Implementation of German Language Chatbot for Local Wisdom Information Services in North Sumatra

Herlina Jasa Putri Harahap[1], Ahmad Bengar Harahap[2], Muhammad Reza[3]

{ herlinajasaputri@unimed.ac.id }

German Language Department, Universitas Negeri Medan, Jalan Willem Iskandar, Pasar V Medan Estate, Percut Sei Tuan, Deli Serdang[1,2], Department of Linguistics, Universitas Pendidikan Indonesia, Jl. Dr. Setiabudhi No. 229 Bandung[3]

Abstract. In the modern digital era, technology plays a pivotal role in the tourism industry, particularly for foreign tourists, including German-speaking visitors traveling to Indonesia, especially North Sumatra. These tourists often seek information about local culture and customs online before their trips. This pre-travel research helps them understand key aspects of their destinations, addressing the challenge of limited available information and the scarcity of knowledgeable sources. This study emphasizes the importance of a robust research methodology in identifying and analyzing the necessary data sources to answer relevant questions and provide effective solutions. The research process involved developing and testing a chatbot using Rasa, an open-source conversational AI framework. The steps included installing Rasa, creating a new project directory, and defining the chatbot dataset, which involved creating essential files such as nlu.md, domain.yml, and stories.md. The training phase utilized machine learning frameworks like TensorFlow or PyTorch, ensuring that the chatbot met the study's objectives and provided accurate, helpful responses for users.

Keywords: Tourism Industry, Local wisdom in North Sumatra, German Chatbot development.

1. Introduction

In the current digital age, technology plays an increasingly vital role in the tourism industry (Poerwanto & Shambodo, 2020). The integration of technology not only enhances the experiences of tourists but also supports them in navigating the complexities of travel. Before embarking on a journey, tourists frequently turn to the internet to search for relevant information about their destinations (Komalasari et al., 2019). This pre-trip research is crucial for obtaining essential details about their intended destinations, helping them make informed decisions about what to expect, where to visit, and how to plan their activities.

However, one common challenge tourists face is the difficulty of accessing accurate, up-to-date, and reliable information about the places they plan to visit. Despite the wealth of re-

ICIESC 2024, September 17, Medan, Indonesia

DOI 10.4108/eai.17-9-2024.2352982

sources available online, the information is often scattered, outdated, or irrelevant, leaving travelers with incomplete knowledge. Moreover, in many cases, there are limited sources or individuals available to provide comprehensive and personalized insights about specific tourist destinations. This information gap not only hampers the efficiency of trip planning but also diminishes the overall travel experience. To address this issue, advancements in artificial intelligence, particularly the development of chatbots, offer a promising solution. Chatbots, powered by natural language processing (NLP), enable seamless communication between users and machines, simulating human-like interactions. By harnessing AI technology, chatbots can assist tourists by answering queries, providing real-time updates, and offering tailored recommendations based on user preferences. According to Al-Jarf (2021), as AI systems become increasingly sophisticated, the ability to understand and process human language becomes an essential feature for ensuring smooth and efficient communication between users and machines. In this way, language serves as a bridge for cultural and social exchange, facilitated by technological advancements.

One of the most accessible and widely-used AI frameworks for chatbot development is RASA. RASA operates on two key components: RASA NLU (Natural Language Understanding) and RASA Core. It is an open-source Python-based library designed to help developers build conversational software that can engage users in meaningful dialogue. The RASA chatbot framework is grounded in the philosophy of making dialogue management systems more intuitive and efficient. By leveraging machine learning, RASA simplifies the process of chatbot creation, enabling developers to bootstrap complex systems with minimal initial data and effort. This approach not only accelerates the development process but also ensures that chatbots can adapt and evolve with continuous learning and updates.

