ICIC Express Letters
Volume **18**, Number **4**, April **2024**

ICIC International ©2024 ISSN 1881-803X
pp. **325–332**

AN ENHANCED BIDIRECTIONAL ENCODER TRANSFORMERS WITH RELATIVE POSITION FOR INDONESIAN SKILL RECOGNITION

Meilany Nonsi Tentua[1], Suprapto[2,*] and Afiahayati[2]

[1]Department of Informatics
Faculty of Science and Technology
Universitas PGRI Yogyakarta
JL. PGRI I Sonosewu No. 117, Yogyakarta 55182, Indonesia
meilany@upy.ac.id

[2]Department of Computer Science and Electronics
Faculty of Mathematics and Natural Sciences
Universitas Gadjah Mada
Sekip Utara, Bulak Sumur, Yogyakarta 55281, Indonesia
afia@ugm.ac.id
*Corresponding author: sprapto@ugm.ac.id

Received April 2023; accepted July 2023

Abstract. *This paper presents an approach to improve Indonesian skill recognition using an enhanced bidirectional encoder transformers with relative position (EBERT-RP). The proposed method aims to overcome the challenges in recognizing Indonesian skills due to the complexity of the Indonesian language and the lack of annotated data. The EBERT-RP model incorporates relative position embeddings, which allow the model to capture the relative positions of tokens in a sentence, and a novel attention mechanism that improves the model's ability to attend the critical information. To evaluate the performance of the EBERT-RP model, we conducted experiments on a dataset of Indonesian skill recognition task. Our results show that the EBERT-RP model outperforms other state-of-the-art models, achieving an F1-score of 90.2% on the test set. Furthermore, we conducted an ablation study to analyze the contribution of the relative position embeddings and the attention mechanism to the model's performance. The results show that the relative position embeddings and the attention mechanism are crucial for high performance.*

Keywords: EBERT-RP, Skill recognition, Relative position, NER, Natural language processing

1. **Introduction.** Natural language processing (NLP) has become increasingly important in recent years, with numerous applications such as text classification [1], sentiment analysis [2], text generation [3], and skill recognition [4]. Skill recognition is a challenging task that involves identifying skills and competencies from a candidate's resume or job application [5]. This task has gained significant attention due to its importance in the recruitment process for many industries [6].

NLP models' success relies heavily on capturing the relationship between two words in a sentence. Bidirectional Encoder Representations from Transformers (BERT) is a pre-trained NLP model with remarkable performance in various tasks [7] and also has shown promising results in NER tasks [10]. BERT multilingual [7], IndoBERT [11], and IndoNLU [12] are pre-trained model Indonesian language. Using domain-adaptive pre-training such language is advantageous in enhancing the performance of the task for both hard skills and soft skill components. However, BERT does not consider the relative

DOI: 10.24507/icicel.18.04.325

position between words in a sentence [8], which is particularly important for languages with complex grammar and word order, such as Indonesian.

The recognition of skills from job postings and resumes is a crucial task in talent management and recruitment processes [9]. Skill recognition is challenging due to the variation in language use, context, and expression of skills across different job postings and resumes [4]. Named entity recognition (NER) has been widely used to identify and extract named entities, such as persons, organizations, and locations, from natural language texts. However, conventional NER methods do not perform well in recognizing skills, which often involve complex expressions and domain-specific terminology. The existing BERT-based models have limited performance in recognizing skills due to the model's inability to capture the complex relationships and dependencies between tokens in a sequence.

There has been a growing interest in incorporating relative position mechanisms into deep learning models for NLP tasks. Several studies have proposed methods to enhance BERT with relative position mechanisms for various NLP tasks, such as NER and relation extraction [13]. These methods have shown improved performance compared to the original BERT model [14]. However, to the best of our knowledge, there has been no study that investigates the use of relative position mechanisms for skill recognition tasks, particularly in the Indonesian language.

In this paper, we propose an enhanced bidirectional encoder transformers with relative position (EBERT-RP) for Indonesian skill recognition. Our model incorporates relative positional encoding to capture the relationship between words in a sentence. We also introduce a new skill recognition dataset for the Indonesian language. We evaluate our proposed model on the benchmark dataset and compare it with other baseline models, demonstrating its superior performance.

2. **Literature Review.**

This section describes the research conducted in skill recognition and the relative position embedding in transformers.

