

# 1. Pendahuluan

## Latar Belakang

Al-Quran merupakan kitab suci sekaligus pedoman hidup bagi seluruh umat muslim (orang yang memeluk agama Islam) [6]. Al-Quran merupakan wahyu atau perkataan Allah Swt. kepada Nabi Muhammad Saw. dengan melalui perantara malaikat Jibril. Al-Quran berisi 30 juz, 114 surat, dan 6.236 ayat. Dalam Al-Quran itu sendiri, Al-Quran menjelaskan bahwa Al-Quran merupakan sebuah petunjuk yang wajib untuk dipahami dan diikuti oleh semua umat muslim. Al-Quran diturunkan dalam kurun waktu sekitar 23 tahun (lebih tepatnya 22 tahun, 2 bulan, dan 22 hari) yang dimulai ketika Nabi Muhammad Saw. berumur 40 tahun pada 17 Ramadan hingga Beliau wafat pada umur 63 tahun.

Ekstraksi entitas bernama adalah salah satu tugas fundamental dalam pemrosesan bahasa alami (*natural language processing/NLP*) yang melibatkan pengidentifikasian dan pengklasifikasian entitas-entitas bernama, seperti orang, organisasi, lokasi, tanggal, dan istilah-istilah khusus lainnya dalam sebuah teks [5]. Entitas-entitas bernama ini mengandung informasi penting dan memiliki peranan yang signifikan dalam pemahaman dan analisis data teks di berbagai bidang.

Berbagai macam pendekatan, sistem, dan model telah dikembangkan untuk melakukan tugas ekstraksi entitas manusia ini, seperti metode statistik, metode berbasis aturan, dan metode yang berbasis pembelajaran mesin. Namun, ketika diterapkan pada terjemahan bahasa Indonesia dari Al-Quran, pendekatan-pendekatan tersebut memiliki beberapa keterbatasan atau tantangan tertentu. Tantangan yang dihadapi mencakup adaptasi aturan, perolehan data yang teranotasi, dan penanganan karakteristik kebahasaan yang khas dari terjemahan Al-Quran. Oleh karena itu, diperlukan sebuah sistem khusus yang dapat disesuaikan dengan terjemahan bahasa Indonesia dari Al-Quran. Sistem ini akan mempertimbangkan konteks yang unik, terminologi yang digunakan, dan nuansa kebahasaan yang menjadi ciri khas dalam terjemahan Al-Quran, sehingga dapat menghasilkan ekstraksi entitas manusia yang lebih akurat dan tepat. Beberapa penelitian terkait telah menyinggung tantangan dan peluang untuk mengembangkan sistem pengenalan entitas manusia untuk teks yang termasuk religius seperti Al-Quran, serta telah mengeksplorasi pendekatan yang khusus untuk berbagai bahasa dan terjemahan.

Contoh hasil ekstraksi entitas manusia dapat dilihat pada tabel 1. Tabel tersebut berisikan terjemahan ayat Al-Quran yang dibagi per kata, ditambah dengan label di sebelah kanan.

words	labels
'(yaitu)'	'O'
'jalan'	'O'
'orang-orang'	'B-PER'
'yang'	'I-PER'
'telah'	'I-PER'
'Engkau'	'I-PER'
'beri'	'I-PER'
'nikmat'	'I-PER'
'kepadanya;'	'O'
'bukan'	'O'
'(jalan)'	'O'
'mereka'	'B-PER'
'yang'	'I-PER'
'dimurkai;'	'I-PER'
'dan'	'O'
'bukan'	'O'
'(pula'	'O'
'jalan)'	'O'
'mereka'	'B-PER'
'yang'	'I-PER'
'sesat.'	'I-PER'

**Tabel 1.** Hasil Ekstraksi Entitas Manusia pada Al-Quran, Surat Al-Fatihah ayat ke-tujuh

Pada kolom label kita dapat melihat ketiga kelas yang digunakan pada penelitian ini, yakni '*B-PER*', '*I-PER*', dan '*O*'. '*B-PER*' merupakan kependekan dari *Beginning of Person entity* yang merupakan tanda awalan bahwa kata tersebut tergabung ke dalam entitas *Person* atau manusia, '*I-PER*' adalah *Inside of person* yang artinya kata tersebut merupakan kata yang termasuk ke dalam entitas manusia dan posisinya tidak berada di depan atau di awal, termasuk di tengah kalimat atau di akhir kalimat.

#### **Topik dan Batasannya**

Penelitian ini difokuskan pada ekstraksi entitas manusia dari Al-Quran terjemahan Bahasa Indonesia. Tujuan penelitian ini adalah untuk mengidentifikasi dan mengekstraksi entitas manusia yang terdapat dalam teks Al-Quran yang telah diterjemahkan ke dalam Bahasa Indonesia. Penelitian ini memiliki batasan-batasan sebagai berikut:

1. Jenis Entitas yang Diekstraksi: Fokus penelitian ini terbatas pada entitas manusia, seperti nama-nama orang, gelar, dan atribut manusia lainnya yang terdapat dalam teks Al-Quran terjemahan Bahasa Indonesia dan tidak akan melakukan ekstraksi pada entitas selain entitas manusia seperti entitas tempat, organisasi ataupun waktu.
2. Metode yang Digunakan: Penelitian ini menggunakan pendekatan berbasis pembelajaran mesin untuk ekstraksi entitas manusia. Metode yang digunakan menggunakan metode *Bi-LSTM* atau *Bi Directional LSTM*.
3. Korpus Data yang Digunakan: Penelitian ini memanfaatkan korpus data yang terdiri dari teks Al-Quran yang telah diterjemahkan ke dalam Bahasa Indonesia. Korpus data tersebut terdiri dari satu atau beberapa terjemahan Al-Quran yang tersedia dalam Bahasa Indonesia.
4. Dataset yang digunakan dibatasi pada terjemahan Al-Quran bahasa Indonesia untuk surat pertama (Surat Al Fatihah) sampai dengan surat ke-enam (Surat Al An'am).

#### **Tujuan**

Tujuan utama dari penelitian ini adalah melakukan analisa dari suatu sistem yang dapat mengekstraksi entitas manusia dari teks Al-Quran terjemahan Bahasa Indonesia. Hasil analisa dari sistem tersebut diharapkan dapat menghasilkan nilai atau skor yang cukup bagus untuk matrik evaluasinya. Matrik evaluasi mencakup akurasi, *f1 score*, *recall*, dan *precision*. Serta memperkirakan hal apa yang harus dilakukan untuk meningkatkan skor dari matriks evaluasi tersebut.