

## 1. Introduction

Hadith is a primary source of Islamic law after the Qur'an. As a primary source of law means every activity carried out must be in accordance with what is regulated in the Qur'an and hadith. But the existence of hadith that originated from the Prophet Muhammad was tainted by the appearance of weak and fabricated hadith where the origin and interpretation are used in various motives [1]. Believing and practicing sharia based on a weak or even fabricated hadith is an act that is condemned by Rasulullah SAW with the threat to take his seat in hell. Weak and fabricated hadiths with good or bad intentions remain counted as an act of counterfeiting so that there should be no action based solely on weak and fabricated hadith [2].

Weak and fabricated hadith is the lowest and the worst level of authenticity compared to other hadith. Fabricated hadith are hadith that are intentionally made or falsified while weak ones are hadith that are the result of unintentional forgery, occurring because there are errors in the chain of narrators of the hadith [3]. Examples of hadith that are often encountered lately are "Perselisihan di antara umatku adalah rahmat" (The difference of opinions between my people is a mercy for my people), whereas according to syarah Al-Albani this hadith has no source and because of that statement, many Muslims after the time of the Imams especially today continue to disagree and debate in many ways that involve a creed and practice, the statement is also contrary to the Qur'an Al-Anfal verse 46 "Dan taatlah kepada Allah dan Rasul-Nya dan janganlah kamu berbantah-bantahan, yang menyebabkan kamu menjadi gentar dan hilang kekuatanmu..." (And obey Allah and His Messenger, and do not dispute and thus lose courage and then your strength would depart, etc.). The verse makes it clear that disputes and animosity are not from God. Al-Albani also said in his syarah [4] that these words will negatively impact Muslims from time to time. Disputes caused by differences between mazhab have reached its climax, even the fanatical followers of the mazhab are not reluctant to disbelieve and branded the followers of other mazhab as heretics.

Currently, the only method of checking back the authenticity of a hadith is by asking a religious expert or by searching the hadith manually on physical books and electronic documents. Therefore, we need a system to search weak and fabricated hadith as a media to doublecheck whether the information obtained is based on weak and fabricated hadith and save ourselves from the threat of hellfire and religious debates resulting from taking weak and fabricated hadith as the basis of our action.

In recent years, the implementation of natural language processing in the Qur'an and hadith are used to conduct information retrieval [5]. Natural language processing can process and analyze information on documents so that it can be used to search whether a piece of text information that is said to be a hadith can be used as a query as a search parameter for a collection of Indonesian translated documents of weak and fabricated hadith.

There are several studies to conduct information retrieval. Research conducted by Vibkhav et al. [6] was conducted to analyze and explain how the vector space model works. Hanum et al. [7] conduct research on information retrieval in the form of halal product queries from Malay language documents using the latent semantic indexing method and generate an accuracy value of 86%. Jbara et al. [8] researched the classification of texts that existed in 1321 Sahih al Bukhari hadith by classifying classes that have the same topic then the testing is done by comparing with the similarity coefficient table and generate an accuracy of 73%. Humaini et al. [9] design an algorithm for Indonesian translated Qur'an based on the vector space model and TF-IDF which generate 98.70% accuracy.

In previous studies, the use of vector space models has the highest level of accuracy, which is 98.70%. The second best method uses latent semantic indexing with 86% accuracy and then the similarity coefficient table method with 73% accuracy. However, the three related studies use different data sets, research with latent semantic indexing using Malay language data sets, while vector space models and coefficient tables use Indonesian translated data sets. Therefore, the vector space model was chosen because the dataset used is Indonesian translated and is the best method for conducting retrieval documents [10].

Based on related research, this paper proposes to build a weak and fabricated hadith document retrieval system using vector space models and can correct spelling by using the spell checking method Symspell to calculate the increase in accuracy of document retrieval, because the test data used will be in the form of raw text from the internet that contains spelling errors. Vector space models work by representing text information in documents in a frequency index vector [6]. The spell checking method that will be used as a functional additive to improve the accuracy of document retrieval is the symmetric deletion spell checking algorithm or the Symspell method because Symspell can provide faster spelling correction than conventional methods like Peter Norvig [11].