2.1. **Relative position embedding in transformers.**

Transformers [15] that use self-attention mechanisms have been adopted in various NLP tasks due to their parallelism and excellence in modeling very long contexts. Relative position embedding is transformer positional information proposed to improve the weaknesses of absolute position embedding with a sinusoidal function. The relative position was first introduced by adding a vector as directional information from the input element [16]. This vector is embedded in the key matrix in calculating attention values and the value matrix in calculating attention filtering values. It was further developed in Transformer-XL [17] and XLNet [3]. The relative position embedding was improved in Transformer-XL and XLNet by adding a bias parameter in the form of an absolute positional encodings vector to the content-based and location-based attention. The relative position embedding proposed in Transformer-XL and XLNet was used in NER modeling [13]. The modification was to remove the matrix parameter and maintain the bias parameter. In this paper, we utilize relative position encodings to capture the difference in positions between tokens, enabling the model to learn relative positional relationships.

2.2. **Skill recognition.**

Several studies have addressed skill recognition tasks using various methods. BERT-BiLSTM-CRF has been used in studies for skill recognition tasks [5] in English. It was found that the model based on BERT pretraining vector was better.

However, the disambiguation of multi-sense skills is the recognition and normalization of occupational skills in online recruitment. Using word embedding to quantify skills, apply Markov Chain Monte Carlo (MCMC) methods to aggregate vectors into clusters representing respective senses [18]. That clustering algorithm outperforms other clustering algorithms for the disambiguation problem.

The automated approach for skill entity recognition and normalization has essential applications in workforce training and job matching. It can significantly improve the accuracy of identifying and matching relevant skills with skill taxonomic generation and skill tagging [6]. The problem being addressed is an extreme multi-label classification (XMLC) problem [9] and SVM [19], where the binary relevance of thousands of individual skills needs to be determined based on the descriptions provided. The model effectively tackles the issue of missing skills and can help recover relevant skills that may have been overlooked during the job posting process.

SkillNER enables the detection of communities of job profiles based on their shared soft skills and communities of soft skills based on their shared job profiles [4]. This system demonstrates that it can automatically retrieve soft skills from a large corpus efficiently, proving useful for firms, institutions, and workers.

Deep learning methods, such as BERT, can be effective for skill recognition tasks in various languages. Using domain-adaptive pre-training is beneficial in improving the performance of the task in terms of both hard skills and soft skill components. Several studies for skill recognition use domain language adaptive pre-trained, such as Finnish language (FinBERT) [20], and JobSpanBERT [21].

Our proposed approach for Indonesian skill recognition introduces a novel model that utilizes enhanced bidirectional encoder transformers with relative positions. This enhancement involves integrating relative position information into the architecture of the model. Relative position information refers to the positions of words in relation to each other within a sentence or sequence rather than relying on their absolute positions in the input sequence.

3. **Material and Methodology.** In this section, we present our approach for an enhanced bidirectional encoder transformers with relative position (EBERT-RP) for Indonesian skill recognition, depicted in Figure 1. Our model comprises two modules, namely the pretrained language model EBERT-RP and skill recognition model. Each module has four steps: tokenization, pre-processing, proposed model of EBERT-RP, and training model.

