

Publication Date: 15 May 2024

Archs Sci. (2024) Volume 74, Issue 2 Pages 118-123, Paper ID 2024217.
<https://doi.org/10.62227/as/74217>

The Chatbot Systems for Mathematics Education using Knowledge Graphs and Named Entity Recognition

Hasna Salsabila¹ and Isnaini Rosyida^{1,*}

¹Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia.

Corresponding authors: iisnaini@gmail.com.

Abstract The development of chatbot systems for mathematics courses in elementary school has gained significant attention due to their potential to enhance the learning experience. This study proposes a novel approach that combines knowledge graph (KG) and Named Entity Recognition (NER) methods using Neo4j and SpaCy within the Rasa Open-Source v3.0 platform as chatbot frameworks. The knowledge graphs represent mathematical concepts and their relationships, enabling the chatbot to provide accurate and relevant responses to user queries. The NER SpaCy is employed to identify and extract mathematical entities from user inputs, ensuring a precise understanding of the context. The integrations of Neo4j and NER using SpaCy with Rasa Open-Source v3.0 facilitate efficient information retrieval and improve the conversational abilities of the chatbot. Experimental results demonstrate the effectiveness of the proposed approach, showcasing its potential as an educational tool for fifth-grade students in elementary schools.

Index Terms Knowledge graphs, NER, chatbot, mathematics, graph database, Rasa framework.

I. Introduction

IN recent years, there has been a growing interest in developing intelligent chatbot systems to enhance the learning experience in various domains, including mathematics education. These chatbots provide private and interactive conversations that allow students to engage in a dynamic learning process. One crucial aspect of developing a chatbot is the extraction and organization of relevant information. Labadze et al. discussed the benefits of AI chatbots in education [1]. Memon et al. constructed an educational domain-based multichatbot communication system [2]. Further, Okonkwo and Ade-Ibijola also provided chatbot applications in education [3].

A knowledge graph (KG) is a powerful representation of structured data that captures relationships and connections among different entities. It provides efficient storage and retrieval of information, making it ideal for building intelligent chatbot systems [4]. Patsoulis et al. provided chatbot integration with knowledge graphs in e-government [5]. Timon-Reina et al. proposed graph databases and their implementations in the biomedical fields [6]. Moreover, Zou described some applications of the KG [7]. Yoo and Jeong constructed an intelligent chatbot using the BERT model and KG [8]. Agrawal et al. built KG from unstructured texts and implemented it in cybersecurity education [9]. Xue provided life-long education data classification based on a KG [10]. Meloni et al. integrated conversational agents with KG focused on the

scholarly domain [11]. Buhl et al. designed a chatbot based on KG that actively learned from humans [12]. Moreover, Nair and Shivani proposed a question-answering model using KG for remote school education [13].

On the other hand, Named Entity Recognition (NER) is a Natural Language Processing (NLP) technique that identifies and categorizes named entities, such as names of people, places, organizations, and numerical expressions, in a text document [14]. Khairunnisa et al. discussed dataset enhancement and multilingual transfer for NER in the Indonesian language [15]. Yanti et al. proposed an application of NER via Twitter on SpaCy in Indonesian [16]. Ali designed a chatbot with the NER model using an artificial neural network [17]. Therefore, knowledge graphs and NER techniques play a significant role in constructing a chatbot system.

In this article, we construct a chatbot system for fifth-grade elementary school mathematics using Rasa Open Source v3.0 as the chatbot framework. The system incorporates the utilization of knowledge graphs and NER techniques through Neo4j and SpaCy libraries, respectively. By leveraging these tools, the chatbot system aims to provide contextually relevant responses to student queries so that it enhances their learning experience.

This paper is organized into five sections. The second section is preliminaries. The third section discusses the method used in this research. The fourth part provides results and discussion. Finally, the conclusions are given in the fifth

section.

II. Preliminaries

The next subsections provide some basic concepts used in this paper.

