TIFDA JOURNAL
Technology Information and Data Analytic
Journal Website : http://jeptira.id/index.php/tifda
e-ISSN : 3064-0660

Implementation of Natural Language Processing in Digital File Management using Chatbot at The Center of PPKUKM Service

Bagus Tri Mahardika[1*], Rezza Maulana[1]
[1]Program Studi Teknologi Informasi Fakultas Teknik,Universitas DarmaPersada
[1]Taman Malaka Selatan Street No.8, East Jakarta, Indonesia
bagusunsada@gmail.com

Abstract — Digital document management is a significant challenge for government organizations, particularly when the information retrieval procedure remains manual and inefficient. This research seeks to create a chatbot system utilizing Natural Language Processing (NLP) to aid users in locating and reading digital files through natural language instructions. This system was deployed at the Data Center (Pusdatin) of the PPKUKM Office to expedite information retrieval and enhance document management efficiency. The system was constructed utilizing the Laravel framework for the user interface and Python's FastAPI for natural language processing. Features encompass document retrieval by name and date, along with the capability to exhibit document content directly. This research utilized the Waterfall software development methodology, encompassing stages of requirements analysis, system design, implementation, and testing. The final results indicate that the system operates effectively and offers a more adaptable and user-centric interaction experience.

Keywords – Natural Language Processing, FastAPI, Digital Documents, Waterfall

I. INTRODUCTION

In the contemporary digital age, governmental bodies and corporate entities are progressively dependent on electronic document management solutions. Crucial documents, including service letters, decrees, assignment letters, and assorted administrative files, are now frequently saved online to minimize paper consumption and enhance information accessibility. As the annual volume of archived documents rises, issues emerge regarding efficient data retrieval and management.

The local government sector, including the Department of Industry, Trade, Cooperatives, and Small and Medium Enterprises, generates hundreds of papers in many formats each month, encompassing activity reports and service application letters. These records are housed on servers or digital archiving systems lacking a standardized classification framework. Employees must utilize the search function by entering the file name or relevant keywords to locate a certain document. This search strategy is inherently constrained and frequently ineffectual, particularly when users are unaware of the precise terminology or layout utilized in the page.

The issue is exacerbated when documents exhibit diverse styles, uneven formatting, or are categorized ambiguously. Manual document organization and categorization impedes the search process and heightens the likelihood of errors or the omission of critical information. Frequently, time designated for studying document content is instead consumed by the search for it.

Conversely, the implementation of NLP-based chatbots might enhance the efficiency of information services by facilitating natural language conversations, hence simplifying user access to required information [1]. This offers a solution that can augment the intelligence of document management systems. Natural Language Processing (NLP) is a subset of artificial intelligence that allows computers to comprehend and manipulate human language. Natural Language Processing (NLP) in chatbots facilitates the analysis of incoming messages and source materials to deliver optimal and relevant responses. Integrating

NLP into digital file management systems enables searches to rely on both literal keywords and the meaning and context of the document's content. An employee seeking "food aid report" need not know the exact file name or date, as the NLP-based system can comprehend the search intent and correlate it with pertinent documents, even if the document bears a different title, such as "Social Fund Distribution Report".

Puspitasari states that Natural Language Processing (NLP) is a subset of artificial intelligence that allows computers to comprehend, interpret, and generate natural human language. In digital document management, NLP enhances the system's capacity to comprehend the context and significance of user inquiries, hence facilitating more pertinent and precise responses [4]. Salamun (2024) NLP allows information systems to do document searches based on both explicit keywords and human meaning and purpose. This is essential for overseeing extensive and diverse digital archives, especially inside governmental institutions [5].

A significant number of NLP models remain mostly in English. Stresses the necessity for NLP advancement tailored to local situations to yield more precise outcomes, particularly in public services and governmental administration [6][7]. NLP-driven chatbots are engineered to comprehend and address user inquiries in natural language. This technology allows chatbots to conduct text analysis, comprehend user intent, and deliver suitable responses depending on the available data [8].[9].

Implementing NLP through chatbots in digital file management systems can yield new and successful solutions [10]. Users may effortlessly input commands in colloquial English to look for, access, or manage particular documents. This accelerates the document retrieval process while diminishing dependence on inflexible and technical search frameworks.

