

**Python**

**Production of programs  
that analyze Korean PDF  
and extract keywords**

**Proposal**

Date : 2023.11.03

Name : Haegeon Lee

ID : 183014

# **1. Introduction**

## **1) Background**

With the advent of the information age, the era of reading and interpreting a large amount of documents has arrived. If the amount of documents is vast when a particular document is first encountered, fatigue accumulates before reading. Therefore, if the Python program provides keywords that readers should focus on first, the speed increases rapidly and more efficient reading is possible. Therefore, the existence of these programs is necessary.

## **2) Project goal**

It aims to develop a program that analyzes pdf to analyze what key keywords are based on the frequency of words and provides them to users.

## **3) Differences from existing programs**

Existing programs are produced based on English, so there is a big difference in reading Korean pdf and it is impossible to use. Therefore, there is a difference from existing programs because we focus on extracting keywords from Korean pdf file for Koreans.

# **2. Functional Requirement**

## **1) Function 1 – Extract text from file**

- Extract the text file from the pdf file and store it in the list.

## **2) Function 2**

- Remove investigations such as '은','는','이','가' from list token. And import

Morpheme analyzer to make up a list with only nouns.

### 3) Function 3

- Saves the frequency of nouns in dictionary form and outputs the most written value.

### 3. Schedule

	11/3	11/10	11/17	11/24	12/1	12/15	12/22
Write Proposal	----->						
Function 1		----->					
Function 2			----->				
Function 3					----->		