

지능형사이버선거범죄대응시스템 관련자료 공개

1 시스템 소개

패턴인식, 유사성 분석 등 기술을 활용하여 수집한 온라인상 게시물을 위법혐의가 높은 순서에 따라 순차적으로 정렬함으로써 모니터링한 위법 게시물을 신속히 조치할 수 있게 하여 확산을 차단하는 시스템

2 시스템 업무처리 흐름

- 초기학습자료, 검색키워드 입력
- 게시물 수집: 주요 포털 및 커뮤니티 사이트 등
 - ※ 폐쇄형 사이트인 카카오톡, 밴드 등은 수집 불가
- 알고리즘에 의한 분류: 기학습된 데이터 기반 패턴·유사성 분석
- 수집 게시물 분류: 위반혐의가 높은 순서대로 화면표출
- 삭제요청 및 재학습: 위법성 검토 후 삭제요청한 사례를 기반으로 재학습

3 분류 알고리즘

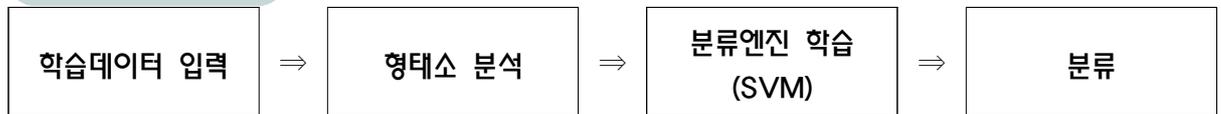
○ 자동분류엔진(DQCAT3)

본 시스템에 활용되는 솔루션으로서 자동 분류요청 클라이언트를 통해서 분류 대상 문서의 정보를 전달받아 형태소 분석기를 통해 텍스트로부터 키워드를 추출, 학습DB를 기반으로 자동 분류 수행

○ 학습모델(SVM)

기계학습의 한 종류로서 패턴인식, 자료분석을 위한 학습 모델이며, 주로 분류와 회귀분석을 위해 사용. SVM 알고리즘은 주어진 데이터 집합을 바탕으로 새로운 데이터가 어느 카테고리에 속할지 판단하는 분류모델

SVM 분류절차



4 공개목록

- 업무처리 흐름도
- 초기학습데이터 목록(2017년이후 임기만료선거에서 삭제요청 자료)
- 검색 키워드
- 분류 알고리즘 소스파일