II. METHODOLOGY

This research encompasses the design and development phase of a Natural Language Processing (NLP) chatbot system for digital file management. The system was engineered with technology capable of comprehending the user's natural language, correlating it with the contents of archived documents, and presenting search results in a pertinent and comprehensible manner. As per The System Development Life Cycle (SDLC) was utilized in the creation of this system, employing the waterfall methodology. The Software Development Life Cycle (SDLC) is a methodical framework organized in a sequential manner to facilitate the development of information systems (software). The Waterfall model is a traditional system development methodology, predominantly utilized by system developers, as seen in Figure 1.

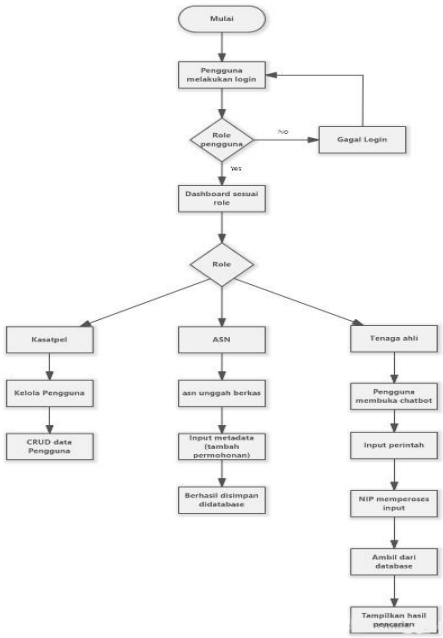

Figure 1. Flowchart algorithm

1. Analysis
   This phase involves examining the system requirements essential for the design and development of the system.

2. Desain Stage Plan
   The system display design process is conducted based on the outcomes of the needs analysis performed during the analysis stage.

3. Implementation Phase Design

   At this juncture, it involves the execution of a plan, action, or policy to attain objectives.

4. Testing Phase Design
   This is a testing phase for the developed system, intended to assess its viability and identify its deficiencies.

5. System Maintenance Phase Design
   This is the concluding phase of software development, during which the system may be modified and enhanced based on user requirements.

III. RESULTS AND DISCUSSION

This section delineates the outcomes of the system's implementation and testing, along by an analysis of the findings. The analysis relies on data acquired from testing to evaluate system performance and determine the degree to which the system fulfills the objectives set during the design phase. The discourse additionally encompasses an analysis of the findings and comparisons with analogous studies, if applicable.

1. Home view of login page

The subsequent page is the login interface, serving as the initial display for program access. It features a window for inputting the email and password required to access the program. Should the email and password be supplied incorrectly, the system will notify the user to input a legitimate email or password. Upon validation, they may access the dashboard page in accordance with their designated duties, as illustrated in Figure 2.

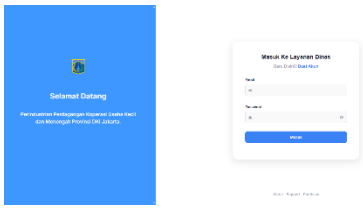

Fig 2. Home Login Page

2. Register page view

The Registration page is where one establishes an account. It features a form for inputting your name, email address, phone number, password, and password confirmation. Upon completion of all data entry and user consent to the terms and conditions, the registration process may proceed by selecting the "Submit" button to establish an account.

3. Service menu page view

This page is as a common repository for Civil Servants (ASN) to submit diverse service requests. This site allows ASN to examine a list of existing service requests, search for specific requests via the search field, and create new service requests using the "Add Request" button. The table presents essential information for each request, including the service type, applicant's name, position, regional government agency (SKPD), email address, WhatsApp number, attachments, implementation date, and a "Action" column for subsequent actions, as illustrated in Figure 3.

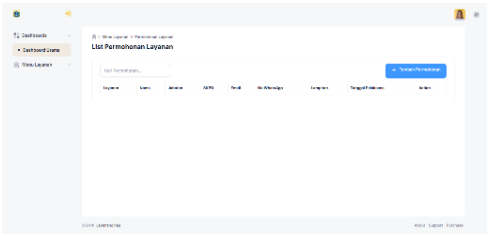

Fig 3. Service menu page

4. File upload page view

This page facilitates the addition of application services and the uploading of files. Here, ASN can choose the category of service or file for upload. This area allows ASN to submit service requests or upload necessary paperwork. This display contains several elements to be completed, including the selection of service/ file category, applicant's name, position, SKPD, email, active WhatsApp number, and a column for uploading necessary file attachments. Furthermore, a column exists to ascertain the implementation date. Upon the accurate completion of all information and the successful upload of files, ASN may either store the application for subsequent processing or terminate the process by selecting the "Close" option, as seen in Figure 4.