A. Named Entity Recognition (NER)

NER is a significant task in Natural Language Processing (NLP). The NER can be defined as a method for information extraction that processes both structured and unstructured documents to identify entities such as people, locations, organizations, or companies. The term “named entity” was first introduced as a crucial component of information extraction at the Sixth Message Understanding Conference (MUC-6) [14]. The MUC-6 categorized named entities into three label types: ENAMEX (people, organizations, locations), TIMEX (dates and times), and NUMEX (money, percentages, quantities) [18].

B. Knowledge Graph (KG)

A graph $G = (V, E)$ consists of a node-set V and an edge set E , where each edge indicates a connection between two nodes in the graph. Whereas a knowledge graph (KG), often called a semantic network, illustrates the connections among real-world elements, including concepts, events, objects, and circumstances, as well as the network of these connections. The term KG was coined since this information is commonly stored in a graph database and portrayed as a graph structure. The three primary parts of a KG are labels, edges, and nodes. A node can be any item, location, or person. The relationship between the nodes is defined by an edge. A customer like IBM or an agency like Ogilvy are two examples of nodes. An edge would be a customer relationship between IBM and Ogilvy [19].

A KG can also be represented in three triples, i.e., head (h), tail (t), and predicate (p), where h and t are elements of V , that indicate subject and object (entities), and p is an element of E that describes a semantic relationship between the subject and object ([20], [21]). For instance, Figure 1 shows a KG with three triples (h , t , and p) as follows: square (“persegi”), triangle (“segitiga”), and trapezoid (“trapesium”) are objects, two-dimensional shape (“bangun datar”) is the subject, and the predicate (p) is “belongs to.” Further, a triangle is also a subject and a square is an object with the relation “contained in,” etc.

In 2012, Google introduced the KG, used as semantic knowledge in search. The function of the Google KG is to identify entities in the text to enhance search results by providing semantically structured summaries and links to connect entities in research queries [6]. This search aims to improve search engines and user research experiences. The most common applications of the KG are in question-answering systems, recommender systems, and information retrieval. In chatbots, a KG is usually built upon existing modules to connect all gathered information, bridging structured and unstructured information, and representing this information in

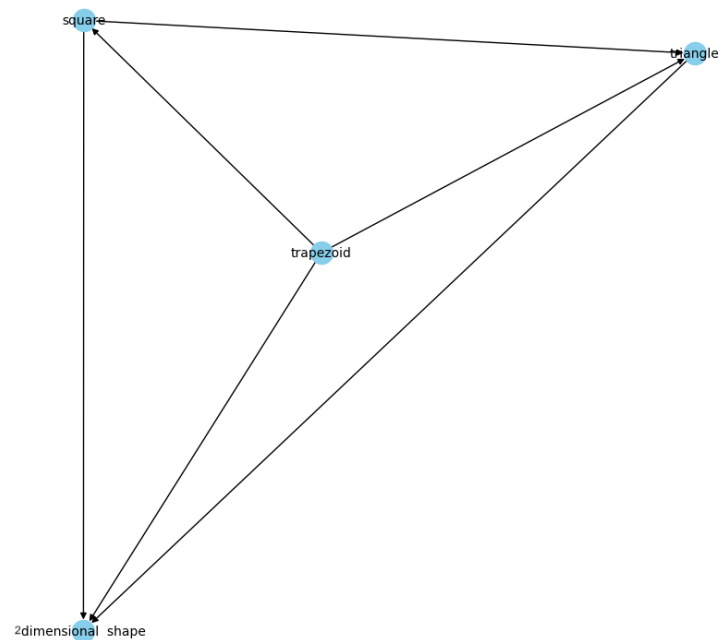


Figure 1: An example of a KG of two-dimensional shape (“bangun datar”).

a clear and structured form to users [5]. Another example of the usage of KG is in Kbot to select responses and provide feedback to the user. User queries result in named entities, and information from these entities is extracted using NER. The extracted results from these entities using NER are used to construct a KG. For example, in Kbot, a user’s question “Who is Alan Turing?” would be categorized as “person” using NER, and “Alan Turing” would be extracted as the main entity [4].

