



# 한이음 파이팅

안녕하십니까.

저희 주제는 빅데이터 영어 뉴스학습 애플리케이션입니다.

---

## 목차는

프로젝트 개발동기

주요 기술 설명

기대효과 순으로

말씀드리겠습니다.

---

## 프로젝트 개발동기

영어뉴스를 활용한 학습은 다양한 장점이 있습니다

첫번째. 다양한 표현법

매일 새로운 이슈를 통해 다양한 표현법을 익힐 수 있습니다.

같은 사건이나 이슈가 조금씩 변형되어 나오기 때문에

표현력을 높이는데 효과적입니다.

두번째. 글의 흐름 파악

기사에는 전형적인 흐름 반복이 있습니다.

주기적으로 기사를 읽다 보면 세밀한 계획을

세우지 않아도 저절로 복습을 하는 효과가 있습니다.

세번째. 트렌디한 어휘

사람들의 입에 자주 오르내리는

최근 이슈를 자연스럽게 접하여

각 분야에 자주 사용되는 필수 어휘를 배울 수 있습니다.

네번째. 배경 지식 증가

꾸준하게 기사를 읽으면 배경지식이 저절로 쌓이고

같은 사건에 대한 다양한 표현을 자연스럽게 습득하게 됩니다.

이러한 장점에도 불구하고  
단지 어렵다는 이유로  
뉴스를 통한 학습을 꺼려하고 있습니다.

---

### 두번째는 강세의 중요성입니다

같은 단어나 문장을 잘못된 강세로 발음하면,  
원어민 입장에서 인식률이 최대 10% 가량 떨어지는 것으로 조사되었습니다.

영어는 강세 언어입니다.  
강세란 한 단어의 특정한 음절을  
다른 음절보다 강하게 발음하는 것인데  
이를 제대로 표현하지 못하면  
아무리 발음이 좋아도  
원어민이 제대로 알아 듣지 못합니다.

이를 바탕으로 우리 팀은 영어 뉴스를 읽으며 얻는 장점들에 더해  
어휘 능력을 위한 강세 학습까지 사용자에게 제공하는 본 프로젝트를 구현하였습니다.

이를 바탕으로 우리 팀은  
영어 뉴스를 읽으며 최신이슈와 영어를 동시에 습득하며, 빅데이터를 활용한 자동 문제 생성 기능과 어휘 능력  
을 위한 강세학습까지 사용자에게 제공하는 본 프로젝트를 구현하였습니다.

---

## 주요 기술 설명

- 시스템 구성도

웹서버는 사용자의 접근이 용이하도록 아파치 웹서버와  
다양한 디바이스 활용이 가능한 HTML5 반응형 웹으로 개발하였습니다.  
배포를 위해 Amazon Web Service도 이용하였습니다.

웹 애플리케이션 서버는  
뉴스 사이트에서 Jsoup 기술로 웹 크롤링하여 영어 뉴스를 수집하는 수집 영역  
CoreNLP를 활용하여 영어 문장과 어휘를 추출하는 자연어 처리 영역  
주요 어휘를 뉴스 데이터와 비교하고 유용한 문장 제공을 위한 빅데이터 분석 영역  
웨이브 넷 기반의 음정 알고리즘을 이용한 음정 분석 영역 으로  
구성되어 있습니다.

DB서버는 데이터의 특성에 따라 MySQL과 MongoDB로 나누어 저장됩니다.

이어서 프로젝트에 사용된 주요 알고리즘을 소개하겠습니다.

- 자연어 처리 알고리즘

CoreNLP를 사용하여 만든 자연어 처리 알고리즘은 영어 뉴스의 원문을 문장과 어절, 어휘의 원형 별로 자연어 처리합니다.

이후 자연어 처리한 각 문장의 원형과

본 앱이 보유한 주요 단어 데이터를 비교하여

자동으로 유용한 문제를 생성합니다.

- 빅데이터 분석 알고리즘

빅데이터 분석 알고리즘은 웹크롤링한 뉴스 데이터를 데이터베이스의 모든 뉴스 데이터와 비교하여

제일 많이 겹치는 단어의 빈도에 따라 단어에 중요도를 부여합니다.

제일 높은 중요도를 부여받은 단어 순으로 사용자에게 제공하여 유용한 단어를 학습할 수 있게 합니다.

- 음성 분석 알고리즘

음성 분석 알고리즘의 수행 과정은 분석, 정규화, 보간으로 이루어져 있습니다.

녹음된 음성을 ParseMouth로 목소리를 분석하고 정규화하여 음정을 가공합니다.

이후 비정상적인 수치인 음정의 아웃라이어를 제거하고 음정의 빈 구간이 존재한다면 선형보간법으로 처리합니다.

마지막으로 원어민의 음성과 비교하여 결과를 출력합니다.