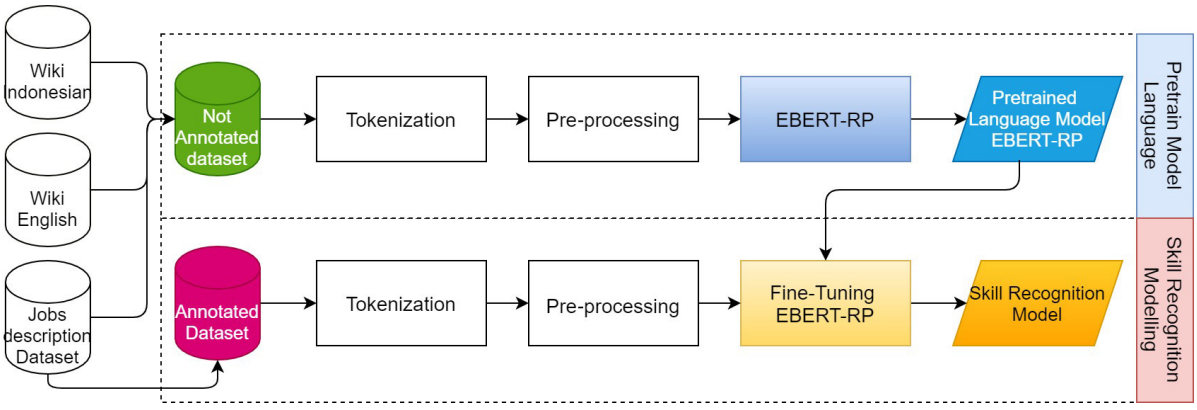

FIGURE 1. The architecture of EBERT-RP for skill recognition

3.1. **Material.** The EBERT-RP pre-training process and skill recognition modeling used a dataset composed of Indonesian language job requirements gathered from various job portals. The collected data was filtered to remove duplicates, resulting in 34,966 job requirements. The dataset was then divided into two groups: one for pre-training the language model and the other for skill recognition modeling.

The dataset used for pre-training the EBERT-RP language model did not include any annotations. To enhance the corpus for pre-training EBERT-RP, we add the Indonesian and English Wikipedia corpus to the job requirements dataset.

The dataset used in the skill recognition modelling follows the BIO tagging format [22] which includes three labels: Beginning, Inside, and Other. The dataset contains 4,394 job requirements pre-processed through word tokenization and manually labeled by an annotator to identify entities related to B-HSkill, S-Skill, B-Tech, I-HSkill, I-SSkill, and I-Tech.

3.2. **Tokenization.** Tokenize the text data into word pieces, also known as sub words. This is done using an algorithm called WordPiece, which breaks words into smaller units based on the frequency of occurrence. The vocabulary set in the EBERT-RP tokenizer is a set of all the unique tokens the model uses to represent text data. These tokens are the input sequences' building blocks fed into the EBERT-RP model. The vocabulary set includes a special set of tokens, such as the [CLS] token to indicate the start of a sequence, the [SEP] token to separate sentences or different parts of a sequence, and the [MASK] token from masking out certain tokens during training. The size of the vocabulary set in the EBERT-RP model has 31,923 of tokens. Table 1 shows a sample of tokenizing a job description.

Table 1. Sample of tokenizing a job description

Job description	Tokenize
Mahir dalam bahasa SQL, HTML, dan VB.	['[CLS]', 'mahir', 'dalam', 'bahasa', 'sql', ',', 'html', ',', 'dan', 'vb', '.', '[SEP]']

3.3. **Pre-processing.** The job requirements dataset is pre-processed by tokenizing the text into individual words using the WordPiece tokenizer. The resulting tokens are then converted to their corresponding token IDs and segment IDs and masked to indicate which tokens are actual words and which are padding. We also add special tokens to indicate the start and end of each sentence in the text. Table 2 shows a sample of pre-processing a job description.

Table 2. Sample of pre-processing a job description

Sentence	Mahir dalam bahasa SQL, HTML, dan VB.
Token IDs:	$[2, 24633, 1878, 2760, 12271, 16, 10727, 16, 1622, 30020, 18, 3]$
Segment IDs:	$[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$
Mask IDs:	$[1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$

The integer IDs assigned to each word or sub-word token in the input text will be referred to as the input word vector t_i and added to the absolute position vector p_i ($i = 1, 2, \ldots, n$ with $n = 256$). A maximum number of words that can be processed in EBERT-RP is $n = 256$. We use Equations (1) and (2) for the absolute position vector p_i.

$$PE_{(pos,2i)} = \sin\left(pos/10000^{2i/d}\right) \tag{1}$$

$$PE_{(pos,2i+1)} = \cos\left(pos/10000^{2i/d}\right) \tag{2}$$

The result of the addition will produce vector x_i. Equation (3) shows the addition process on each word vector.

$$x_i = t_i + p_i \tag{3}$$

where x_i, t_i, p_i, and i are output vectors in the position embedding layer: word vector, absolute position vector, and position vector, respectively.

3.4. **Proposed model.** The EBERT-RP model is based on the BERT architecture but includes several modifications. We add a relative position embedding layer to the model to help it recognize the relative positions of skills in the text. They also modify the input layer to include a token position embedding layer, which helps the model recognize the positions of individual words in the text.

