

1. Pendahuluan

BAB I Berisi permasalahan serta tujuan yang ingin dicapai pada penelitian ini serta beberapa perencanaan yang akan dilakukan dari mengidentifikasi masalah sampai laporan akhir.

Latar Belakang

Al-Qur'an merupakan mukjizat abadi dalam islam dengan pengetahuan dan teknologi yang semakin maju saat ini [1]. Al-Qur'an terdiri dari 114 Surah, 30 Juz dan 6236 Ayat. Setiap ayat Al-Qur'an memiliki makna yang berbeda [2]. Setiap ayat dalam al-qur'an memiliki kesamaan makna antar ayat lainnya yang disebut dengan sinonim set. Sinonim set adalah kumpulan dari satu atau kata-kata yang memiliki arti atau sinonim yang sama. Setiap kata yang memiliki makna yang sama dapat menggantikan bagian dari penggunaan kata tanpa mengubah arti atau makna kalimat [3], bahasa Indonesia juga memiliki database yang dibentuk menjadi situs web untuk mencari sinonim dengan referensi tesaurus bahasa Indonesia [4]. Sinonim yang valid di WordNet hanya berfungsi untuk dua kata yang memiliki hubungan sinonim, jika kedua kata tersebut tidak memiliki hubungan sinonim, kedua kata tersebut tidak valid [5].

Adapula sistem pencatatan sinonim dari kata-kata yang serupa yaitu WordNet, tetapi sistem word-Net yang ditemukan hanya menggunakan pencarian bahasa Inggris [6]. WordNet tidak hanya berisi kata-kata yang sinonim tetapi juga mengandung makna yang berlawanan [7]. Tidak hanya WordNet mampu memuat kata-kata yang memiliki arti yang sama, bahasa Indonesia juga memiliki buku sinonim dari dua kata atau lebih yang memiliki arti yang sama. Tesaurus berisi kosakata bahasa yang sesuai dengan hubungan kedekatan. Tesaurus dapat memudahkan pembaca untuk menemukan kata-kata [8]. Tesaurus dapat membantu pengindeksan mencari deskriptor untuk kata kunci dari dokumen [9].

Algoritma Clustering Seperti Hierarchical Clustering bisa menjadi salah satu algoritma yang dapat membangun sinonim set dengan data set yang berasal dari Ayat Ayat Al-Qur'an. Algoritma Hierarchical Clustering dipilih berdasarkan fungsinya untuk membentuk cluster dari setiap titik data dan dapat membangun cluster dengan membedakan setiap objek yang dipilih dalam mode top-down atau bottom-up. Kelebihan dari algoritma Hierarchical clustering adalah dapat menyajikan pengelompokan data menggunakan pengukuran jarak, ini sangat cocok untuk diterapkan untuk mengukur jarak antara kata-kata dalam kumpulan data dan juga hierarchical clustering melakukan pengelompokan hierarkis yang dapat membangun satu set sinonim yang sebelumnya tidak diketahui berapa banyak sinonim yang akan ditemukan[10]. Sebelumnya sudah membandingkan dengan algoritma K-Means Clustering tetapi ternyata tidak cocok untuk diterapkan pada dataset ini dalam menemukan set sinonim. karena pada awal harus menentukan K dalam K-Means Clustering dan mengetahui berapa Sinonim Set yang akan dihasilkan oleh sistem.