C. Rasa Open Source

An open-source machine learning framework called Rasa is used to create chatbots and AI agents. It is an open-source library that provides a framework for creating Python-based programming language chatbots [22]. Rasa consists of two components, i.e., Rasa Natural Language Understanding (NLU) and Rasa Core. Both components have their respective roles in the chatbot system. In the latest version, they merged into one system called Rasa. On the one hand, Rasa NLU is utilized to interpret messages sent by users. On the other hand, conversations and decision-making are handled by Rasa Core [2]. Various researchers have developed chatbots using the Rasa framework, such as in ([23], [24]). Rasa’s most recent release is called Rasa X, which is a tool that supports the creation, enhancement, and implementation of AI assistants driven by the Rasa framework. Rasa NLU has some additional features, i.e., SpaCy and tensorflow. The SpaCy is a free and open-source package for advanced NLP in Python.

III. Methods

The methods used to construct the chatbot are as follows.

- 1) **Data collection:** the technique involves data scraping to gather text from various online article sources. This text contains information related to vocabulary associated with elementary school mathematics material. The purpose of text data scraping in this research is to expedite the process of collecting text data about mathematics material. This scraping process utilizes the BeautifulSoup and Newspaper3k libraries in the Python language. Data scraping facilitates the retrieval of text from online articles in large quantities. The collected text data includes material titles and content. The gathered text is then stored in JSON file format or saved in a Pandas data frame with a CSV file format, and text preprocessing is performed to obtain structured data.
- 2) **Designing the chatbot:**
 - 2.1 The use of KG in this research serves as a graph database. It assists in the conversation system of the chatbot, primarily in the process of retrieving information from user text messages. The KG pertains to the subject matter of 5th-grade mathematics in elementary school. It consists of nodes and edges. Nodes in the KG represent named entities related to the subject matter of 5th-grade mathematics. Edges in the KG represent relationships between each node, such as characters, formulas, variety, etc.
 - 2.2 Further, the NER plays a role in extracting noun entities related to elementary school mathematics material. These entities represent nodes in a knowledge graph. In the knowledge graph, there are edges connecting these nodes. This research employs the SpaCy library to train the NER model according to the language used [15], which is "Indonesian language", as the SpaCy library has not yet provided a pre-trained NER model in the "Indonesian language". For instance, if a student, as a user, inputs a text message into the chatbot, denoted as Q , where " Q : what is the subject matter of two-dimensional shapes ("bangun datar")?," SpaCy is used to create the NER model to extract information from the entity "bangun datar" into the graph database. The resulting extracted entity is then stored in the KG, displaying information related to two-dimensional shapes (see Figure 1).
 - 2.3 Since we do not use Rasa Enterprise, a paid version of Rasa, the Facebook Messenger feature is utilized to display conversations happening in the chatbot and is integrated into the Rasa framework. This integration process utilizes the software ngrok which provides a server URL for connecting Rasa with the Facebook Messenger.

IV. Results and Discussion

Using KG and NER methods, along with Neo4j and SpaCy library, to build a chatbot system for the mathematics course in the fifth-grade elementary school with Rasa Open Source v3.0 as the chatbot framework, showed promising results. In this section, we present the outcomes of our experiments and discuss their implications.

The KG plays a crucial role in organizing and representing the domain-specific knowledge required for the chatbot. Neo4j, a graph database, creates and manages the KG efficiently. We extract mathematical concepts and their relationships from various sources, such as textbooks, online educational materials, and curriculum guidelines. These entities are modeled as nodes in the graph, while their relationships are represented as edges. The KG provides a structured representation of the mathematical knowledge, allowing the chatbot to access and utilize it effectively. The KG in Neo4j is depicted in Figure 2. Meanwhile, the graph database in Neo4j is shown in Figure 3.



Figure 2: The KG in Neo4j.

Furthermore, we employ NER with SpaCy to identify relevant entities within user queries. The NER model is trained on a dataset containing annotated examples of mathematical entities such as numbers, operations, units of measurement, and mathematical terms. By incorporating NER, the chatbot could accurately recognize and extract mathematical entities from user input, enabling it to provide more precise and contextually appropriate responses. The training process is described in Figure 4.

