# CHAPTER I

# Introduction

## 1.1 Background

Nowadays, Korean culture becomes very popular among the Indonesian young people. There are a lot of Korean things that spread out in Indonesia. This might affects their habits. Many young people who act like Korean actress try to imitating from fashion, speech, and hairstyle. However, they just want to learn Korean culture of history, traditional culture, and language. If we talk about Korean language, we could not forget the Korean Alphabet. It is due to Korean has their own alphabet.

The Korean alphabet, also known as Hangul, or Chosongul (officially transcribed Han-geul in South Korea and Chosŏn'gŭl in North Korea), is the native alphabet of the Korean language. It was created during the Joseon Dynasty in 1443. Now, it has been being the official script of both South Korea and North Korea, and co-official in the Yanbian Korean Autonomous Prefecture of China's Jilin Province. Hangul is a featural alphabet of 24 consonant and vowel letters. However, instead of being written sequentially like the letters of the

Latin alphabet, Hangul letters are grouped into blocks, such as 한 han, each of which transcribes a syllable. Although the syllable 한 han may look like a single character that actually composed of three letters: ㅎ h, ㅏ a, and ㄴ n. Each syllabic block consists of two to five letters, including at least one consonant and one vowel. These blocks then arranged horizontally from left to right or vertically from top to bottom. The number of mathematically possible blocks is 11,172, though there are far fewer possible syllables allowed by Korean phonotactics, and not all phonotactically possible syllables occur in actual Korean words. For a phonological description, we could see the Korean phonology.

With the explanation above, people who used Latin letter learning the Korean Alphabet is not easy, so it is necessary to the existence of technology such as Optical Character Recognition (OCR) that could help people immediately reading the Korean Letter.

Optical Character Recognition, usually abbreviated to OCR, is the mechanical or electronic conversion of scanned images of handwritten, typewritten or printed text into machine-encoded text. It is widely used as a form of data entry from some sort of original paper data source, whether documents, sales receipts, mail, or any number of

printed records. This is a common method of digitizing printed texts so that they can be electronically searched, stored more compactly, displayed on-line, and used in machine processes such as machine translation, text-to-speech and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

Early versions need to be programmed with images of each character, and worked on one font at a time. The "Intelligent" systems with a high degree of recognition accuracy for most fonts are commonly now. Some systems are capable of reproducing formatted output that closely approximates the original scanned page including images, columns and other non-textual components.

In this project, I have made an application that can help people who want to learn Korean Alphabet which implemented Optical Character Recognition. Using Java as programming language, this program needs JPEG file as input data, and after the image processed, as an output, it will create string of Latin Letter result.

## 1.2 Scope

This project created with Java language programming based on Heuristic Algorithm and Topological Skeleton Algorithm implementation. In data structure, this project using Linked List. This project is focused on detect Korean character picture from any size, any color, and tilt up to a certain limit. The problem limitations on this project can be explained as follows.

1. The System just approve the input of image that have JPEG formatted.
2. The character which want to recognize must be perpendicular, if it has tilt, the threshold is 1 degree.
3. If the image comes from the camera then the Brightness and Contrast must be maximum.
4. No big dust in this project.(Two or more colors background)
5. The number of Korean character that can be detected depends on database updated.

## 1.3 Objective

The purpose of this Project are :

1. To facilitate in the further research on Korean Dictionary based on Korean Optical Character Recognition (OCR).
2. To simplify people reading Hangeul (Korean Letter) by Optical Character Recognition (OCR) application.
3. To help people who want study more about Hangeul.