Penelitian ini akan menerapkan Hierarchical Clustering untuk membangun grouping sinonim set menggunakan bahasa inggris dan bahasa arab, Dataset seluruh ayat quran dihasilkan dari openburhan.net . Hasil dari sinonim set akan divalidasi menggunakan Gold Standar. Gold Standar adalah jenis evaluasi yang diperlukan dalam menyelaraskan hasil program komputer, Gold Standar berguna untuk pendapat ahli yang dapat diterima sebagai referensi yang akurat [11]. Tujuan penelitian ini adalah menghasilkan kumpulan pengelompokan sinonim menggunakan algoritma hierarchical clustering dengan input kata Inggris dan output Arab serta meningkatkan akurasi hasil clustering. Kumpulan sinonim yang dihasilkan oleh sistem adalah kata yang memiliki arti yang sama dengan kata yang dimasukkan. maka sistem akan mengelompokkan untuk memfilter hasil ke dalam bahasa Arab saja dan membuatnya menjadi beberapa level yang dibentuk oleh sebuah cluster. Cluster yang ditampilkan berarti bahwa grup pertama yang muncul berarti kata-kata dalam grup tersebut memiliki arti yang dekat, dan grup berikutnya yang muncul berarti bahwa jarak kesamaan tidak terlalu dekat. Tujuan penelitian ini dibuat berdasarkan permasalahan yang ada yaitu banyaknya penelitian untuk membangun sinonim set namun belum ada yang menghasilkan sinonim set dengan menggunakan metode grouping sinonim set. Manfaat grouping sinonim set adalah dapat mendapatkan hasil

sinonim berdasarkan jarak kedekatan maknanya dari kata yang diinputkan. Penelitian ini memiliki batasan yaitu hanya memproses kosakata kata benda

Penelitian ini juga merupakan penelitian lanjutan dari sebuah artikel berjudul "Pengembangan Sinonim dari Al-Qur'an Vocabulary with WordNet Approach" oleh Laras Gupitasari [12]. Perbedaannya dengan artikel sebelumnya adalah artikel ini mengembangkan proses baru yaitu Pengelompokan Sinonim Set yang bertujuan mengelompokkan kata-kata dari sekelompok sinonim yang dihasilkan berdasarkan makna yang sama dan setelahnya hasil sinonim akan dikelompokkan dalam bahasa Arab saja dan akan menampilkan pengelompokan bahasa Arab sesuai dengan jarak kedekatan makna dengan kata-kata yang dimasukkan, artikel ini menggunakan input bahasa inggris dan output bahasa arab sedangkan artikel sebelumnya menggunakan input bahasa arab dan output bahasa inggris arab serta artikel ini meningkatkan hasil clustering dari artikel sebelumnya maksudnya ialah hasil clustering yang dihasilkan oleh artikel sebelumnya belum terlalu menghasilkan banyak kata yang terkait.

Topik dan Batasannya

Dari latar belakang yang sudah dipaparkan, topik serta batasan yang diangkat dalam penelitian ini sebagai berikut :

1. Kosa Kata Al-Qur'an

Batasan topik yang diangkat berupa kosa kata dari Al-Qur'an, namun tidak semua kata Al-Qur'an menjadi topik dalam penelitian. kelas kata yang diangkat menjadi topik yaitu kelas kata terbuka yang hanya memfilter output berupa kata benda.

2. Input dan Output Sinonim Set

Inputan dari sistem berupa kata yang ingin diprediksi makna yang sama. Output dari sistem berupa menampilkan seluruh kosa kata yang bersinonim dari kata yang diinputkan, yang nantinya akan dikelompokkan sesuai dengan kesamaan makna. Contoh input dan output sebagai berikut :

Tabel 1. Output Sistem

Input Kata	Terjemahan	Output Sinonim Set
Day	Day Date Tomorrow	نهر، يوم، سبع، عشر، كبر، ذا، فصل، أذف نقر، قنو، فتل، رطب، عرجن غدو
Name	Name Brand	سمو، أله، حسن قبس، وسم

3. Rencana pengujian Sistem Proses pengujian menggunakan metode Hierarchical clustering untuk membuat himpunan sinonim set dari bahasa inggris dan bahasa arab. Adapula WordNet yang digunakan untuk mencari makna sama antar kata dari bahasa inggris.

Tujuan

Tujuan dari permasalahan yang ingin dicapai dalam penelitian ini sebagai berikut :

1. Pengelompokan data sinonim set berdasarkan jarak kedekatan makna yang sama antar lema dari kata yang diinputkan
2. Pembangunan sistem sinonim set dengan Bahasa Inggris dan Bahasa Arab