

제목 : OCR 및 Average Hash 기법을 사용한 이미지 기반 유해사이트 판별 방법

1. 논문 주제 및 배경:

유해한 정보를 포함하는 웹사이트를 식별하고 분류하기 위한 시스템을 제안함.

2. 제안된 방법:

- OCR(광학 문자 인식)과 Average Hash 기법을 활용하여 유해 사이트를 식별하는 시스템을 제안했음.
- 도박 배너 광고의 유사한 이미지를 반복적으로 사용하는 특징을 활용하여 배너 광고 이미지의 평균 해시 값을 분석함.
- Easy OCR을 사용하여 배너 광고에 작성된 문구가 유해한지 여부를 판단함.

3. 실험 및 평가:

- 제안된 방법의 성능을 평가하기 위해 사이트 이름을 입력할 때 해당 사이트의 배너 광고 이미지를 수집하고 분석하는 프로그램을 생성함.
- 프로그램 실행 시 수집된 정보는 데이터베이스에 저장되어 예상되는 유해 사이트를 검색하는 데 활용함