Furthermore, RASA's open-source nature allows for flexibility in implementation, making it an attractive option for developers looking to create robust, scalable chatbot solutions. A study by Josphineleela et al. (2023) highlights the effectiveness of RASA in delivering reliable conversational experiences, with the framework's machine learning-based architecture allowing for continuous improvement and refinement of chatbot interactions. As AI continues to advance, frameworks like RASA will play an increasingly pivotal role in shaping how tourists interact with technology, ultimately enhancing their travel experiences by providing timely, relevant, and personalized information. In conclusion, the integration of AI-powered chatbots in the tourism sector offers significant potential for improving the availability and accuracy of information for travelers. By bridging the information gap and providing real-time support, chatbots can revolutionize how tourists plan, navigate, and experience their journeys, ensuring a more seamless and enjoyable travel experience.

One of the key principles in chatbot development is intent recognition and entity extraction. Intent recognition refers to a chatbot's ability to discern the purpose or intention behind a user's message, whether delivered through text or voice. For instance, a user might ask for details about a tourist destination, request food recommendations, or seek directions to a particular location. The chatbot's role is to correctly interpret and respond to these intentions.

Entity extraction, on the other hand, focuses on the chatbot's ability to identify and comprehend specific pieces of information within the user's input. This could include recognizing the name of a tourist attraction, pinpointing a specific date for travel, or understanding the user's dietary preferences when making food recommendations. For example, when a user asks for restaurant options, the chatbot should be able to understand that “vegetarian” refers to a food preference or that “Bali” is a specific travel destination (Thodge et al., 2023).

In a recent study, a tourism-focused chatbot was developed with the capacity to provide detailed information on the rich ethnic diversity of North Sumatra. This system goes beyond just answering basic queries by generating comprehensive responses, complete with descriptive texts and images, to enhance the user experience. By leveraging AI (Artificial Intelligence), the chatbot is designed to simulate human-like conversations, responding to users' inquiries about various tourism services with accuracy and contextually relevant details. For example, if a user asks about traditional festivals or the best local cuisine, the system can provide not only the names of events or dishes but also background information, dates, and even images, all within a single conversation thread (Rizkiyani et al., 2021). This integration of intent recognition and entity extraction allows the chatbot to interact in a way that feels natural and informative, making it a powerful tool for tourism services. Through AI-driven dialogue, the chatbot becomes a virtual guide capable of catering to personalized requests, enhancing the tourism experience by delivering real-time, dynamic information.

The North Sumatra Tourism and Local Wisdom Chatbot Service successfully passed alpha testing with a 100% accuracy rate. In a system accuracy test involving 5 users, the system achieved a 90% accuracy rate, while beta testing with 30 respondents yielded a 93% satisfaction rate, with respondents either agreeing or strongly agreeing. These results were obtained by evaluating whether the chatbot provided responses that aligned with user input. Additionally, a questionnaire was administered to 30 respondents to assess the reliability of the data used. The reliability test of the questionnaire returned a score of 0.955 (very high), and the validity test confirmed the questionnaire's validity for use in this final project. Based on the conducted tests, it can be concluded that the system operates as intended (Nugraha & Sebastian, 2021).

In developing this chatbot, several methodologies were employed, notably Natural Language Processing (NLP), which focuses on enabling machines to better understand human language (Regin et al., 2022; Anggraeni et al., 2019). In addition to NLP, the RASA framework was used to build the chatbot, along with the BERT algorithm, which aids computers in understanding ambiguous language by using surrounding text to establish context. The system was also evaluated using a Non-Response-Rate test, achieving an accuracy score of 85%, indicating that the system performs according to its intended objectives (Rachmadi, 2020).

One of the tools and components that enables developers to build interactive and intelligent chatbots using natural language processing (NLP) and dialogue modeling is the Rasa framework. This popular open-source framework is widely used for developing chatbots and virtual assistants, offering a variety of tools and components that simplify the creation of such systems. Despite its capabilities, the use of chatbots built on the Rasa framework has not yet been widely adopted. Therefore, research and practical implementation focusing on the use of the Rasa framework to provide local wisdom information services in North Sumatra are necessary.

