

## 1. Pendahuluan

Bagian ini berisikan bahasan lebih detail dari abstrak mengenai latar belakang, topik dan batasan, dan tujuan penulisan tugas akhir ini. Terdapat juga rencana susunan penulisan tugas akhir ini.

## Latar Belakang

Saat ini telah banyak penelitian linguistik dan *Natural Language Processing* (NLP) yang memanfaatkan Princeton WordNet (PWN) [7] sebagai bahan penelitian seperti pada pembangunan WordNet untuk Bahasa Turki [3], Bahasa Arab [5], Bahasa China [15], pembangunan Synonym Set untuk Bahasa Indonesia [13], dan masih banyak penelitian lainnya.

Penelitian ini memiliki tujuan untuk membangun *synonym set* atau *synset* Bahasa Arab pada Al-Quran dengan memanfaatkan *lexical semantic similarity* pada PWN. Pada penelitian sebelumnya mengenai pembangunan WordNet Bahasa Arab [5] telah berhasil membangun *lexical database* WordNet dengan fokus Bahasa Arab modern. Pada penelitian ini akan berfokus pada kosa kata dalam Al-Quran.

Al-Quran memiliki surah sebanyak 114 surah dan ayat berjumlah 6.236 ayat. Dari sekian banyak ayat yang ada, setiap ayat memiliki kaitan dengan sebagian ayat yang lainnya. Begitupun dengan kata dalam setiap ayat juga memiliki kaitan dengan sebagian kata yang lainnya, salah satu kaitannya yaitu kesamaan makna dari setiap kata. Bagi pembaca yang kurang memahami Bahasa Arab dan hanya membaca beberapa kata saja dalam Al-Quran, dapat menyebabkan kurangnya pemahaman mengenai arti sebenarnya dari setiap kata dalam AlQuran.

Setiap kata dalam Al-Quran dapat memiliki makna lebih dari satu, contohnya adalah kata آية (āyat) yang memiliki arti *sign* (QS. 2:39:4) dan *verse* (QS. 2:99:4). Lalu kata كِتَابِي (kitābī) yang memiliki arti *book* (QS. 2:2:2), *scripture* (QS. 2:23:8), dan *record* (QS. 11:6:15). Selain kata yang memiliki makna lebih dari satu, sebagian kata juga memiliki arti yang sama atau dekat, contohnya adalah kata رَبِّ (rab) yang memiliki arti *Lord* (QS. 1:2:3) dan إِلَهٍ (ilāh) yang memiliki arti *God* (QS. 2:133:17) dimana *Lord* dan *God* memiliki kedekatan makna yaitu Tuhan.

Sumber daya bahasa sangat penting untuk penelitian komputasi linguistik khususnya *Natural Language Processing* (NLP). Namun untuk korpus Bahasa Arab Al-Quran saat ini masih tergolong kurang walaupun beberapa tahun terakhir telah ada pengembangan korpus korpus Bahasa Arab yang bebas akses. Korpus-korpus tersebut dapat dikatakan belum cukup mendorong para ahli bahasa untuk menerapkannya pada studi berbasis korpus mereka [2]. Oleh karena

itu penelitian ini menjadi penting untuk membangun *synonym set* yang dapat digunakan sebagai *prototype* untuk membangun tesaurus Al-Quran. *Synonym set* yang terbentuk juga dapat digunakan untuk lebih memahami kosa kata – kosa kata pada Al-Quran.

Pendekatan utama dari penelitian ini adalah WordNet dan kamus monolingual yaitu tesaurus. Seperti yang kita tahu, WordNet merupakan *lexical database* Bahasa Inggris yang populer dan banyak digunakan oleh peneliti bidang komputasi linguistik dan NLP. Tesaurus juga merupakan kamus monolingual yang menyediakan hubungan sinonim antar kata dalam suatu bahasa [11] dan dapat diakses dengan bebas pada [thesaurus.com](http://thesaurus.com) atau [www.lexico.com](http://www.lexico.com). Tersedia juga tesaurus dalam bentuk offline [11], [14].

### **Topik dan Batasannya**

Berdasarkan latar belakang yang telah dijabarkan, topik-topik yang terkait dari tugas akhir ini diantaranya yaitu:

#### 1. *Synonym Set*

Hasil akhir dari tugas akhir ini adalah berupa himpunan atau *synonym set* kata yang memiliki kedekatan arti.

#### 2. Tesaurus Al-Quran

*Synonym set* yang dibangun dapat digunakan untuk menambah korpora Al-Quran salah satunya yaitu Tesaurus Al-Quran. Tesaurus Al-Quran berarti berbentuk himpunan kata dengan kata *arabic* sebagai entri dan dimana antar kata memiliki kedekatan arti.

#### 3. WordNet

Menghitung kedekatan antar kata menggunakan *PATH similarity* pada WordNet yaitu mengukur kedekatan antar kata dengan menghitung jumlah node pada jalur terpendek antar kata.

Adapun batasan-batasan untuk penulisan tugas akhir ini diantaranya:

1. Jenis kata yang digunakan hanya kata benda (noun).
2. Terjemahan kata yang digunakan untuk mengukur kedekatan kata adalah terjemahan Bahasa Inggris.

### **Tujuan**

Tujuan dari penulisan tugas akhir ini diantaranya:

1. Membangun *synonym set* sebagai *prototype* untuk pembuatan korpus Al-Quran.
2. Menghitung akurasi hasil *synonym set* yang terbentuk menggunakan F-Measure.

### **Organisasi Tulisan**

Bagian-bagian pada penulisan ini terdiri dari: Bagian 1 mengenai jabaran masalah. Bagian 2 membahas teori-teori yang terkait dengan penelitian ini. Bagian 3 mengenai sistem yang dibangun. Bagian 4 membahas hasil penelitian beserta evaluasi menggunakan F-Measure. Bagian 5 berisi kesimpulan akhir dari penelitian ini.