**주제 : 우리회사관련 부정게시물 찾아내기**

**1. 배경 및 필요성**

평판이 좋지 않은 몇몇 기업들이 있다. 대부분 회사내 갑질, 서비스 부진, 제품이상이 원인인데 그것은 기업 내부 규정부터 바뀌어야 하는 것이라 기업 임원들의 생각변화가 필요하다.

하지만 사실상 대부분의 기업은 그렇지 않고 직원들에게 글들을 보면서 원인을 분석하고 당장 직면한 문제를 해결하라고 한다. 위의 일을 하는 부서가 따로 있는 기업도 있지만 게시물들이 너무 많아서 시간이 오래 걸리는 문제가 있다고 한다.

우리는 인터넷 포털 사이트 ‘naver’의 각 매체 (네이버 뉴스, 네이버 블로그, 네이버 카페 등)로부터 해당 기업이 제시한 키워드로 검색한 기준의 게시물들을 크롤링 해와서 그 게시물에 기업이 부정 키워드라고 정한 키워드가 포함돼 있으면 부정 게시물로 자동 판별해서 알려주는 프로그램을 제안한다.

**2. 목적 및 접근방법**

사용자(ex. 기업직원)가 좀 더 빠르고 한눈에 게시물들을 확인할 수 있고, 미리 판별된 부정 게시물도 따로 모아서 볼 수 있다. 이를 통해 일처리의 신속성 향상을 돕는 것을 목적으로 한다.

* 게시물들을 크롤링 해온 다음 제목, 글, URL 등으로 카테고리를 나눠서 데이터베이스에 저장한다.
* 저장된 데이터를 R 프로그램으로 불러온다.
* 게시글 내용에 부정적인 키워드가 존재하는지 판단 후 부정 / 일반 게시물로 나눈다.
* 부정적인 게시글을 따로 모아서 사용자에게 필요한 카테고리만 요약해서 데이터 프레임을 생성한다.
* 해당 프레임을 엑셀로 변환해서 사용자에게 제공한다.

**3. 데이터 획득계획**

‘naver’로부터 각 매체에서 크롤링하여 데이터베이스에 데이터를 획득한다.

**예시**

**<게시물 크롤링>**

설명 : ‘naver’의 게시물 내용들을 긁어와서 데이터베이스에 저장하는 것.

참고 : 컬럼중에 rate => 부정 flag ( 3 : 보통게시물 2 : 부정게시물)

* detect\_keyword => 글 내용중에 어떤 부정 키워드가 있었는지
* title => 글 제목
* contents => 글 내용
* url => 글 url





**<부정적인 게시물로 판단했을 때 >**