2. Method

Research Object Formulation:

a. Research Objects: Information services on local wisdom in North Sumatra and their utilization by tourists.
b. Research Population: Tourists visiting North Sumatra.
c. Research Sample: Tourists who agree to participate in interviews and fill out questionnaires.

Data Collection Methods:

a. Interviews: Conduct direct interviews with tourists visiting North Sumatra, using either structured or semi-structured interview guides.
b. Questionnaires: Distribute questionnaires to tourists for further data collection. These can be distributed via online platforms, such as email or online survey tools, or handed out at tourist locations.

System Development Method:

1. Research Instruments

Proper research instruments are essential for the success of this study. These include both software and hardware components, which are as follows:

Software:

a. ASUS laptop operating system
b. Python, the programming language used in conjunction with the Rasa.ai conversational platform
c. Visual Studio Code, the text editor employed for development

Hardware:

a. Asus A407M Laptop
b. Asus NT140WHM-N44 V8.0 NT140WHM-N44 V8.3 LCD

2. Chatbot System Design

The design of the chatbot system will be developed in stages. The first step involves input from the user in the form of questions. When a user asks about tourist attractions, the question will be processed by Rasa.ai, which has been selected by the researcher for developing this chatbot. Rasa.ai is an open-source machine learning framework designed for intelligent text-based or voice conversations. It can comprehend user inputs, engage in conversations, and integrate with various communication platforms and APIs. Rasa.ai operates at level 3 of conversational AI, part of a 5-level system. At this level, it can understand context, handle users changing their minds, and respond to unexpected questions. The Rasa.ai framework includes several key components, such as:

a. Intents

Intents represent the user's intention behind a message. For example, if a user starts a conversation with "hi," they are signaling the intent to initiate a chat.

b. Entities

Entities are specific pieces of information that can be extracted from a user's message. For instance, after greeting the bot, the bot might ask the user for their contact details, which serve as key identity information for the conversation.

c. Slots

Slots function as the bot's memory. They store any relevant information that needs to be retained during the conversation, such as the user's identity or other critical details.

d. Responses

Responses are the bot's replies to the user, based on the user's input. For the bot to deliver an accurate response, it must first understand the user's query and what it intends to convey.

3. Process Design

Following the chatbot system design, the process flow for building a chatbot is outlined, as illustrated in the diagram.

