OOP Project 10조 요약본

팀명: C\_oala

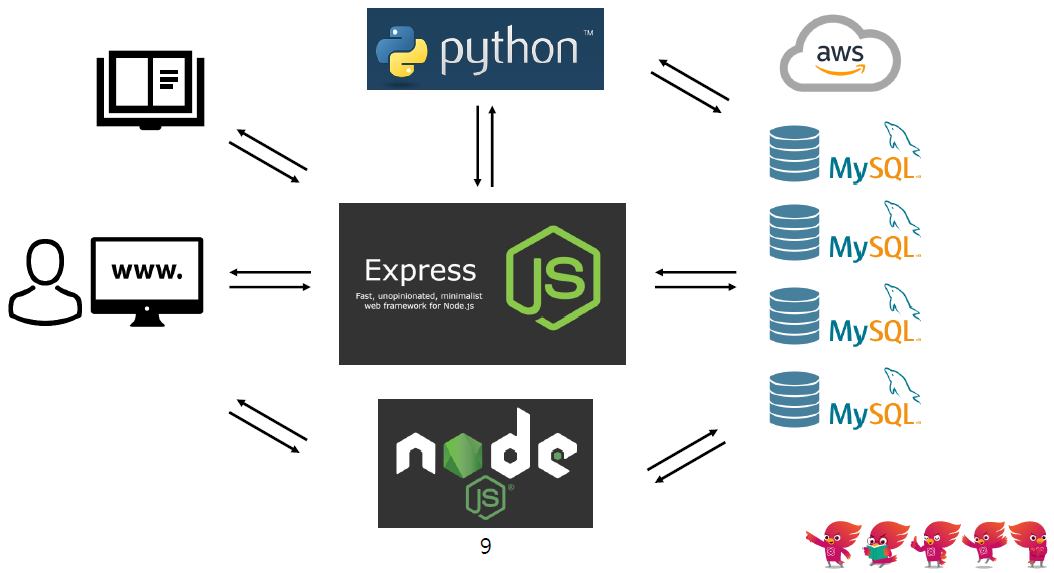
팀원: 김찬영, 김진모, 엄태규, 이상윤, 최서영, 하경민

주제: 텍스트 마이닝을 통한 논문 데이터 제공 대시보드 개발

