

# Sistem Tanya Jawab Berbahasa Indonesia dengan Memanfaatkan Natural Language Inference = Utilizing Natural Language Inference for Indonesian Question Answering System

Muhammad Ravi Shulthan Habibi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920535382&lokasi=lokal>

---

## Abstrak

Sistem tanya jawab merupakan salah satu tugas dalam domain natural language processing (NLP) yang sederhananya bertugas untuk menjawab pertanyaan sesuai konteks yang pengguna berikan ke sistem tanya jawab tersebut. Sistem tanya jawab berbahasa Indonesia sebenarnya sudah ada, namun masih memiliki performa yang terbilang kurang baik. Penelitian ini bereksperimen untuk mencoba meningkatkan performa dari sistem tanya jawab berbahasa Indonesia dengan memanfaatkan natural language inference (NLI). Eksperimen untuk meningkatkan sistem tanya jawab berbahasa Indonesia, penulis menggunakan dua metode, yaitu: intermediate-task transfer learning dan task recasting sebagai verifikator. Dengan metode intermediate-task transfer learning, performa sistem tanya jawab berbahasa Indonesia meningkat, hingga skor F1-nya naik sekitar 5.69 dibandingkan tanpa menggunakan pemanfaatan NLI sama sekali, dan berhasil mendapatkan skor F1 tertinggi sebesar 85.14, namun, peningkatan performa dengan metode intermediate-task transfer learning cenderung tidak signifikan, kecuali pada beberapa kasus khusus model tertentu. Sedangkan dengan metode task recasting sebagai verifikator dengan parameter tipe filtering dan tipe perubahan format kalimat, performa sistem tanya jawab berbahasa Indonesia cenderung menurun, penurunan performa ini bervariasi signifikansinya. Pada penelitian ini juga dilakukan analisis karakteristik pasangan konteks-pertanyaan-jawaban seperti apa yang bisa dijawab dengan lebih baik oleh sistem tanya jawab dengan memanfaatkan NLI, dan didapatkan kesimpulan bahwa: performa sistem tanya jawab meningkat dibandingkan hasil baseline-nya pada berbagai karakteristik, antara lain: pada tipe pertanyaan apa, dimana, kapan, siapa, bagaimana, dan lainnya; kemudian pada panjang konteks 100 dan 101-150; lalu pada panjang pertanyaan 5 dan 6-10; kemudian pada panjang jawaban golden truth 5 dan 6-10; lalu pada keseluruhan answer type selain law dan time; terakhir pada reasoning type WM, SSR, dan MSR.

.....

The question-answering system is one of the tasks within the domain of natural language processing (NLP) that, in simple terms, aims to answer questions based on the context provided by the user to the question-answering system. While there is an existing Indonesian question-answering system, its performance is considered somewhat inadequate. This research conducts experiments to improve the performance of the Indonesian question answering system by utilizing natural language inference (NLI). In order to enhance the Indonesian question-answering system, the author employs two methods: intermediate task transfer learning and task recasting as verifiers. Using the intermediate-task transfer learning method, the performance of the Indonesian question-answering system improves significantly, with an increase of approximately 5.69 in F1 score compared to not utilizing NLI at all, achieving the highest F1 score of 85.14. However, the performance improvement with the intermediate-task transfer learning method tends to be non-significant, except in certain specific cases and particular models. On the other hand, employing the task recasting method as a verifier with filtering parameter type and sentence format change type leads to a decline in the performance of the Indonesian question-answering system, with the significance of this performance

decrease varying. Additionally, this research conducts an analysis on the characteristics of context-question-answer pairs that can be better answered by the question-answering system utilizing NLI. The findings conclude that the question-answering system's performance improves compared to its baseline across various characteristics, including different question types such as what, where, when, who, how, and others. Furthermore, it improves with context lengths 100 and 101-150, question lengths 5 and 6-10, as well as answer lengths (golden truth) 5 and 6-10. Additionally, it performs better in overall answer types excluding law and time, and lastly, in reasoning types WM, SSR, and MSR.