구축

고민정
 동국대학교 컴퓨터공학과
 e-mail:mjgo@dongguk.edu

Building Casebase for the Content Analysis in Internet Community

Ko Min Jung
 Dept of Computer Engineering, Dongguk University

요 약

인터넷의 대중화로 커뮤니티에 콘텐츠가 대량으로 생성되고 있고, 이들은 또 다른 측면의 여론으로 그 비중이 커져가고 있다. 이러한 변화에 기존에 구축된 웹 데이터베이스를 기준으로 생성된 인터넷 커뮤니티 콘텐츠를 처리하고자 할 경우, 새로이 생성된 언어와 패턴으로 인하여 분류 및 관리에 어려움이 많다. 본 연구는 이를 해결하기 위해서 실시간으로 웹로봇을 활용하여 새로운 사례와 데이터를 수집하고, 이를 사례별로 분류한 사례베이스를 구축하여 대량의 커뮤니티 콘텐츠 분석이 가능하다.

1. 서론

최근 인터넷 커뮤니티가 활성화됨에 따라서 다양한 네티즌의 의견이 사용자 게시판에 생성되고, 이를 기반으로 하는 새로운 형태의 네티즌 여론으로 자리 잡고 있다[4]. 여기서 각 인터넷 커뮤니티들은 올바른 여론 주도를 위하여 게시된 콘텐츠를 분석 및 분류하기 위한 데이터베이스가 필요하다[1][7]. 현재는 각 인터넷 사이트 관리자가 수작업으로 등록된 콘텐츠 중에서 비방이나 광고성 내용들만을 삭제하는 방식으로 관리되고 있다[4]. 이를 해결하기 위해서 본 연구에서는 새로운 사례를 실시간으로 수집하여 사례베이스로 구축하는 방안을 제시한다.

여기서는 현재 사용되는 신조어들과 새로이 생성되는 콘텐츠 구조들을 신속하게 분석할 수 있는 사례베이스 구축을 위한 방법과 절차를 제시한다. 이를 위해서 사례베이스를 구축하는데, 이는 기존 데이터베이스의 분류기준을 따르지 않고, 분석 대상인 콘텐츠의 내용에 따라서 긍정사례, 부정사례, 중립사례, 금치어사례 등으로 분류하여 구성한다. 또한 각 해당 표현에 실시간으로 생성되는 서브 사례들을 웹로봇을 이용하여 수집하고, 구축하여 검색이 가능하도록 설계한다.

2. 관련연구

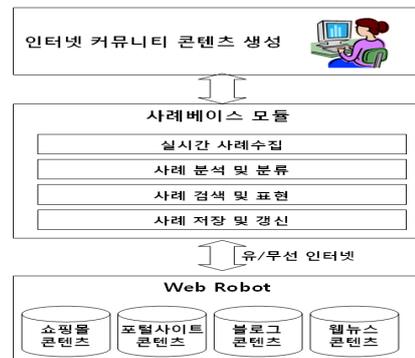
본 연구와 관련된 기존의 연구로서는 사례기반 추론, 지능형 에이전트 연구와 지능형 로봇이 있다.

사례기반 추론은 인공지능의 한 방법으로 주어진 새로운 문제를 과거의 유사한 사례를 바탕으로 해결하는 기반이다[6][8]. 본 연구에서는 이러한 사례기반의 추론을 인터넷 커뮤니티 콘텐츠에 활용한다.

사례베이스를 이용한 추론의 이론적 배경은 유사한 문제는 언제나 다음에 일어날 확률이 높다는 방법론을 활용하여 새로운 문제를 해결하는 대표적인 이론이다[8]. 인터넷 커뮤니티에서 생성되는 콘텐츠의 경우도 기존에 게시물을 참조하여 이를 수정하거나 보완하여 재생산되는 경우가 많다[3][5]. 이와 같이 인터넷 사이트들의 콘텐츠는 기존의 콘텐츠를 재가공하거나 참조하여 새로운 콘텐츠를 생성한다는 특성을 갖는다고 할 수 있다. 이는 본 연구에서 기존의 사례를 통해서 새로운 콘텐츠를 분석하는 방법에 대한 근거가 된다.

3. 콘텐츠 분석을 위한 사례베이스 구축

본 연구에서 구축할 사례베이스를 기반으로 인터넷 커뮤니티 콘텐츠 분석을 위한 전체 구조도는 다음 (그림 1)과 같다.



(그림 1) 사례베이스 기반 콘텐츠 분석을 위한 구조도