We have seen notable improvements in the chatbot's performance after conducting thorough testing and evaluation. The chatbot's capacity to comprehend and analyze mathematical queries is improved by the combination of the KG

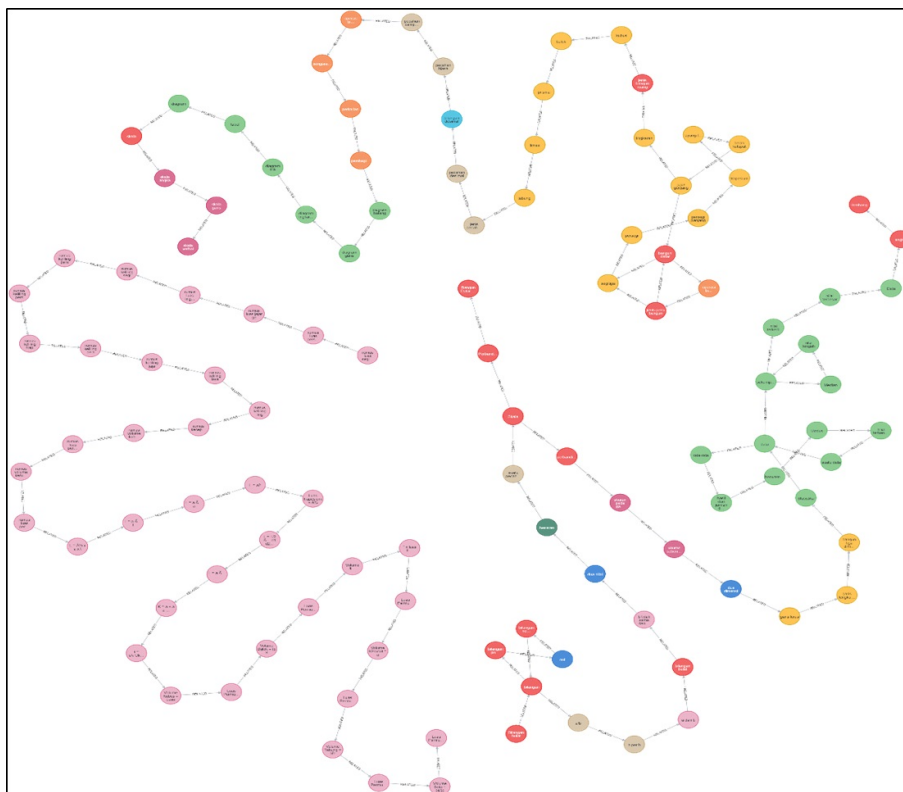


Figure 3: Graph database in Neo4j.

```

===== Initializing pipeline =====
[2023-07-01 19:11:53,536] [INFO] Set up nlp object from config
[2023-07-01 19:11:53,567] [INFO] Pipeline: ['tok2vec', 'ner']
[2023-07-01 19:11:53,583] [INFO] Created vocabulary
[2023-07-01 19:11:53,583] [INFO] Finished initializing nlp object
[2023-07-01 19:11:57,114] [INFO] Initialized pipeline components: ['tok2vec', 'ner']
█ Initialized pipeline

===== Training pipeline =====
█ Pipeline: ['tok2vec', 'ner']
█ Initial learn rate: 0.001

```

E	#	LOSS TOK2VEC	LOSS NER	ENTS_F	ENTS_P	ENTS_R	SCORE
0	0	0.00	37.29	0.45	0.94	0.30	0.00
1	200	174.96	5907.32	37.53	40.55	34.92	0.38
2	400	252.80	4253.95	72.47	73.84	71.15	0.72
4	600	267.13	3451.08	84.34	85.59	83.13	0.84
6	800	245.39	2719.35	91.53	91.58	91.47	0.92
9	1000	295.41	2404.23	94.77	95.40	94.15	0.95
12	1200	363.02	2102.93	96.33	96.01	96.65	0.96
16	1400	313.75	1706.52	97.38	97.29	97.47	0.97
21	1600	292.33	1750.68	97.95	98.02	97.88	0.98
27	1800	258.78	1442.66	98.34	98.25	98.44	0.98
35	2000	280.80	1320.97	99.13	99.03	99.22	0.99
44	2200	271.64	1069.02	99.27	99.37	99.18	0.99
55	2400	131.20	739.99	99.42	99.59	99.26	0.99
66	2600	109.80	583.10	99.33	99.33	99.33	0.99
77	2800	505.58	1242.20	99.50	99.44	99.55	0.99
88	3000	88.09	495.68	99.46	99.59	99.33	0.99
99	3200	188.48	491.90	99.57	99.59	99.55	1.00
111	3400	2314.96	2113.80	99.35	99.29	99.40	0.99
122	3600	342.78	678.96	99.50	99.55	99.44	0.99
133	3800	241.67	600.55	99.63	99.70	99.55	1.00
144	4000	120.92	428.35	99.59	99.66	99.52	1.00
155	4200	106.88	397.26	99.52	99.70	99.33	1.00

