

I. INTRODUCTION

QURAN is the Moslem holy book that was revealed by Allah SWT to the Prophet Muhammad SAW. Every Moslem is obliged to read and understand the meanings of the Quran and apply it in life to be on the straight path. The Quran contains many historical stories, metaphors, and implied descriptions of the universe and its contents. Because of that knowledge, the Quran is much researched by everyone. The problem is that there is a lot of information contained in the Quran so that ordinary people have difficulty understanding the Quran in its entirety.

To overcome this problem, the author proposes a sequence chunking method to extract and provide important information about the Quran. Sequence chunking is a method to chunk sentences by gathering words that are considered important using neural networks. This group of words then become a phrase. This method is divided into two tasks: segmentation to identify the scope of the chunk explicitly and labeling to label each chunk as a unit based on the results of segmentation.

In this paper, the author will implement the Bidirectional Long Short-Term Memory (Bi-LSTM) model to chunk inside-outside-beginning (IOB) tags from the Quran [1]. This model will transform sequence chunking into a sequence labeling problem. The Bi-LSTM will be used to do both segmentation and labeling to correctly identified the tags. The predicted results will be then evaluated with the real tags to calculate system performance. The dataset used by the author is surah 1 (Al-Fatihah) to surah 5 (Al-Maidah) from Saheeh International Quran obtained from *quran.com*. The author chose the Saheeh International because it provides convenience to the reader and it brings closer to original meanings [2]. The dataset is still not in the appropriate form, therefore, some preprocessing needs to be done.