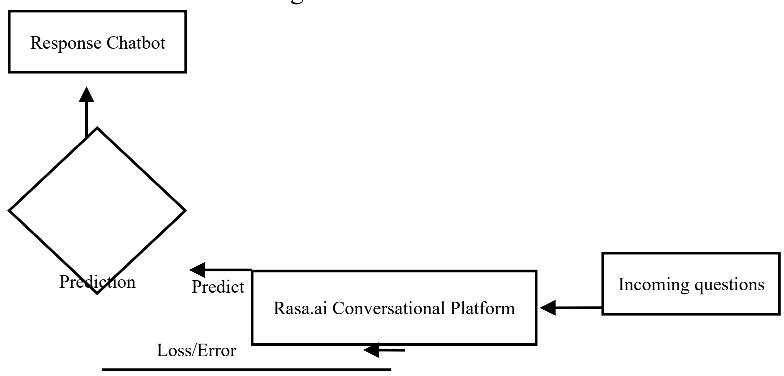

Fig. 1. Process Design

Figure 1 illustrates the data processing flow within the chatbot system. Before responding to user queries or statements, the chatbot requires the user to input questions related to tourist attractions and local wisdom in North Sumatra. The chatbot operates on source code that integrates external programming logic, enabling it to generate responses based on the user's input. It then makes predictions based on a precompiled dataset provided by the author and delivers responses tailored to the user's inquiries. Data Analysis Method: The data analysis in this study was conducted using relevant theories derived from the data collected during the research process. In line with the principles of descriptive research and a qualitative approach, data analysis was performed throughout the study. As stated by Sugiyono (2023), "data analysis in qualitative research is carried out before, during, and after fieldwork." This ensures that data analysis is a systematic process of searching, compiling, and categorizing the collected data to produce easily comprehensible conclusions. For this study, the researcher employed the data analysis techniques outlined by Saldana, Miles and Huberman (2014), who describe the key stages of data analysis as data reduction, data display, and conclusion drawing/verification. Accordingly, the steps taken by the researcher in analyzing the data were as follows:

A. Data Collection

The first step involves researchers entering the research environment to collect data. This stage is crucial for obtaining information about the issues occurring in the field. Researchers gather data essential for developing a chatbot that can effectively respond to frequently asked questions. The data collection process includes gathering relevant supporting information, conducting interviews and discussions with stakeholders, and observing various questions and answers. These observations typically focus on information about tourist attractions and local wisdom in North Sumatra, derived from multiple visited locations.

B. Data Reduction

The second step is data reduction, where researchers distill the collected data by summarizing, selecting key points, and concentrating on aspects pertinent to the implementation of the chatbot using the Rasa Framework for tourism information services and local wisdom in North Sumatra. This reduction process clarifies the data, facilitating easier collection of further information.

C. Data Presentation

The third step involves presenting the research data descriptively. This structured presentation allows researchers to understand the findings in written form, making it easier to interpret the data, draw conclusions, and decide on further actions.

D. Drawing Conclusions

The final step involves deriving conclusions from all the results, beginning with data reduction and data presentation, which are conducted concurrently. This approach simplifies the data analysis process. Consequently, the conclusion serves to summarize the findings of the study, providing clarity on what transpired after the analysis.

3. Results and Discussion

The results of the research involve using the Command Prompt platform to test the chatbot development process designed to provide information about tourism in North Sumatra and its local wisdom. The training and testing process using Rasa.ai involves the following steps:

1. Install Rasa

Rasa.ai is an open-source machine learning framework for developing intelligent text-based or spoken conversations. It can interpret user input, engage in conversations, and integrate with communication platforms and APIs.

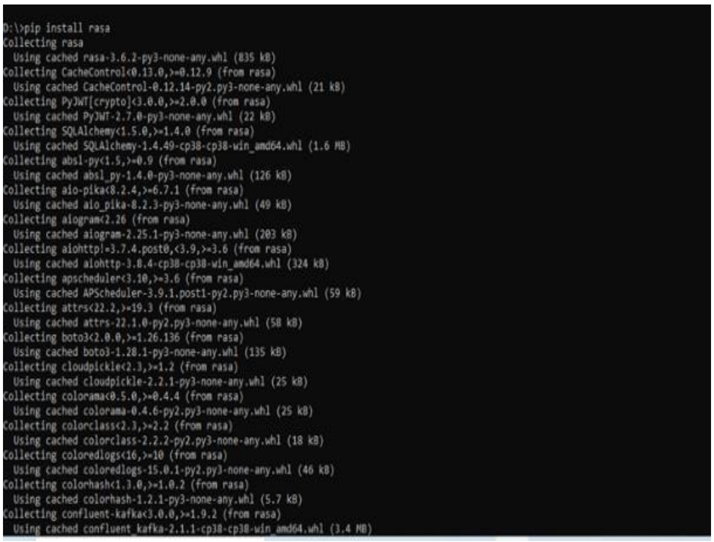

Fig. 2. Install Rasa

2. Create a New Directory

Before creating a chatbot dataset, you need to set up a new project directory.

