|  |
| --- |
| **1. 주제**  **글꼴 분석을 통한 DB구축과 글꼴 추천 시스템**  **(나)분반, 7팀, 20211732** |

|  |  |
| --- | --- |
| **2. 요약**  디지털 시대에 미디어 콘텐츠의 중요성이 대두되며 경쟁력을 높여줄 글꼴의 영향력 또한 높아지고 있다. 하지만 글꼴은 저작권의 보호를 받고, 이 경계 또한 모호해 일반 사용자가 사용범위를 구분하기는 쉽지 않다. 따라서 사용자가 입력한 글꼴 파일의 저작권 관련 사항과 비슷한 글꼴을 추천하는 시스템 개발을 기획하고자 한다.  HOG 기술을 사용하여 글꼴 파일의 특징을 추출하고, 저작권 관련 사항을 함께 저장하여 데이터베이스를 구축한다. 또한 비슷한 글꼴 파일을 추천하기 위해 계층적 군집화 기술을 통해 데이터베이스의 데이터들을 군집화한다.  이번 프로젝트를 통해 글꼴 정보의 접근에 어려움을 느끼는 일반 사용자들이 손쉽게 저작권 관련한 정보를 확인하고, 비슷한 글꼴을 사용하는 등 글꼴을 사용하면서 느낀 어려움을 해결 할 수 있다. | **3. 대표 그림**    (그림1) 글꼴 검색 순서도 |

|  |
| --- |
| **4. 서론**  디지털 시대인 현재 미디어 콘텐츠는 사람들에게 중요한 부분으로 자리 잡았다. TV 예능과 드라마, 광고, 유튜브 영상과 같이 모든 미디어 콘텐츠 분야에서 사람들의 이목을 끌기 위해 상황에 맞는 적절한 글꼴(타이포그래피)을 사용하고 있다. 글꼴을 적절하게 사용하면 가독성을 높이거나 상황의 이해를 도와 콘텐츠의 경쟁력을 높이는 데 도움이 된다. 최근 ‘Netfilix’에서 공개하여 전 세계적인 열풍을 불고 있는 ‘오징어 게임’ 또한 한글 글꼴을 활용하여 만든 오프닝으로 이목을 끌었다.  이렇게 디지털 시대에 글꼴의 영향력은 증가하고 있지만, 일반 사용자가 글꼴을 사용하기란 쉽지 않다. 글꼴은 저작물로 저작권의 보호를 받으며, 비영리∙영리에 따라 사용 범위도 다르다. 글꼴의 저작권을 침해한 자는 저작권법 제136조 제1항에 따라 5년 이하의 징역 또는 5000만 원 이하의 벌금에 처할 수 있다. 한국저작권위원회에 따르면 2019년 글꼴 파일 저작권 상담이 3,886건으로 글꼴과 관련하여 상당히 많은 분쟁이 일어나는 것을 알 수 있다. 해당 분쟁에서 글꼴의 저작권 합의 수단의 하나로 해당 글꼴이 포함된 패키지를 구매하는 것이 있는데, 폰트 한두 개를 사용하였지만 수십만 원에서 수백만 원에 달하는 글꼴 패키지를 구매해야 합의를 해준다는 조건을 내 거는 경우도 적지 않다. 글꼴은 종류도 다양하고 개인부터 기업까지 제작 업체도 달라서 저작권에 대한 전문지식이 없는 경우 어떤 것이 저작권 침해에 해당하는지조차 파악이 어려운 실정이다. 글꼴의 정보가 정확하게 알려지지 않아 일반 사용자는 글꼴의 저작권 관련 내용을 찾는 것도 어렵다. 또한 사용자의 마음에 드는 글꼴을 찾았다 할지라도 사용자의 의도에 따라 사용할 수 없으면 다시 글꼴을 찾는 힘을 들여야 하며 비슷한 글꼴을 다시 찾는 일도 쉽지 않다.  따라서 이런 문제를 해결하기 위해 한글 글꼴을 분석하여 데이터베이스를 구축하고, 사용자가 사용할 글꼴 파일을 입력하면 사용 범위와 가격 등 글꼴 정보를 볼 수 있도록 한다. 또한 사용자가 등록한 글꼴 파일과 비슷한 글꼴을 추천하여 일반 사용자가 글꼴에 쉽게 접근할 수 있도록 한다. |