1. **프로젝트 개요**

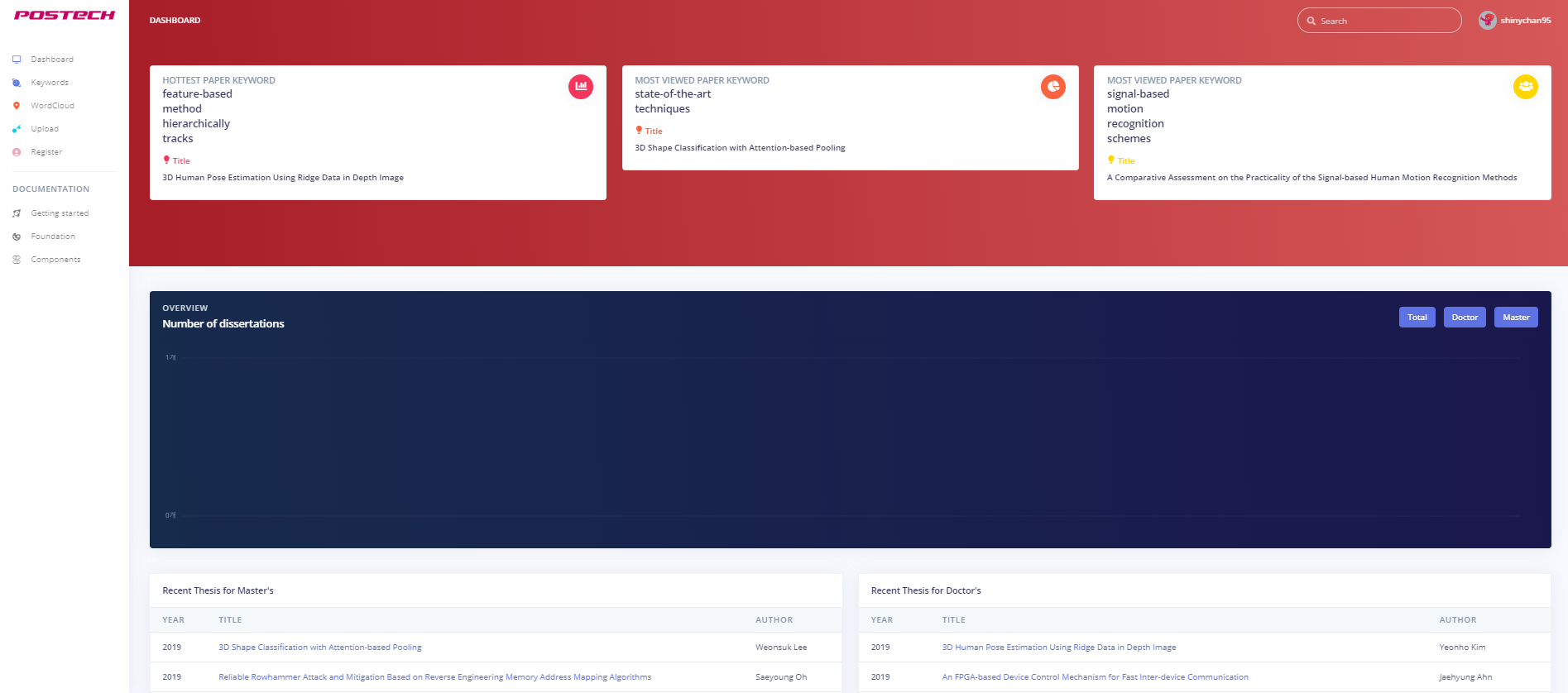
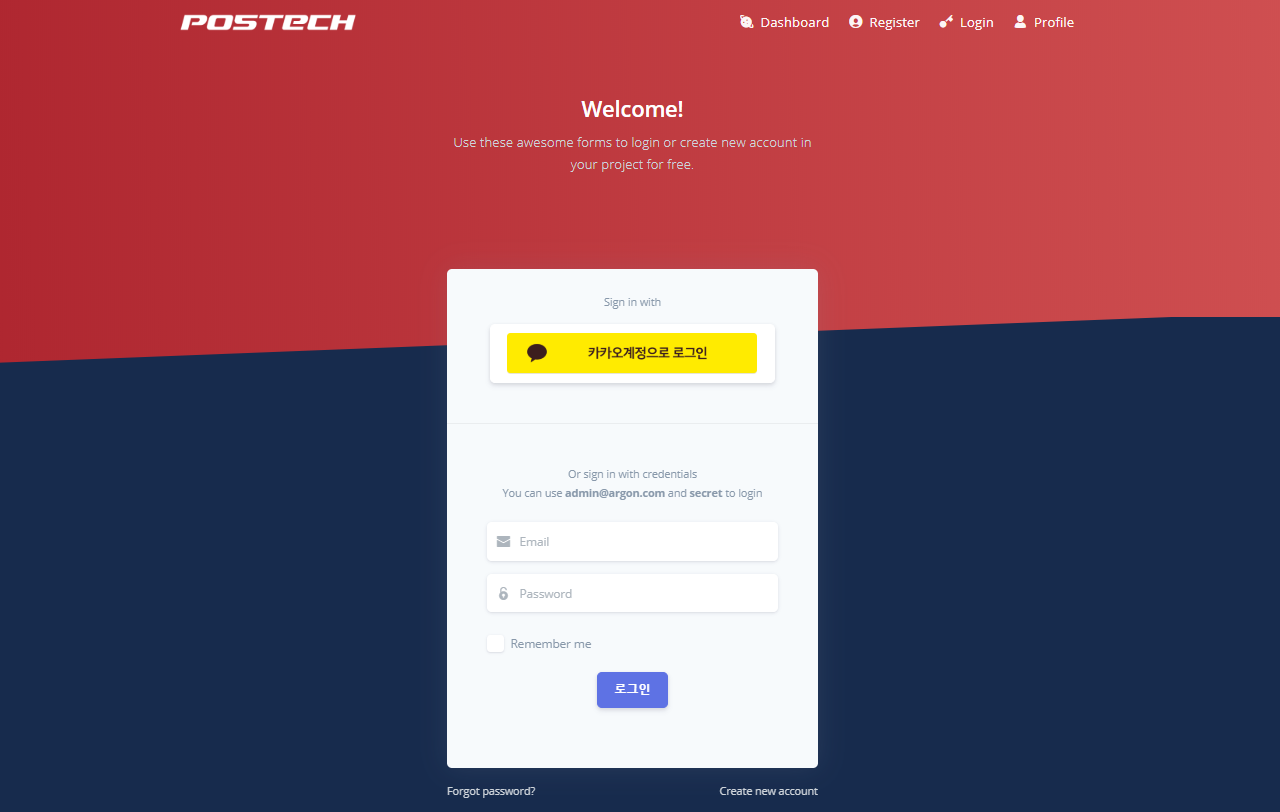
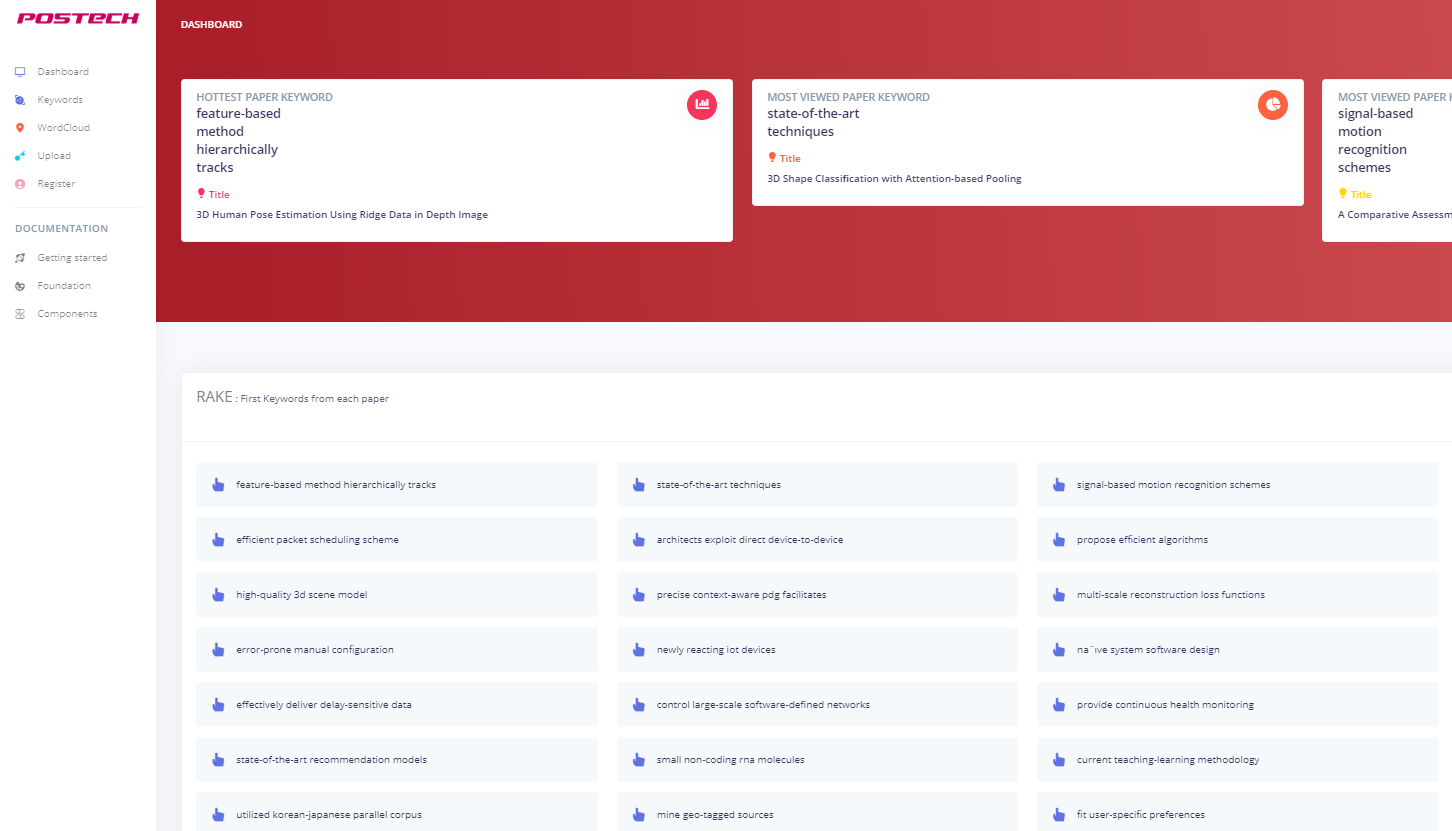
* OOP 개념을 활용하여 사회에 기여할 수 있는 응용 프로그램을 만드는 것을 주 목적으로 한다. 그 기여로써 대상을 고등학생으로 지정하고, 고등학생들이 포스텍에 대한 정보를 얻는데 있어서, 실제 논문 데이터를 통해서 요약되면서 핵심이 될 수 있는 키워드를 제공하는 대시보드를 구현하였다.

