

Pembangunan *Synonym Set* untuk WordNet Bahasa Indonesia dengan Menggunakan Metode Komutatif

Dina Juni Restina ^{#1}, Moch. Arif Bijaksana ^{*2}, Kemas Muslim Lhaksmana ^{#3}

[#] *Department of Informatics, Telkom University
Bandung, West Java, Indonesia*

¹ dinajr@student.telkomuniversity.ac.id

² arifbijaksana@telkomuniversity.ac.id

³ kemasmuslim@telkomuniversity.ac.id

Abstract

In NLP (Natural Language Processing) many lexical semantic linkages can be applied, for example is the WordNet application. The WordNet has been built using a commutative method in Developing synset (synonym set). The synonym set has been built has a commutative attribute which means that if a word1 has a synonym of word2, then word2 must also have a synonym of word1, it can be said that the synonym set is valid, such as the implementation of Princeton WordNet. WordNet was first created in the English language, before the researchers translated into various languages such as Japanese, Arabic, Turkish and Indonesian and other languages. For this reason, it takes the development of WordNet to participate in helping other researchers, so in the future The Indonesian WordNet would have more complete vocabulary. WordNet has been built focusing on synset extraction, which is the first stage of WordNet development before the stages of relations between synset and gloss words. The development of synonym set uses the Indonesian Thesaurus dictionary as the source of words. The value of F-measure from the development of synonym set using the commutative method is 66 percent which is the value of the evaluation of the suitability of the synset produced by the system and the synset made by linguists based on the specified gold standard.

Keywords: Commutative method, Synset dan WordNet Bahasa Indonesia.

Abstrak

Dalam NLP (*Natural Language Processing*) banyak keterkaitan semantik leksikal yang dapat diaplikasikan, salah satunya adalah aplikasi WordNet. WordNet yang dibangun menggunakan metode komutatif dalam proses pembangunan *synonym set (synset)*-nya. *Synset* yang akan dibangun harus memiliki sifat komutatif baru dapat dikatakan *synset* yang bernilai valid, yang berarti jika sebuah kata w_1 memiliki sinonim w_2 , maka w_2 juga harus memiliki sinonim w_1 , seperti yang terjadi pada *Princeton WordNet*. WordNet pertama kali dibuat dalam bahasa Inggris, sebelum para peneliti menerjemahkan kedalam berbagai bahasa seperti bahasa Jepang, bahasa Arab, bahasa Turki serta bahasa Indonesia dan bahasa lainnya. Untuk itu dibutuhkan pembangunan WordNet untuk turut serta membantu peneliti lain agar kedepannya WordNet Bahasa Indonesia yang ada memiliki kosa kata yang lebih lengkap. WordNet yang akan dibangun akan berfokus pada ekstraksi *synset* yaitu tahapan pertama pembangunan WordNet sebelum tahapan relasi antar *synset* dan *gloss* kata. Pembangunan *synset* ini menggunakan Tesaurus Bahasa Indonesia sebagai sumber kata. Nilai *F-measure* dari pembangunan *synset* dengan menggunakan metode komutatif adalah sebesar 66 persen yang merupakan nilai evaluasi kesesuaian *synset* yang dihasilkan oleh sistem dan *synset* yang dibuat oleh ahli bahasa berdasarkan *Gold Standard* yang ditentukan.

Kata Kunci: Metode komutatif, *Synset* dan WordNet Bahasa Indonesia.