캡스톤 디자인1 주제 선정 보고서

2015104194 이규호

2014104112 신주민

1. 요약

프로젝트는 유튜브(주로 강의 영상)에 업로드 되는 자막을 가져와서 이를 형태소 단위로 나누고, graph centrality를 텍스트에 적용한 textrank 알고리즘을 활용하여 요약하여 사용자에게 제공하는 서비스입니다. 서비스는 웹으로 보여집니다.

1. 선정 이유

2010년 이후 youtube는 아주 발전 가능성이 높은 컨텐츠이며, 현재도 많은 발전이 이루어지고 있습니다. 또한 한국어 텍스트의 요약 알고리즘은 영어에 비해 발전도가 많이 더딥니다. 저희는 이 두 가지 점을 캐치하여 많은 사람들이 유튜브를 통해 새로운 지식을 획득하는데, 더 효과적으로 돕기 위해서 해당 프로젝트를 선정했습니다.

1. 개요

자막은 savesubs를 통해 자동으로 추출합니다. 추출된 자막을 soyNLP를 통해 형태소 단위로 나눕니다. 나누어진 형태소들을 하나의 node로 생각하고, 문장 내에서 함께 나온 단어들을 연관성이 있다고 판단하고, edge로 연결합니다. 이를 이용하여 graph centrality 를 분석합니다. 어떤 단어가 중요한 단어인지 파악하고, 그 단어를 기반으로 텍스트들을 조합하여 텍스트 요약 알고리즘을 설계합니다. 해당 서비스는 간단한 웹사이트를 개발하여 사용자들이 이용할 수 있도록 합니다.