

Abstrak

Karena Al-Qur'an adalah kitab suci utama bagi umat Muslim, penelitian ekstraksi informasi tentang teks-teks Al-Qur'an, terutama dalam bentuk Ekstraksi Entitas Manusia, merupakan tugas penting untuk pemahaman lebih lanjut Al-Qur'an beserta tafsirnya. Tantangan dalam mengekstrak entitas manusia dari teks Al-Qur'an adalah banyak ayat yang memiliki struktur yang kompleks, sehingga penting untuk membangun sistem yang dapat mengekstrak entitas secara otomatis, akurat, dan cepat. Ekstraksi Entitas Manusia pada Al-Qur'an sendiri adalah tugas yang bertujuan untuk mengekstrak entitas manusia dalam sebuah kalimat atau ayat, seperti nama seseorang, nama kelompok, dll pada teks Al-Qur'an. Contoh input diambil dari potongan Surat Al-Baqarah ayat 46 yang berbunyi "Yaitu orang-orang yang meyakini bahwa mereka akan menemui tuhan mereka dan bahwa mereka akan kembali kepadanya" dari input tersebut sistem ekstraksi entitas manusia diharapkan dapat mengidentifikasi entitas manusia yaitu "orang-orang yang meyakini bahwa mereka akan menemui tuhan mereka". Saat ini, penelitian Ekstraksi Entitas Manusia untuk Al-Qur'an belum banyak dilakukan, hanya beberapa algoritma dengan hasil yang tersebar. Dalam penelitian ini, kami menggunakan beberapa model supervised yaitu Conditional Random Field (CRF), BiLSTM-CRF, dan model pretrained deep learning transformer IndoBERT, kemudian kami akan menerapkan dan melakukan analisis komparatif untuk kinerja beberapa model tersebut. Kami menemukan bahwa model berbasis deep learning, yaitu BiLSTM-CRF, berkinerja terbaik dalam mengekstraksi entitas manusia, sementara model berbasis probabilistik, yaitu CRF, mengalami kesulitan dalam mengekstraksi entitas manusia, khususnya entitas manusia bersarang.

Kata kunci : Al-Quran, Ekstraksi Entitas Manusia, Supervised Learning, Studi Komparatif, Entitas Bersarang