Figure 4: The training custom model NER with SpaCy.

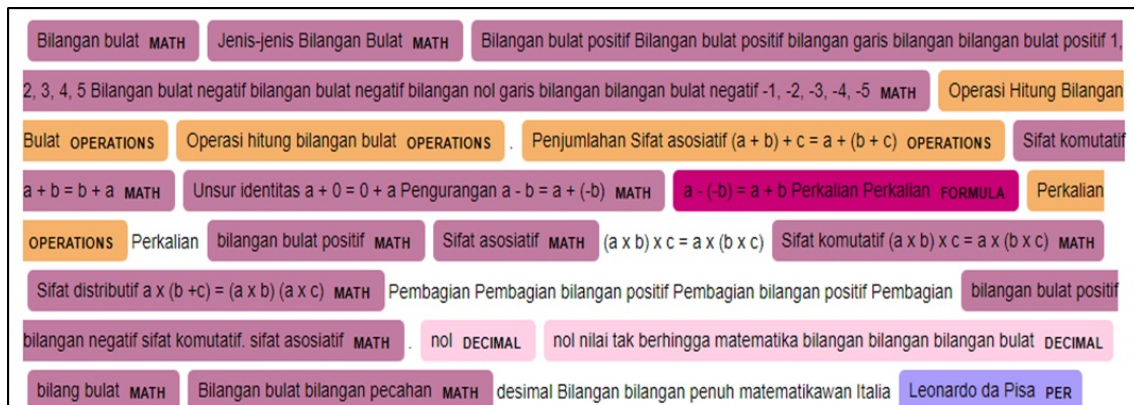


Figure 5: The NER using displaCy.

and NER approaches. The chatbot could accurately identify mathematical entities mentioned in user queries, enabling it to retrieve relevant information from the KG. This resulted in more accurate and informative responses provided by the chatbot. The appearance of NER with SpaCy using displaCy is shown in Figure 5. DisplaCy is the quickest tool to visualize named entities.

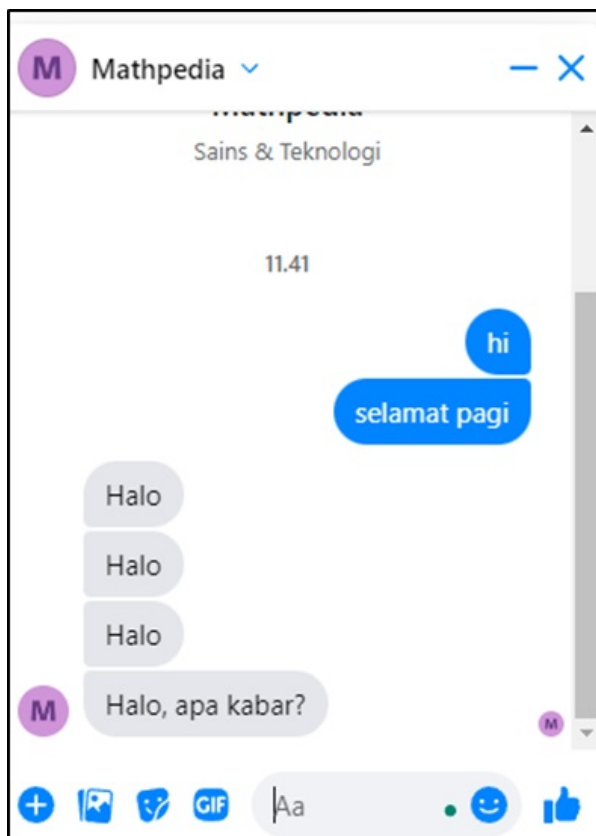


Figure 6: The chatbot implementation result on the Facebook Messenger.

