

## **Abstrak**

*Part of speech tagging (POS tagging)* yang merupakan bagian dari *Natural Process Language (NLP)*. *POS tagging* merupakan suatu proses untuk melakukan pelabelan pada sebuah kata dalam suatu kalimat sesuai dengan kelas katanya secara otomatis. Terdapat bermacam macam metode *taggers* dalam *POS tagging*, masing masing metode *taggers* memiliki karakteristik tersendiri dalam penerapannya. Penelitian ini metode yang digunakan yaitu *Conditional Random Fields* dan *Hidden Markov Model*. Pelatihan kedua model metode tersebut menggunakan korpus bahasa Indonesia dan teks berita Bahasa Indonesia sebagai data uji untuk mengetahui metode mana yang paling efisien berdasarkan hasil dari nilai akurasi dan waktu pelatihan masing masing model. Metode yang memiliki nilai terbaik adalah metode CRF dengan nilai *Accuracy* 97.68 pada evaluasi data *test corpus* dengan waktu pelatihan 146.90s, kemudian ada metode HMM yang memiliki nilai *Accuracy* tertinggi dengan nilai 94,25% dan waktu pelatihan yang relatif lebih singkat yaitu 32.45s dan untuk kalimat sampel berita Bahasa Indonesia yang berjumlah 116 token, metode CRF menghasilkan sebesar 90.05% *Accuracy* lebih tinggi dari metode HMM yang menghasilkan 79.31% *Accuracy*.

Kata kunci : Part of Speech Tagging, Efisien, Conditional Random Fields, Hidden Markov Model

**Abstract**—Part of speech tagging (POS tagging) is a part of Natural Process Language (NLP). POS tagging is the process of automatic labeling of a word in a sentence according to the word class. There are various tagger methods in POS tagging, each tagger method has its own characteristics in its application. The research method used is Conditional Random Fields and Hidden Markov Model. The training of the two method models uses the Indonesian language corpus and Indonesian news texts as test data to determine which method is the most efficient based on the results of the accuracy and training time of each model. The method that has the best value is the CRF method with an accuracy value of 97.68 on the evaluation of the corpus test data with a training time of 146.90 seconds, then there is the HMM method which has the highest accuracy value with a value of 94.25 % and shorter training time relatively shorter at 32.45 seconds and for the sample Indonesian news sentences containing 116 tokens, CRF method produces 90.05% accuracy which is higher than the HMM method which produces 79.31% accuracy.

**Keyword:** Part of speech Tagging, Natural Process Language, Efficient, Conditional Random Fields, Hidden Markov Model