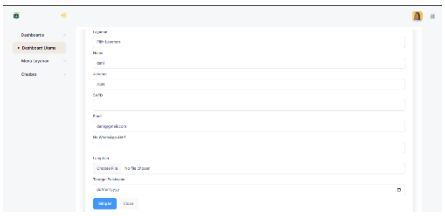

Fig 4. Login pages

5. Page chatbot

This portal serves as a chatbot interface designed for locating files uploaded by public servants. This page facilitates the search for previously submitted PDF files, particularly for civil personnel. Users can engage with the chatbot by entering requests or inquiries in the designated fields, such as, "Display the May service request file." Upon entering the request, the user only clicks the "Send" button, prompting the chatbot to process the request and either display or direct the user to the pertinent file. This function is highly beneficial for the effective management and retrieval of critical papers, as illustrated in Figure 5.

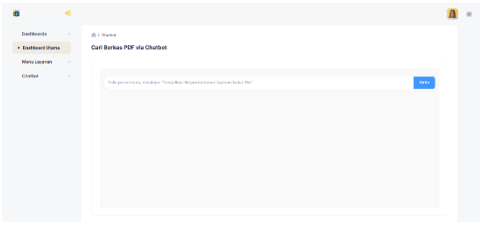

Fig 5. Page Chatbot

6. Search page by name

The file search functionality by name enables users to locate papers by inputting the applicant's name via standard text entry. The system employs Natural Language Processing (NLP) technology to automatically identify and correlate keywords associated with the provided name, regardless of its completeness or formal presentation, as illustrated in Figure 6.

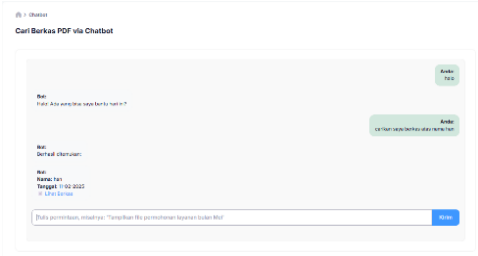

Fig 6. Search page by name

7. Page Fitur summary pdf

The Summary PDF feature has been evaluated to confirm its capacity to autonomously and effectively condense document content. Evaluations of various document types, including application letters, activity reports, and meeting minutes, demonstrate that the system can produce pertinent summaries with high precision, encapsulating essential information such as the purpose, names of involved parties, and the date of the event. The text extraction method with PyMuPDF and summarization via an NLP model was efficient, averaging under five seconds per page. The summary findings were carefully checked, confirming that the algorithm can generate meaningful summaries while preserving the essential context of the original content, as illustrated in Figure 7.

Fig 7. Feature Summary

8. Intent Labeling results

The intent labeling method serves as the preliminary phase in developing an NLP-based classification model. Each user statement is examined and classified into a specific, preset intent. This labeling seeks to clarify the intent or intention of the user's message, enabling the system to respond properly. This study involves a total of [number] data points labeled with [number of intents] distinct intents, including greeting, show_all_requests, read_pdf_content, summarizepdf, among others, as illustrated in Figure 8.

```
{"text": "woi", "label": "greeting"}
{"text": "haii", "label": "greeting"}
{"text": "yo bro", "label": "greeting"}
{"text": "assalamualaikum bot", "label": "greeting"}
{"text": "bisa bantu?", "label": "greeting"}

{"text": "boleh lihat seluruh permohonan?", "label": "show_all_requests"}
{"text": "munculkan semua permohonan layanan", "label": "show_all_requests"}
{"text": "apa saja permohonan yang masuk", "label": "show_all_requests"}
{"text": "daftar semua permintaan", "label": "show_all_requests"}
{"text": "tampilkan semuanya", "label": "show_all_requests"}
```

