

Ekstraksi Informasi Beasiswa dari Media Sosial menggunakan BiLSTM-CRF

Muhammad Rizki Ramadhan Setiawan¹, Ade Romadhony², Hasmawati³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹mrizkiramadhans@students.telkomuniversity.ac.id, ²aderomadhony@telkomuniversity.ac.id,

³hasmawati@telkomuniversity.ac.id

Abstrak

Sosial media merupakan tempat dimana orang-orang berkumpul dan saling bertukar informasi. Dari informasi tersebut dapat muncul berbagai macam peluang seperti beasiswa yang dikeluarkan oleh lembaga pendidikan. Peluang ini dapat banyak ditemukan pada sosial media seperti Twitter. Namun kebanyakan informasi yang dikeluarkan menggunakan format tersendiri sehingga menjadi tidak terstruktur dan menghambat upaya pengolahan informasi yang terkait. Melihat cepatnya informasi berlalu dan banyaknya kompetisi dalam meraih peluang tersebut, efisiensi menjadi faktor penting dalam mengumpulkan dan memproses informasi. Untuk mengatasi permasalahan tersebut, maka dilakukan ekstraksi informasi untuk mengubah informasi tidak terstruktur menjadi terstruktur menggunakan metode *Bidirectional Long-Short Term Memory* dengan *Conditional Random Fields* (BiLSTM-CRF). Metode ini digunakan karena dapat memberikan konteks informasi dari masa lalu dan masa depan pada teks sehingga sesuai untuk mengatasi tugas ekstraksi informasi. Tujuan penelitian ini adalah melakukan ekstraksi informasi dengan mengimplementasikan model BiLSTM-CRF untuk melakukan proses klasifikasi informasi yang diekstraksi sesuai dengan kategori pengelompokkan yang ditetapkan sehingga data yang terkumpul menjadi terstruktur dan mudah untuk dibaca. Hasil yang didapatkan dari implementasi model tersebut adalah nilai performansi dengan *Precision* 90%, *Recall* 51%, dan *F1-Score* sebesar 54%.

Kata kunci : beasiswa, twitter, *sequence labelling*, *BiLSTM-CRF*, ekstraksi informasi

Abstract

Social media is a place where people gather and exchange information. From this information, various opportunities can emerge, such as scholarships issued by educational institutions. This opportunity can be found on social media such as Twitter. However, most of the information released uses its own format so that it becomes unstructured and hampers the processing of related information. Given the speed at which information passes and there is a lot of competition for these opportunities, efficiency is an important factor in gathering and processing information. To overcome this problem, information extraction is carried out to change unstructured information into structured using the Bidirectional Long-Short Term Memory method with Conditional Random Fields (BiLSTM-CRF). This method is used because it can provide context of information from the past and future in the text so that it is suitable for solving the task of extracting information. The purpose of this research is to extract information by implementing the BiLSTM-CRF model to classify the extracted information according to the defined grouping category so that the collected data becomes structured and easy to read. The results obtained from the implementation of the model are performance values with 90% Precision, 51% Recall, and 54% F1-Score.

Keywords: scholarship, twitter, sequence labeling, BiLSTM-CRF, information extraction