A sequence input $x = (x_1, x_2, \ldots, x_m)$ with $x_i \in \mathbb{R}^{d_x}$ and m as the length of the sentence will be projected using three matrices $W^Q \in \mathbb{R}^{m \times d_q}$, $W^K \in \mathbb{R}^{m \times d_k}$, and $W^V \in \mathbb{R}^{m \times d_v}$ to extract feature representations Q, K, and V, which are called query, key, and value with $d_k = d_q$. The Q, K, and V matrices are calculated using Equation (4).

$$Q = x_i W^Q, \quad K = x_i W^K, \quad V = x_i W^V \tag{4}$$

Relative positional embedding is done by adding a vector $a_{ij}^V, a_{ij}^K \in \mathbb{R}^{d_a}$, where a_{ij}^V and a_{ij}^K are the direction information of the input elements x_i and x_j, respectively. The vector a_{ij}^K is embedded in the matrix $K = x_j W^K$ to interact with matrix $Q = x_i W^Q$ in calculating the attention score. Meanwhile, the vector is embedded in matrix $V = x_j W^V$ in calculating the attention filtration. Equations (5) and (6) show the modified attention filter and attention score with relative position embedding. Equation (7) shows the interaction of relative position with matrix Q.

$$z_i = \sum_{j=1}^{n} \alpha_{ij} \left(x_j W^V + a_{ij}^V \right) \text{ where } \alpha_{ij} = \frac{\exp e_{ij}}{\sum_{k=1}^{n} \exp e_{ik}} \tag{5}$$

$$e_{ij} = \frac{x_i W^Q \left(x_j W^K + a_{ij}^K \right)^T}{\sqrt{d_z}} \tag{6}$$

$$e_{ij} = \frac{x_i W^Q \left(x_j W^K \right)^T + x_i W^Q a_{ij}^{K^T}}{\sqrt{d_z}} \tag{7}$$

3.5. **Training.** We train EBERT-RP models from scratch based on the aforementioned configuration. We train EBERT-RP based on a masked language model objective as previous research has done [7], randomly we select 15% of tokens, and then substitute 80% of these tokens with "[MASK]", substitute another random 10% token, and take care 10% of the tokens unspoiled. We use a transformers encoder with 12 hidden layers (dimension = 768), 12 attention heads, and 3 hidden feedforward layers (dimension = 3,072). The only difference is the maximum sequence length, fixed at 256 tokens based on the average number of tokens in job requests. We also train BERT (without relative position) with the same dataset and configuration used in EBERT-RP to compare the performance of EBERT-RP.

3.6. **Fine-tuning.** The EBERT-RP pretrained language model is used to identify skill entities in job requirements. A linear layer and SoftMax activation are added in the NER skill model to identify entities. We use one of the hyperparameters suggested in research before [7] as the optimal hypermeter value in fine-tuning the model, for example, learning rate is 5e-3, epoch = 1, 2, 3, 4 or 5, and batch size = 32.

4. **Result and Discussion.** The EBERT-RP language models were trained using the NVIDIA A100-SXM machine. The training process used hyperparameters such as a learning rate of 1e-04, a dropout rate of 0.1, a batch size of 64, and 300 epochs with 448,500 steps. The resulting model has 110M parameters.

The EBERT-RP model and BERT (without relative position) are fine-tuned on the annotated job requirements dataset for modeling Indonesian skill recognition. We use a batch size of 32 and train the model for 11 epochs with an Adam optimizer.

We evaluate the performance of our proposed model EBERT-RP on the test set using precision, recall, and F1-score at the level of individual tokens and entities. Table 3 shows

the performance of EBERT-RP model at the token level outperforms BERT without a relative position. The lable "O" has the highest performance (F1-score = 96%), as it is owned by most of the tokens in the dataset because of this entity. The lable "B-Tech" has an F1-score = 85%, because a single word commonly owns this entity. Overall, this model has performance F1-score = 73.5% at token level.

Table 3. Performance model at the token level

Lable	BERT (without relative position)			EBERT-RP		
	Precision	Recall	F1-score	Precision	Recall	F1-score
B-HSkill	59%	65%	**62%**	69%	53%	60%
B-SSkill	71%	87%	78%	76%	86%	**81%**
B-Tech	84%	84%	84%	88%	81%	**85%**
I-HSkill	65%	52%	58%	72%	51%	**60%**
I-SSkill	68%	54%	61%	80%	54%	**64%**
I-Tech	73%	59%	65%	83%	60%	**69%**
O	95%	95%	95%	94%	97%	**96%**

Table 4 shows performance model at the entity level. The lable "B-SSkill" has the highest performance (F1-score = 97%), and the lowest performance is "I-Tech" (F1-score = 82%). The proposed model achieved an F1-score of 90.2% at entity levels.

