1. Introduction

Hadith is one of the guidelines of Muslims that is useful as a complement to the Qur'an which contains the words, behavior, and deeds of the Prophet Muhammad; it was conveyed through Sanad or a series of narrators' names [1]. Basically, the Hadith is written in Classical Arabic, which is a formal language used for more than 1500 years in Islamic teachings[2]. There are around 241,665 Hadiths which have been successfully recorded into 40 books. Each of these books contains a discussion of di erent things. Hadith has been translated into various languages including Indonesian, and the Hadith has undergone a transformation from the original handwriting on a paper into a digital text document. For Muslims the validity of a Hadith is very important, where it determines whether a Hadith can be applied as a guideline or not. The narrator's name in the Hadith is one of the most important components in determining the validity of a Hadith, which each narrator has a di erent level of credibility in conveying a Hadith, and the study aimed at this thing is known as ilm al-rijal (biographical evaluation; literal: knowledge of men) [3], but with the large number of Hadiths available, causing the process of manually determining the validity of a Hadith based on the narrator's name becomes di cult , especially in Indonesian Hadith translations.

Named Entity Recognition (NER) is a method that aims to find entities in a text document, in this case the entity includes the name of the person, location, organization, etc. Based on the techniques and approaches, NER can be divided into three main types, including rule-based, machine learning, and hybrid [4]. At the beginning of its develop-ment, most NER systems were built with a rule-based approach or by utilizing rules made manually by humans. This approach results in a relatively good performance but requires more e ort to achieve this. Therefore, the supervised-learning type of Machine Learning approach is one of the approaches that most often used to automatically induce rule-based systems[5]. The simple overview of the NER system includes: first, the system will receive input in the form of text documents such as telah menceritakan kepada kami Husain bin Muhammad berkata, telah menceritakan kepada kami Sulaiman bin Qarn (had told us Husayn ibn Muhammad said, had told us Sulayman bin Qarn), then from the text documents found several named entities, including Husain bin Muhammad and Sulaiman bin Qarn. The name entity that has been found can be used for further processing of text documents such as document indexing, which can help search for information based on the entity from the document.

This study will discuss the implementation of the Named Entity Recognition to the Indonesian translation of the Hadith collection to find the names of narrators from each Hadith. In this study 200 Hadiths from 9 di erent books and consisting of 2241 narrator names will be used as a dataset. Named Entity Recognition can be seen as a multi-class classification problem in which a Named Entity can consist of a series of words in which each word needs a label to indicate which entity belongs to the set of words[6]. With a variety of entity forms and the amount of data used, this study will use a supervised-learning approach, and to maximize performance from the NER system, Support Vector Machine (SVM) is chosen as a classifier model that is known to have good generalization capabilities in classifying data and ability to deal with high-dimensional data. The narrator name produced by the NER system will then be used as an index of the Hadiths that have been narrated by the narrator using the Inverted Index method.

────────