Moreover, the chatbot's utilization of the KG facilitates intelligent responses based on the relationships among mathe-

matical concepts. It could provide explanations, examples, and step-by-step solutions by traversing the graph and retrieving connected nodes and edges. This feature proves to be highly beneficial for students, as it enhances their understanding of mathematical concepts and problem-solving techniques. The chatbot implementation result is given in Figure 6.

Overall, the integration of the KG and NER methods using Neo4j and SpaCy in the development of the chatbot system yields promising results. The chatbot demonstrates improved accuracy and effectiveness in understanding and responding to mathematical queries from fifth-grade elementary school students. The combination of Rasa open source v3.0 as the chatbot framework, along with these methods, showcases the potential for developing an intelligent chatbot system in the field of mathematics education.

V. Conclusions

Following the outcome, it was clear that using the Rasa Framework, the Named Entity Recognition (NER) method, and knowledge graphs (KG) to build a conversation system in a mathematics chatbot was a successful method that made it easier for users to interact with the chatbot. Through the implementation of the Rasa framework, the chatbot could provide a more natural and responsive conversational experience. The use of the NER method in extracting information from text related to mathematics subject matter and user queries in the chatbot has proven capable of identifying and extracting important entities such as mathematical concepts, formulas, variables, and other related keywords. This enables the chatbot to understand and respond more accurately and relevantly to user queries. As a knowledge base for the chatbot, using the KG as a model and visualization of subject matter entities that students often talk about was very helpful because it gave structured and logically connected information. The KG allowed the chatbot to provide deeper and more detailed explanations of mathematical concepts as well as display relationships between these concepts. With the presence of the KG, the chatbot could provide more comprehensive answers and assist users in understanding mathematics subject matter.

Based on the results of this research, several suggestions could be given for the development of a conversation system in a mathematics chatbot using KG and NER, as follows: expanding the range of recognized entities in the system to enhance the quality and accuracy of entity recognition. Further development is needed to identify and extract more specific mathematical entities. To accomplish this, it is necessary to use more pertinent sources and datasets, create more intricate conversational interactions for responding to user inquiries, and improve the chatbot to interact with users by asking follow-up questions, offering practice problems, or presenting case studies.

Further, it is important to evaluate and improve the chatbot system continuously. Collecting feedback from users and regularly monitoring the chatbot's performance can help identify weaknesses and address existing shortcomings. Integration with other platforms and services to enhance accessibility and chatbot utilization should be reconsidered, especially integration with frequently used platforms or services such as instant messaging applications or online learning platforms. This can enhance the utilization of the chatbot in the field of education. It is expected that the development of a conversation system in a mathematics chatbot using KG and NER could be continuously improved since it will provide greater benefits in supporting interactive and effective mathematics learning.