Table 4. Performance model at the entity level

Lable	BERT (without relative position)			EBERT-RP		
	Precision	Recall	F1-score	Precision	Recall	F1-score
B-HSkill	85%	90%	87%	88%	92%	**90%**
B-SSkill	94%	97%	96%	95%	98%	**97%**
B-Tech	92%	92%	92%	94%	95%	**94%**
I-HSkill	85%	80%	82%	89%	87%	**88%**
I-SSkill	94%	82%	88%	93%	86%	**90%**
I-Tech	80%	69%	74%	88%	76%	**82%**

We compare the performance of our model with other baseline pre-trained language models, including BERT without relative position embedding, BERT [7], IndoBERT [11], and IndoNLU [12]. Table 5 shows performance comparison between other pretrained Indonesian language models and EBERT-RP.

Table 5. Comparison of models' performances

		Token level			Entity level		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Baseline	BERT	81%	75%	78%	87.9%	85.9%	86.3%
	IndoNLU	52%	30%	31%	58.3%	49.0%	53.2%
	IndoBERT	81%	76%	78%	87.4%	85.7%	86.5%
	BERT (without relative position)	74%	71%	72%	88%	85%	87%
Ours	EBERT-RP	80.2%	68.8%	73.5%	**91.1%**	**89.0%**	**90.2%**

The results showed that adding relative position embedding can significantly improve the model's recognition of Indonesian skills. The token level F1-score achieved by the proposed model was 73.5%, which is lower than that of BERT and IndoBERT. This discrepancy can be attributed to BERT and IndoBERT being trained on larger datasets than

EBERT-RP. However, at the entity level, our proposed model surpasses the performance of the baseline models with a precision of 91.1%, a recall of 89%, and an F1-score of 90.2%. This improvement is because EBERT-RP incorporates relative positional encoding to capture the relationship between words in a sentence. Additionally, EBERT-RP is trained using a corpus of job descriptions, so that skill recognition can be better captured.

5. **Conclusion.** This paper proposes an enhanced bidirectional encoder transformers with relative position (EBERT-RP) for Indonesian skill recognition. We incorporate a relative positional encoding layer into the BERT architecture to enhance the model's ability to capture the relationship between words in a sentence.

Our experimental results demonstrate that EBERT-RP outperforms baseline models such as BERT [7], IndoNLU [12] and IndoBERT [11] in terms of precision (91.1%), recall (89%) and F1-score (90.2%) for entity level. The ablation study shows that the relative positional encoding layer contributes significantly to the model's performance.

Our work contributes to the development of natural language processing techniques for low-resource languages such as Indonesian. Incorporating relative positional encoding is a promising approach for improving the performance of language models in languages with complex syntax and grammar. There are various potential areas for future development, for example, 1) a direction could involve solely utilizing job descriptions in the dataset during pre-training of the language model EBERT-RP without incorporating additional corpora like the wiki corpus, and 2) exploring cross-lingual skill recognition involves researching methods to expand the model's capabilities beyond the Indonesian language and enable it to identify skills in multiple languages.

REFERENCES

[1] A. W. Haryanto, E. K. Mawardi and Muljono, Influence of word normalization and chi-squared feature selection on support vector machine (SVM) text classification, *Proc. of 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life*, pp.229-233, DOI: 10.1109/ISEMANTIC.2018.8549748, 2018.

[2] M. R. Yaakub, M. Iqbal, A. Latiffi and L. S. Zaabar, A review on sentiment analysis techniques and applications, *Joint Conference on Green Engineering Technology & Applied Computing*, DOI: 10.1088/1757-899X/551/1/012070, 2019.

[3] Z. Yang, Z. Dai, Y. Yang and J. Carbonell, XLNet: Generalized autoregressive pretraining for language understanding, *Proc. of the 33rd International Conference on Neural Information Processing Systems*, pp.5753-5763, 2019.