Fig 8. Intent Labeling Results

9. Evaluation Results

Upon completion of the model training process, the subsequent stage is evaluation to assess the model's efficacy in accurately classifying intentions. The assessment is performed utilizing test data that has been previously segregated from the training data. The assessment techniques employed consist of quantifying accuracy, precision, recall, and F1-score for each intent. This evaluation seeks to assess the model's efficacy in discerning user intent and to pinpoint the strengths and flaws of the produced system, as seen in Figure 9.

	precision	recall	f1-score	support
cek_asn_baru	1.00	1.00	1.00	3
closing	1.00	1.00	1.00	2
greeting	1.00	1.00	1.00	3
read_pdf_content	1.00	1.00	1.00	2
search_by_criteria	1.00	1.00	1.00	5
search_by_topic	1.00	1.00	1.00	5
show_all_requests	1.00	1.00	1.00	3
summarize_pdf	1.00	1.00	1.00	2
accuracy			1.00	25
macro avg	1.00	1.00	1.00	25
weighted avg	1.00	1.00	1.00	25

Fig 9. Evaluation Result

IV. CONCLUSION

The research and implementation results indicate that Natural Language Processing (NLP) technology can be effectively utilized to enhance digital document management. Incorporating NLP into the system facilitates data retrieval and processing using natural language input, eliminating the necessity for technical commands, thus enhancing the system's efficiency and adaptability. Moreover, the deployment of NLP-based chatbots has demonstrated efficacy in assisting users from diverse backgrounds, including civil officials, task force leaders, and specialists, in locating and retrieving digital documents. Chatbots can comprehend commands expressed in standard language and can also read and present document contents directly, hence enhancing convenience, speed, and accessibility in system utilization.

REFERENCES

[1] M. Furqan, S. Sriani, and M. N. Shidqi, “Chatbot Telegram Menggunakan Natural Language Processing,” Walisongo Journal of Information Technology, vol. 5, no. 1, pp. 15–26, Jun. 2023, doi: 10.21580/wjit.2023.5.1.14793.

[2] M. Akbar Nasution et al., “Implementasi NLP Dalam Pembuatan Chatbot Customer Service Publisher Jurnal Studi Kasus LARISMA,” 2024.

[3] Simatupang Simatupang and Sianturi Setiawan, “Perancangan Sistem Informasi Pemesanan Tiket Bus pada PO. Handoyo Berbasis Online,” 2020.

[4] A. Puspitasari, A. N. Paradhita, Y. W. Tineka, Sulistyowati, N. K. S. Noriska, and Haryanto, “Natural Language Processing (NLP) Technology for Chatbot Website,” Jurnal Penelitian Pendidikan IPA, vol. 10, no. SpecialIssue, pp. 319–324, Aug. 2024, doi: 10.29303/jppipa.v10ispecialissue.8241.

[5] S. Salamun, Aldi Aprialdo, and Sukri, “Optimasi Chatbot dengan Pemanfaatan Natural Language Processing,” Jurnal Komputer Terapan, vol. 10, no. 1, pp. 17–26, May 2024, doi: 10.35143/jkt.v10i1.6181.

[6] J. Khatib Sulaiman, M. Abid Nadzif, and U. Stikubank Semarang, “Penggunaan Teknologi Natural Language Processing dalam Sistem Chatbot Untuk Peningkatan Layanan Informasi Administrasi Publik,” Indonesian Journal of Computer Science Attribution, vol. 13, no. 1, pp. 2024–1227, 2024.

[7] B. Patra, “Natural Language Processing in Chatbots: A Review,” Turkish Journal of Computer and

Mathematics Education, Dec. 2020, doi: 10.61841/turcomat.v11i3.14655.
[8] R. Fachrizal, "Pengembangan Aplikasi Chatbot Berbasis NLP (Natural Language Processing)," 2024.
[9] Ponkumar, D. D. N., Duraisamy, G., Teja, D. D., Rohith, A. V. S., Ashok, C., & Tejesh, O. S. S. N. (2024). G-bot: Illuminating Government Programs with Intelligent Assistance. 1270–1275. https://doi.org/10.1109/icdici62993.2024.10810920
[10] R. Dreyling, T. Koppel, T. Tammet, and I. Pappel, "Challenges of Generative AI Chatbots in Public Services -An Integrative Review," Jan. 2024, doi: 10.2139/ssrn.4850714.