1. **프로그램 개요**



* Node.js를 통해 웹 서버를 운영하며, DB와의 연동을 통해 데이터를 실시간으로 가져와 시각화 해준다. 키워드만 제공하는 것이 아니라 그래프 및 Word Cloud 등 데이터 제공에 있어서 다양한 기법과 알고리즘을 적용해보았다. 또한 로그인 기능을 통해 사용자가 관심있게 본 논문을 파악하여 대시보드 너머의 기능을 수행한다. 추가로 사용자가 PDF를 웹 페이지를 통해 입력할 때, 데이터가 자동으로 추가되어 웹 페이지 내 내용이 업데이트 되도록 하여 동적인 서비스로 구성하였다.

1. **UI 설계**

* 위 사진과 같이 로그인 화면, 메인 화면, keyword 페이지, word cloud 페이지로 구성하여, 다양한 정보를 제공할 수 있도록 하였다.

1. **결과 및 한계**
   1. 결과적으로 다양한 알고리즘을 통해 키워드를 뽑아보고 이를 분석하고 최적으로 다듬어 DB에 저장하였고, 이 데이터를 실시간으로 제공하는 대시보드를 완성할 수 있었다. 완성할 수 있었던 것에는 알고리즘, 프론트 엔드, 백 엔드 개발에 있어서 모두 열심히 기여한 덕이 크다.
   2. 알고리즘의 경우 TF, TF-IDF, RAKE, CRF, LDA 등 텍스트 마이닝으로 유명한 알고리즘의 대다수를 사용해보면서 각 알고리즘의 특징을 익히고 이를 적용했을 때 어떤 결과가 나오는지 분석할 수 있었다. 이를 통해 논문의 abstract만으로 분석했을 때, RAKE와 CRF가 적절한 알고리즘이라고 판단하였다.
   3. 백 및 프론트 개발의 경우, 모두 익숙하지 않은 자바스크립트와 Node.js라는 언어와 프레임워크를 사용하여 어려움이 많았다. 특히 기능 구현에 대한 여러 모듈 사용에 대한 정보가 부족했던 것 그리고 비동기 처리에 대한 이해없이 개발을 시작했다는 것이 큰 어려움으로 작용하였다. 하지만, 노력 끝에 이를 극복해내어 기능을 구현할 수 있었다.
   4. 하지만, 프로젝트의 한계로 사용자의 실제 사용에 있어서 정보의 접근성 및 사용자 피드백을 통한 수정이 이루어지지 않은 점이 있다. 더 나은 시각화 및 데이터 제공 방식에 대해 고민한다면 더 좋은 서비스가 될 수 있을 것이라 생각한다. 또한 구현에 초점을 두어 대시보드에 적절한 기술과 프레임 워크를 사용했는지 의문을 가지며 개발적으로 학습이 필요함을 절실히 느꼈다.