[4] S. Fareri, N. Melluso, F. Chiarello and G. Fantoni, SkillNER: Mining and mapping soft skills from any text, *Expert Syst. Appl.*, vol.184, no.7, 115544, DOI: 10.1016/j.eswa.2021.115544, 2021.

[5] L. Cao and J. Zhang, Skill requirements analysis for data analysts based on named entities recognition, *The 2nd International Conference on Big Data and Informatization Education*, pp.64-68, DOI: 10.1109/ICBDIE52740.2021.00023, 2021.

[6] M. Zhao et al., SKILL: A system for skill identification and normalization, *Proc. of Natl. Conf. Artif. Intell.*, vol.29, no.2, pp.4012-4017, 2015.

[7] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol.1, pp.4171-4186, DOI: 10.18653/v1/N19-1423, 2019.

[8] A. Qu, J. Niu and S. Mo, Explore better relative position embeddings from encoding perspective for transformer models, *Proc. of 2021 Conf. Empir. Methods Nat. Lang. Process.*, pp.2989-2997, DOI: 10.18653/v1/2021.emnlp-main.237, 2021.

[9] A. Bhola, K. Halder, A. Prasad and M.-Y. Kan, Retrieving skills from job descriptions: A language model based extreme multi-label classification framework, *Proc. of the 28th International Conference on Computational Linguistics*, no.4, pp.5832-5842, DOI: 10.18653/v1/2020.coling-main.513, 2021.

[10] Y. Chen, J. Mikkelsen, A. Binder, C. Alt and L. Hennig, A comparative study of pre-trained encoders for low-resource named entity recognition, *Proc. of the Annual Meeting of the Association for Computational Linguistics*, pp.46-59, DOI: 10.18653/v1/2022.repl4nlp-1.6, 2022.

[11] F. Koto, A. Rahimi, J. H. Lau and T. Baldwin, IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP, *Proc. of the 28th International Conference on Computational Linguistics*, pp.757-770, DOI: 10.18653/v1/2020.coling-main.66, 2021.

[12] B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar and A. Purwarianti, IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding, *Proc. of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pp.843-857, 2020.

[13] H. Yan, B. Deng, X. Li and X. Qiu, TENER: Adapting transformer encoder for named entity recognition, *arXiv Preprint*, arXiv: 1911.04474, 2019.

[14] N. P. T. Ha, T. N. T. Nguyen, E. Salesky, S. Stueker, J. Niehues and A. Waibel, Relative positional encoding for speech recognition and direct translation, *INTERSPEECH*, pp.31-35, 2020.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, Attention is all you need, *Advances in Neural Information Processing Systems*, vol.2017-Decem, pp.5999-6009, 2017.

[16] P. Shaw, J. Uszkoreit and A. Vaswani, Self-attention with relative position representations, *Proc. of 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol.2, pp.464-468, DOI: 10.18653/v1/n18-2074, 2018.

[17] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le and R. Salakhutdinov, Transformer-XL: Attentive language models beyond a fixed-length context, *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*, pp.2978-2988, DOI: 10.18653/v1/p19-1285, 2019.

[18] Q. Luo, M. Zhao, F. Javed and F. Jacob, Macau: Large-scale skill sense disambiguation in the online recruitment domain, *Proc. of 2015 IEEE International Conference on Big Data*, pp.1324-1329, DOI: 10.1109/BigData.2015.7363890, 2015.

[19] R. Boselli, M. Cesarini, F. Mercorio and M. Mezzanzanica, Classifying online job advertisements through machine learning, *Futur. Gener. Comput. Syst.*, vol.86, pp.319-328, DOI: 10.1016/j.future.2018.03.035, 2018.

[20] M. Chernova, *Occupational Skills Extraction with FinBERT*, Ph.D. Thesis, Metropolia University of Applied Sciences, 2020.

[21] M. Zhang, K. N. Jensen, S. D. Sonniks and B. Plank, SkillSpan: Hard and soft skill extraction from English job postings, *Proc. of 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp.4962-4984, DOI: 10.18653/v1/2022.naacl-main.366, 2022.

[22] N. Alshammari and S. Alanazi, An Arabic dataset for disease named entity recognition with multi-annotation schemes, *Data*, vol.5, no.3, pp.1-8, DOI: 10.3390/data5030060, 2020.