|  |
| --- |
| **5. 본론**  글꼴 분석을 위해서는 글꼴의 데이터베이스를 구축하는 일이 선행 되어야 한다. 글꼴의 DB를 구축할 때에는 글꼴을 분류∙검색을 위해 특징 추출을 해야 한다. 해당 글꼴로 만들 수 있는 모든 글자에서 특징을 추출하면 정확도가 높은 DB를 만들 수 있다. 하지만 위치에 따라 글꼴 모양이 달라지지 않는 영어나 숫자와 다르게, 한글 글꼴은 2,350자를 분석해야 모든 정보를 알 수 있다. 따라서 대부분의 한글 요소를 포함하고, 형태 변별 요소가 많은 글자를 대표 글꼴로 선정하여 글꼴 분석에 할애하는 시간을 단축한다.  각 글꼴에서 글자 영상을 렌더링하여 특징을 추출하면 노이즈나 왜곡이 발생하지 않아 특징 추출에 용의하다. 렌더링 한 영상에서 특징을 추출하는 기술로는 HOG(Histogram of Oriented Gradients) 특징을 사용한다. HOG는 대상 영역을 일정 크기의 셀로 분할하고, 셀마다 edge 픽셀(밝기값이 일정 값 이상인 픽셀)들의 방향에 대한 히스토그램을 구하고 이들 값을 일렬로 연결한 벡터값으로, edge의 방향 히스토그램이라 할 수 있다. 이 기술을 사용해 고유한 윤곽선 정보를 갖는 글꼴의 특징을 추출할 수 있다. 특징을 추출한 파일을 글꼴 DB에 저장할 때 파일 이름, 글꼴 이름을 저장하고, 해당 글꼴의 저작권 관련 사항과 같은 정보를 함께 출력할 수 있도록 글꼴 정보도 함께 저장한다.  사용자가 등록한 글꼴 파일과 비슷한 글꼴을 추천하기 위해서는 DB의 데이터들을 일정한 기준으로 분류해야 한다. 데이터를 분류하는 기술로는 계층적 군집화(Hierarchical Clustering)를 사용한다. 계층적 군집화는 계층적 트리 모형을 이용해 개별 개체들을 순차적, 계층적으로 유사한 개체를 통합하여 군집화를 수행하는 알고리즘이다. 이 기술을 수행하려면 모든 개체 간의 거리와 유사도가 필요하다. 따라서 두 관측치 사이의 최단거리를 계산하는 유클리디안 거리(Euclidean Distance)로 거리를 계산하고, 두 벡터의 상관계수를 계산하는 피어슨 유사도(Pearson Similarity)로 유사도를 계산한다. 계층적 군집화를 수행하고 결과를 시각화하는 덴드로그램(Dendrogram)을 생성하여 하나의 군집에 1~7개의 개체가 포함되도록 트리를 잘라 전체 데이터를 군집으로 나눈다. 이렇게 나눈 군집을 통해 사용자가 등록한 글꼴이 포함된 군집을 출력하여 사용자에게 비슷한 글꼴을 추천할 수 있다.  위의 과정을 수행하고 나면 글꼴 DB와 데이터 군집화까지 사용자가 글꼴 파일을 입력하기까지 선행되어야 할 것들은 완료되었다. 이후 사용자가 정보를 알고 싶은 글꼴 파일을 입력하면, DB 구축에서의 특징 추출과 같이 대표 글꼴 영상을 렌더링 하고, HOG 기술을 수행하여 영상의 특징을 추출한다. 추출한 특징을 DB에 저장된 기존 데이터와 특징을 비교하여 같은 특징인 글꼴의 파일 이름, 글꼴 이름, 글꼴 정보 데이터를 출력한다. 또한 입력한 글꼴의 데이터 군집 위치를 확인하여 해당 데이터가 속해 있는 군집의 다른 글꼴 데이터를 출력해 비슷한 특징을 가진 글꼴을 추천한다.    (그림 3) 데이터 군집화  (그림 4) 글꼴 검색 순서도  (그림 2) DB 구축 |

|  |
| --- |
| **6. 결론**  이번 보고서에서는 HOG 기술을 사용하여 글꼴 영상의 특징을 분석하여 글꼴 데이터베이스를 구축하고 이 DB에 특징 데이터와 파일 이름, 글꼴 이름, 글꼴 정보를 함께 담아 사용자의 정보 검색에 용이하게 하는 시스템과 데이터를 계층적 군집화 기술을 수행해 군집 형성하여 사용자가 입력한 글꼴과 비슷한 글꼴을 추천하도록 하는 시스템을 제안했다.  이후에는 딥러닝 기술을 사용하여 DB를 구축하는 시간을 줄이고, 대표 글꼴의 특징 만 추출하는 것이 아닌 한글 글꼴 모두의 특징을 추출할 수 있게 하여 DB의 정확도를 높이고 싶다. 또한 현재는 사용자가 입력하는 글꼴 파일을 가지고 있어야 하지만, 텍스트가 포함된 이미지에서 글꼴을 인식하고 특징을 추출하여 사용자가 이미지 속 텍스트에 사용된 글꼴의 종류를 알 수 있도록 시스템을 개발하고 싶다. |

**7. 출처**

[1] 김두식 외 (2017), 한글 폰트의 품질 검사 및 폰트 정보 검색을 위한 영상 인식 기술 개발, 한국전자통신연구원.

[2] 전자연 외 4인 (2020), 글꼴 유사도 판단을 위한 Faster R-CNN 기반 한글 글꼴 획 요소 자동 추출, 한국미디어학회논문지, 제23권 8호, p953-964.

[3] 박문호 외 3인 (1997), 인공지능 : 인쇄된 한글 문서의 폰트 인식, 정보처리학회논문지, 제4권 8호, p2017-2024.

[4] 심성하, 이강배, 박성호 (2020), 딥 러닝 기법을 활용한 이미지 내 한글 텍스트 인식에 관한 연구, 한국융합학회논문지, 제11권 6호, p1-6.