ABSTRAK

Ekstraksi informasi dari poster acara merupakan sebuah tantangan karena dokumen ini memiliki tata letak yang bervariasi dan mengandung elemen teks serta visual yang saling berintegrasi. Pendekatan yang ada sekarang dinilai kurang fleksibel terhadap variasi format tata letak dokumen yang bergaya. Oleh karena itu, penelitian ini bertujuan mengembangkan sistem ekstraksi informasi jadwal dari poster acara seminar dengan memanfaatkan pemahaman multimodal yang mampu beradaptasi pada variasi tata letak yang ekstrim.

Metode *pre-processing* yang digunakan meliputi normalisasi gambar hingga ukuran 1000 piksel, dilanjutkan dengan pengenalan teks menggunakan Google Cloud Vision API untuk mengekstraksi teks poster pada level kata, kemudian dilakukan pelabelan berdasarkan informasi jadwal acara yang selanjutnya disimpan ke Google Calendar. Untuk kebutuhan pelabelan yang lebih detail, label dipecah menggunakan skema BIO agar entitas yang berbeda tetapi memiliki label sama dapat dibedakan, dengan proses pelabelan BIO didasarkan pada hasil pengenalan teks Google Cloud Vision API pada level paragraf. Selanjutnya, dilakukan *fine-tuning* model LayoutLMv3 untuk memprediksi label kata dengan mempertimbangkan kombinasi informasi teks, tata letak, dan elemen visual, yang mencakup tahap inisialisasi model *pre-trained*, pelatihan, serta evaluasi performa. Model dikembangkan dan dilatih menggunakan *python* serta pustaka Hugging Face, sementara sistem ekstraksi informasi diimplementasikan dalam bentuk antarmuka berbasis *streamlit*. Evaluasi performa model dilakukan menggunakan metrik *precision*, *recall*, dan *f1-score*, serta dilengkapi dengan pengujian sistem untuk mengukur akurasi model pada data nyata yang diperoleh melalui pengambilan gambar poster menggunakan kamera handphone.

Hasil penelitian menunjukkan bahwa model yang dikembangkan mencapai nilai macro precision sebesar 91,15%, macro recall sebesar 86,37%, dan macro F1-score sebesar 87,87% pada data pengujian, dengan konfigurasi *learning rate* 1,5e-5, *batch size* 8, dan jumlah 100 *epoch*. Selain itu, pengujian model sistem dilakukan pada skenario berbeda, seperti jarak pengambilan gambar 15cm, jarak pengambilan gambar 25cm, dan pencahayaan kurang, menghasilkan nilai akurasi rata-rata akurasi pada setiap skenario di sekitar 80-90%.

Kata Kunci: LayoutLMv3, OCR, ekstraksi informasi, poster acara, multimodal, *token classification*

ABSTRACT

Information extraction from event posters is a challenging task due to their highly variable layouts and the integration of textual and visual elements. Existing approaches are often insufficiently flexible to handle changes in document styled spatial formats. Therefore, this study aims to develop an event schedule information extraction system from seminar posters by leveraging multimodal understanding that can adapt to extreme layout variations.

The preprocessing stage includes image normalization to a resolution of 1000 pixels, followed by text recognition using Google Cloud Vision API to extract poster text at the word level. The extracted text is then labeled based on event schedule information, which is subsequently stored in Google Calendar. To enable more detailed annotation, labels are further decomposed using the BIO labeling scheme to distinguish between different entities that share the same label, with BIO labeling performed based on paragraph-level text recognition results from Google Cloud Vision API. Subsequently, the LayoutLMv3 model is fine-tuned to predict word-level labels by considering a combination of textual, layout, and visual information, encompassing model initialization from a pre-trained checkpoint, training, and performance evaluation. The model is developed and trained using Python and the Hugging Face library, while the information extraction system is implemented as a Streamlit-based interface. Model performance is evaluated using precision, recall, and F1-score metrics, and additional system-level testing is conducted to assess model accuracy on real-world data captured using a smartphone camera..

The experimental results show that the proposed model achieves a macro precision of 91.15%, a macro recall of 86.37%, and a macro F1-score of 87.87% on the test dataset, using a learning rate of 1.5e-5, a batch size of 8, and 100 training epochs. Furthermore, system-level testing under different scenarios—such as image capture distances of 15 cm and 25 cm, as well as low-light conditions—yields average accuracy values ranging from approximately 80% to 90% across all scenarios.

***Keywords**: LayoutLMv3, OCR, information extraction, event poster, multimodal document understanding, token classification*