a. D:\>mkdir my_wisata_project
b. D:\>cd my_wisata_project

3. Define the Chatbot Dataset

This study uses the Rasa.ai framework for chatbot development. Rasa is an open-source machine learning framework that includes Rasa Core and Rasa NLU. Rasa NLU processes natural language inputs, while Rasa Core manages dialogue and makes decisions based on provided knowledge. When developing chatbots with TensorFlow or PyTorch using Rasa, it's essential to be aware of the required specifications to achieve the desired results. Rasa Core features Dialogue Management, which involves several components. Researchers should be familiar with these components, as they include various files created to support chatbot conversations.

a. Data/nlu.md

The nlu.md file contains sample texts that represent various user questions relevant to the chatbot. To support chatbot conversations, researchers have prepared a file named nlu.md in the data folder. This file provides the necessary knowledge to the chatbot. Researchers have defined a total of 58 intents or topics, with 247 questions used to determine whether user queries match the chatbot's predefined intents.

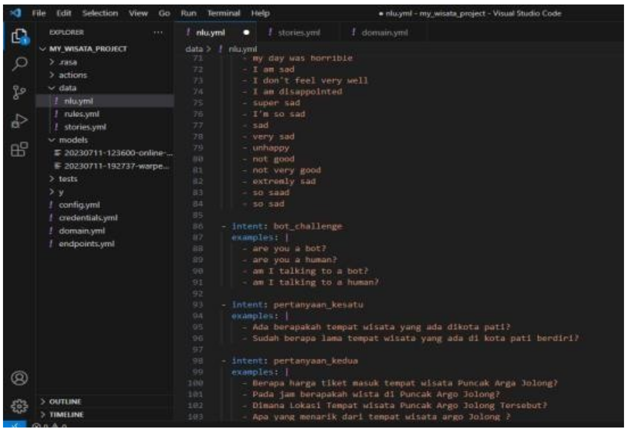

Fig. 3. Question Data Definition

b. Domain.yml

The domain.yml file is used to define intents, entities, slots, actions, and templates. Alongside the nlu.md file, the domain.yml file specifies each intent. Each intent can only provide response, which will be given to the user based on their question.

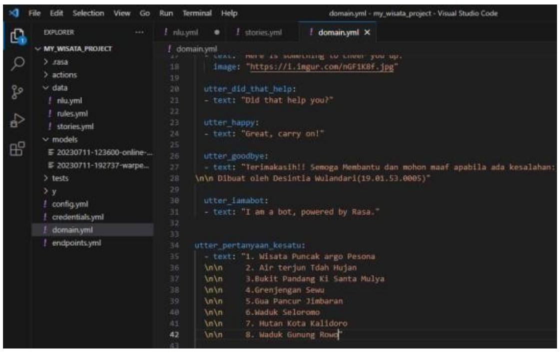

Fig. 4. Domain.yml File Definition

c. Data/stories.md

The stories.md file outlines the conversations the bot will have using the intents defined in the nlu.md file.

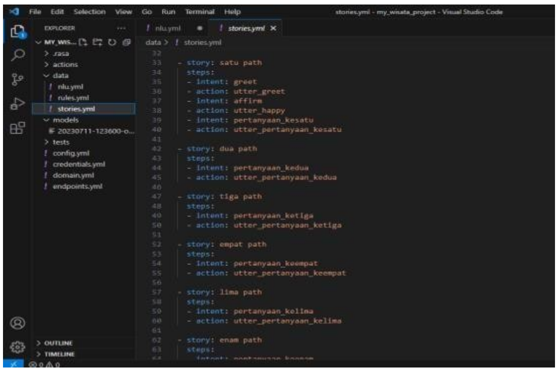

Fig. 5. Definition of Stories.md File

d. Training Model

Training the chatbot model is essential because it ensures the bot can provide appropriate responses to user queries. This training is performed once all data, such as questions and answers related to tourism information and local wisdom in North Sumatra, is prepared

4. Conclusion

In the discussion, researchers utilize the Anaconda platform to evaluate the chatbot testing process, aiming to ensure it effectively addresses user inquiries regarding tourism and local wisdom in North Sumatra. This platform facilitates a structured environment for testing and refining chatbot functionalities. To develop chatbots using frameworks like TensorFlow or PyTorch with Rasa, engineers must adhere to specific guidelines and provisions to achieve the desired outcomes. These guidelines are crucial for ensuring that the chatbot performs accurately and meets the expectations set forth in this study.

