ABSTRAK

Frasa kunci adalah gabungan kata yang mewakili konsep atau garis besar dari suatu dokumen. Frasa kunci digunakan untuk membantu pembaca dalam mengetahui pokok bahasan dari dokumen. Sayangnya terdapat publikasi ilmiah yang memiliki frasa kunci yang tidak relevan terhadap isi dari dokumen atau tidak memiliki frasa kunci. Berdasarkan permasalahan tersebut maka dalam tugas akhir akan dibuat sistem yang dapat melakukan ekstraksi frasa kunci pada publikasi ilmiah secara otomatis dari pdf. Dalam menentukan frasa kunci pada dokumen, akan diusulkan untuk menggunakan pembobotan tf-idf dan deep belief network sebagai metode pembelajaran dengan nilai sentimen sebagai salah satu fitur pembelajaran. Selain nilai sentimen, akan digunakan posisi section sebagai fitur pembelajaran. Posisi section akan ditentukan dengan menggunakan karakteristik font. Deep belief network diusulkan untuk mengetahui efek dari penggunaan deep learning terhadap ekstraksi frasa kunci. Seluruh pengujian yang dilakukan akan menggunakan dataset milik NUS terkait publikasi ilmiah dengan judul "Keyphrase Extraction in Scientific Publications". Berdasarkan hasil penelitian didapat hasil bahwa penggunaan deep belief network akan menghasilkan model pembelajaran dengan akurasi yang lebih tinggi dibandingkan dengan menggunakan regeresi logistik sebesar 4,33%. Penggunaan analisa sentimen sebagai fitur pembelajaran dapat memberikan peningkatan akurasi terhadap model pembelajaran sebesar 4,17%. Sistem ekstraksi frasa kunci yang dibagun menghasilkan f-measure sebesar 13,22%

Kata kunci: Deep Learning, Deep Belief Network, Ekstraksi Frasa Kunci, Fitur Sentimen, Pemrosesan Dokumen, Tf-Idf

ABSTRACT

Kerphrases are combination of words which represent concept or main idea in document. Keyphrases are used to aid reader's understanding regarding to main topic in document. Unfortunately, there are scientific publications which have keypharse that doesn't represent content of document or have no keyphrase at all. Based on the problem, in this work will be built an automatic keyphrase extraction system for scientific publication in pdf format. In order to determine keyphrases, proposed to use TF-IDF weighting and deep belief network as learning method with sentiment value as one of the learning feature. Besides sentiment value, will be used section position as learning feature. Section position will be determined using font characteristics. Deep belief network is proposed in order to find out the effect of using deep learning in keyphrase extraction. The entire testing conducted will use dataset belongs to NUS regarding scientific publication titled "Keyphrase Extraction in Scientific Publications. Based on result, using of deep belief network will bring higher accuracy for learning model compared of using logistic regeresion in 4.33%. The use of sentiment analysis also gives enhancement to the accuracy of learning model in 4.17%. Proposed keyphrase extraction system has 13.22% for f-measure in top-5.

Keywords: Deep Learning, Deep Belief Network, Document Processing, Keyphrase Extraction, Sentiment Feature, Tf-Idf

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR FORMULA

DAFTAR NOTASI/ LAMBANG

Jenis	Notasi/ Lambang	Nama	Arti
Use Case		Aktor	Menggambarkan aktor atau pengguna aplikasi.
Use Case		Case	Menggambarkan proses atau aksi yang dapat dilakukan oleh aktor pada aplikasi.
Use Case		Association	Menggambarkan komunikasi antara *use case* dan aktor yang berpartisipasi (asosiasi).
Activity Diagram		Initial Node	Menandakan dimulainya aktivitas pada sebuah sistem.
Activity Diagram		Activity	Menandakan aktivitas apa yang akan dilakukan oleh pengguna aplikasi.
Activity Diagram		Final Node	Menandakan akhir aliran proses sistem

DAFTAR SINGKATAN

ANN	Artificial Neural Network
DNN	Deep Neural Network
MLP	Multilayer Perceptron
PoS	Part of Speech
RBM	Restricted Bolztman Machine
TF-IDF	Term Frequency–Inverse Document Frequency
UML	Unified Modelling Languange