References

- [1] Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), 56.
- [2] Memon, Z., Aghian, H., Sarfraz, M. S., Hussain Jalbani, A., Oskouei, R. J., Jalbani, K. B., & Hussain Jalbani, G. (2021). Framework for educational domain-based multichatbot communication system. *Scientific Programming*, 2021, 1-9.
- [3] Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033.
- [4] Ait-Mlouk, A., & Jiang, L. (2020). KBot: a Knowledge graph based chatBot for natural language understanding over linked data. *IEEE Access*, 8, 149220-149230.
- [5] Patsoulis, G., Promikyridis, R., & Tambouris, E. (2021, November). Integration of chatbots with Knowledge Graphs in eGovernment: The case of Getting a Passport. In *Proceedings of the 25th Pan-Hellenic Conference on Informatics* (pp. 425-429).
- [6] Timón-Reina, S., Rincón, M., & Martínez-Tomás, R. (2021). An overview of graph databases and their applications in the biomedical domain. *Oxford Database*, 1-22.
- [7] Zou, X. (2020). A Survey on Application of Knowledge Graph. *Journal of Physics: Conference Series*, 1487, 012016 (2020).
- [8] Yoo, S.Y., & Jeong, O. (2019). An Intelligent Chatbot Utilizing BERT Model and Knowledge Graph. *Journal of Society for e-Business Studies*, 24(3), 87-98.
- [9] Agrawal, G., Deng, Y., Park, J., Liu, H., & Chen, Y.C. (2022). Building Knowledge Graphs from Unstructured Texts: Applications and Impact Analyses in Cybersecurity Education. *Information*, 13(11), 526.
- [10] Xue, Y. (2023). Knowledge Graph Based Recommendation by Adversarial Learning Algorithm in Application of Lifelong Education Data Classification. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 1-23.
- [11] Meloni, A., Angioni, S., Salatino, A., Osborne, F., Recupero, D.R., & Motta, E. (2023). Integrating Conversational Agents and Knowledge Graphs Within the Scholarly Domain. *IEEE Access*, 11, 22468-22489.
- [12] Buhl, D., Szafarski, D. Welz, L., & Lanquillon, C. (2023). Conversation-Driven Refinement of Knowledge Graphs: True Active Learning with Humans in the Chatbot Application Loop. In: Degen, H., Ntoa, S. (eds) Artificial Intelligence in HCI, Lecture Notes in Computer Science, vol 14051.
- [13] Nair, L.S. & Shivani, M.K. (2022). Knowledge Graph Based Question Answering System for Remote School Education. In: 2022 *International Conference on Connected Systems and Intelligence (CSI)*, p.1-5.
- [14] Goyal, A., Gupta, V. & Kumar, M. (2018). Recent Named Entity Recognition and Classification techniques: A systematic review. *Computer Science Review*, 29, 21-43.
- [15] Khairunnisa, S.O., Chen, Z., & Komachi, M. (2023). Dataset Enhancement and Multilingual Transfer for Named Entity Recognition in the Indonesian Language. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(6), 1-21.
- [16] Yanti, R.M., Santoso, I., & Suadaa, L.H.. (2021). Application of Named Entity Recognition via Twitter on SpaCy in Indonesian (Case Study: Power Failure in the Special Region of Yogyakarta). *Indonesian Journal of Information Systems*, 4(1), 76-86.
- [17] Ali, N. (2020). Chatbot: A Conversational Agent Employed with Named Entity Recognition Model using Artificial Neural Network. *ArXiv: abs/2007.04248*.
- [18] Sharnagat, R. (2014). Named Entity Recognition: Literature Survey. Center for Indian Language Technology, preprint, retrieved February 5, 2024 from <https://www.cfilt.iitb.ac.in/~cfiltnew/resources/surveys/rahul-ner-survey.pdf>.
- [19] IBM. (2024). What is a knowledge graph?, Retrieved February 15, 2024 from <https://www.ibm.com/topics/knowledge-graph>.
- [20] Duong, H.T., Ho, V.H., & Do, P. (2023). Fact-checking Vietnamese Information Using Knowledge Graph, Datalog, and KG-BERT. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(10), 1-23.
- [21] Zhong, L., Wu, J., Li, Q., Peng, H., & Wu, X. (2024). A Comprehensive Survey on Automatic Knowledge Graph Construction. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 56(4), 1-62.
- [22] Bocklisch, T., Faulkner, J., Pawlowski, N., & Nichol, A. (2017). Rasa: Open Source Language Understanding and Dialogue Management. *NSIP 2017 Conversational AI Workshop*, p. 1-9.
- [23] Fonseca, J. & Rodrigues, F. (2023). ChatBot for student service based on RASA framework, retrieved March 5, 2024 from <https://www.researchsquare.com/article/rs-2771200/v1> preprint.
- [24] Fauzia, L., Hadiprakoso, R.B., & Girinoto. (2021). Implementation of Chatbot on University Website Using RASA Framework. In: 2021 4th International Seminar on *Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, pp. 373-378.

...