In Rasa Core, Dialog Management plays a central role in the chatbot's ability to manage conversations. Before delving into Dialog Management, it is essential to understand its components, which typically involve several key files. Researchers create and manage these files to support and enhance chatbot interactions. For defining the chatbot's conversational data, researchers utilize a file named nlu.md within the data folder. This file is instrumental in imparting knowledge to the chatbot, helping it understand and process user inputs effectively.

In conjunction with nlu.md, the domain.yml file is employed to define various intents. Each intent in this file corresponds to a specific type of user query and is designed to generate a single, relevant response. These responses are then delivered to users based on their inquiries, ensuring that the chatbot provides accurate and contextually appropriate answers. By carefully preparing these files and understanding their roles, researchers aim to optimize the chatbot's performance, ensuring it can effectively assist users with information about tourism and local wisdom in North Sumatra.

References

[1] Al-Jarf, R. (2021). Communicating and interacting with college students through a website Chatbox. International Journal of Management studies and Social Science Research (ijmsssr), 3(5), 106-114.

[2] Anggraeni, M., Syafrullah, M., & Damanik, H. A. (2019, May). Literation Hearing Impairment (I-Chat Bot): Natural Language Processing (NLP) and Naïve Bayes Method. In Journal of Physics: Conference Series (Vol. 1201, No. 1, p. 012057). IOP Publishing.

[3] Josphineleela, R., Kaliappan, S., Natrayan, L., & Bhatt, U. M. (2023, February). Intelligent Virtual Laboratory Development and Implementation using the RASA Framework. In 2023 7th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1172- 1176). IEEE.

Komalasari, R., Pramesti, P., & Harto, B. (2019). Teknologi Informasi E- Tourism Sebagai Strategi Digital Marketing Pariwisata. Altasia Jurnal Pariwisata Indonesia, 2(2).

[4] Nugraha, K. A., & Sebastian, D. (2021). Chatbot Layanan Akademik Menggunakan K-Nearest Neighbor. Jurnal Sains dan Informatika, 7(1), 11-19.

[5] Poerwanto, P., & Shambodo, Y. (2020). Revolusi industri 4.0: Googelisasi industri pariwisata dan industri kreatif. Journal of Tourism and Creativity, 4(1), 59-72.

[6] Rachmadi, M. F. (2020). Analisis Optimalisasi Teknologi Digital di Era Revolusi Industri 4.0 dalam Mengembangkan Kawasan Industri Pariwisata Halal guna Meningkatkan Perekonomian Lokal Kecamatan Gunungpati Kota Semarang. Jurnal DinamikA, 1(1), 39-53.

[7] Regin, R., Rajest, S. S., & Shynu, T. (2022). An automated conversation system using natural language processing (nlp) chatbot in python. Central Asian Journal of Medical and Natural Science, 3(4), 314-336.

[8] Rizkiyani, D. R., Sujatmoko, K., & Akhyar, F. (2021). Implementasi Virtual Costumer Service Dengan Robotic Process Automation (RPA) Dan Kecerdasan Buatan. eProceedings of Engineering, 8(6).

[9] Saldana., Miles & Huberman. (2014). Qualitative Data Analysis. America: Sage Publications.

[10] Sugiyono. (2023). Metode penelitian kuantitatif, kualitatif dan R&D. Bandung: Alfabeta.

[11] Thodge, A., Harsh, M., & Cyril, C. P. D. (2023, May). Web-based chatbot for basic financial and mortgage services. In 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN) (pp. 1-5). IEEE.