

## ABSTRAK

Ketersediaan informasi yang melimpah pada twitter menjadi masalah tersendiri jika informasi yang diterima sebenarnya tidak terlalu dibutuhkan. Untuk itu diperlukan teknik untuk memilah informasi yang dikenal sebagai ekstraksi informasi. Ekstraksi informasi ini berkaitan erat dengan *Named Entity Recognition* (NER). NER merupakan komponen dasar dari ekstraksi informasi yang berfungsi untuk mengenali entitas pada kumpulan teks berbahasa alami. Skripsi ini meneliti tentang NER pada twitter dengan metode perceptron untuk mengenali entitas nama orang dan lokasi. Pelaksanaan penelitian dimulai dari pembuatan data set, eksperimen, evaluasi hasil eksperimen dan penarikan kesimpulan. Eksperimen dilakukan untuk mendapatkan kondisi terbaik pada perceptron dalam mengenali entitas. Parameter yang diuji adalah nilai *threshold*, kelompok data set yang digunakan sebanyak tiga kelompok (nama orang, nama lokasi dan gabungan), jenis data set (*preprocessing* dan *non-preprocessing*), jumlah data *training* dan teknik pengujian. Hasil eksperimen memperlihatkan bahwa perceptron dapat mengenali entitas nama orang dan lokasi yang ditunjukkan dengan nilai *f-measure* yang baik pada setiap tipe entitas dengan memperhatikan parameter yang diuji.

Kata kunci: Ekstraksi Informasi, NER, Perceptron

## **ABSTRACT**

*The availability of twitter information that overflow, can cause problem if the information that received is not much needed. Therefore, the technique to sort the information is required, it known as information extraction. Information extraction is closely related to Named Entity Recognition (NER). NER is a basic component of information extraction, it function is to identify the entities the set of common language text. The core of the research on this thesis is to determine how perceptron method can be used to recognize the entity names and location on twitter. The implementation of research starts from generate of data sets, experiments, and evaluation of experimental result and conclusion. That experiment carried out to obtain the best condition of perceptron recognition entities. The parameters that tested is a value of threshold, group of data sets that used are (people name, location name, and combination of both), the type of data set (preprocessing and non preprocessing), the amount of data training and the technical testing. Result of the experiment shows that perceptron can recognize entities of people name, and location that shown but f-measure value, which can be good to all the type of entities with shown the parameters that tested.*

*Keywords: Information Extraction, Named Entity Recognition, Perceptron*