- 음성 분석 알고리즘

음성 분석 알고리즘은 파이썬 플라스크를 활용하여 만들었습니다.

음성을 녹음하였을 때 간헐적으로 생기는 치찰음이 모음으로 인식되는 것을 방지하고자

비정상적인 수치의 아웃라이어가 발생한 음높이의 결과를

중앙값과 비교, 정해진 수식에 따라 판단 후 제거하여

보다 높은 정밀한 음성 분석이 이루어지게 합니다.

---

## 기대효과(30초) (비용, 강세, 뉴스)

기대효과는 크게 세가지입니다.

첫번째. 자동으로 생성된 문제는 제작 및 관리의 비용을 절감시켜

저렴한 서비스 제공을 통해 **저소득층**의 접근을 높일 수 있습니다.

두번째. 발음 및 강세를 분석 후 시각화하여

어휘 능력이 필요한 사용자에게 정확한 발음 및 강세의 교정을 제공합니다.

마지막으로, 당일 뉴스의 메인 토픽과 관련된 영어 표현 및 어휘를 학습하여 영어 면접이 필요한 취업 준비생 및 이직자들의 어학능력에 도움을 줍니다.

시연을 통해 보여드리겠습니다.

---

로그인 화면)

빅데이터를 활용한 영어학습 애플리케이션 THE를 소개합니다.

메인화면)

본 페이지는 메인 화면으로,

각 기사의 타이틀을 보여주고, 사용자가 뉴스사를 선택할 수 있게 합니다.

— 뉴스사 넘김 넘김 넘김

— 뉴스사 클릭

오늘의 뉴스)

뉴스사를 선택하면 해당 사이트의 기사를

데이터베이스에서 가져와 사용자에게 제공합니다.

현재 2011년 부터 2020년도의 뉴스 원문 데이터를 보유하고 있습니다.

— next

오늘의 문장)

본 페이지는 문장을 학습하는 페이지입니다.

CoreNLP로 자연어 처리한 단어들을 WordPool과 비교하여

유용한 문장을 사용자에게 번역과 함께 제공합니다.

— next

오늘의 문제)

본 페이지는 앞서 학습한 문장과 더불어 유용한 단어를 학습할 수 있는 화면입니다.

문장에서 가장 중요한 단어를 빈칸으로 처리하여 문제의 답을 입력하게 합니다.

— 답 입력

답을 입력했을 때, 틀린 답과 알맞은 답을 다르게 표시합니다.

— next

오늘의 녹음)

본 페이지는 강세학습 화면입니다.

앞서 학습한 문장의 정답을 강조하고

녹음 전 원어민의 발음과 강세를 학습할 수 있습니다.

— voice버튼 클릭(끝나고 바로 녹음 시작)

—녹음 시작

사용자에게 받은 음성 데이터는 플라스크에 구현된 음성분석알고리즘을 통해 평가 하게 됩니다.

— 강세 평가

녹음된 자신의 목소리를 들을 수 있으며,

원어민의 목소리와 비교할 수 있습니다.

— 내 목소리 중간까지 듣고 원어민 목소리 듣기 클릭, 원어민 목소리 재클릭(전체 목소리 출력)

사용자는 자신과 원어민의 강세 중 어느 부분이 틀렸는지 음성과 그래프로 확인할 수 있습니다.

— 점수 보여주기

하단에는 녹음된 음성과 원어민의 음성을 비교하여

강세의 유사성, 강세의 명확성, 발음의 정확성을 분석하여 점수형태로 사용자에게 제공합니다.

— 번역하기 클릭

오늘의 번역)

본 페이지는 번역 학습을 제공합니다.

— 조금 입력

번역의 모범 답안과 사용자가 입력한 문장을

형태소 단위로 분석 후 비교하여 유사도를 보여 줍니다.

— 전부 입력

이러한 학습이 한 사이클로 진행되며, 추가적으로 문제를 더 풀 수 있습니다.

— 문제 더 풀기 클릭

— 단어장 클릭

단어장)

본페이지는 단어장을 제공하는 화면입니다.

— 뉴스사 클릭

사용자는 자신이 학습하고자 하는 사이트를 선택합니다.

데이터베이스에 선택한 사이트에 해당하는

단어 데이터의 뜻과 의미를 학습할 수 있습니다.

— 전부 넘어감

앞선 단어들 중 틀린 단어들을 다시 학습할 수 있습니다.

— 틀린 단어 학습하기

복습하기)

— 복습하기 클릭

마지막으로 복습하기 화면입니다.

— 뉴스사 클릭

뉴스사를 선택하면, 사용자가 최근 학습한 뉴스 기사를 복습할 수 있습니다.

이상으로 시연을 마치겠습니다. 감사합니다.