

Building Synonym Sets for English WordNet with Robust Clustering Using Links Method

Sarah Suryaningsih¹, Moch Arif Bijaksana², Widi Astuti³.

^{1,2,3} Program Studi Teknik Informatika, Universitas Telkom
email: sarahsy@student.telkomuniversity.ac.id¹,
email: arifbijaksana@telkomuniversity.ac.id²,
email: widiwdu@telkomuniversity.ac.id³

Abstrak

WordNet Bahasa Inggris adalah sinonim penting yang ditetapkan untuk menyajikan kesamaan makna antara kata-kata. Kumpulan Sinonim dibangun menggunakan Oxford Thesaurus yang diakses melalui lexico.com, yang merupakan bagian dari basis data leksikal yang akan digunakan. Setelah menggunakan proses ekstraksi melalui Oxford Thesaurus itu akan menghasilkan set sinonim dengan makna yang sama antara kata-kata. Perbedaan antara WordNet dan kamus biasa adalah bahwa kata tersebut saling berhubungan dengan kata-kata lain. Salah satu metode yang digunakan untuk pendekatan ini adalah metode Robust Clustering Using Links, yang merupakan nilai kesamaan dan set sinonim yang telah dibuat untuk digunakan untuk membangun database leksikal. Oleh karena itu tujuan utama pengembangan WordNet bahasa Inggris adalah untuk menghasilkan set sinonim yang akurat menggunakan teknik clustering. Perhitungan evaluasi akan menggunakan metode *F-Measure* dan akan menggunakan *gold standard* untuk metode perhitungan. Dengan metode ROCK, ada peningkatan akurasi *output* dari *input dataset*. Membangun wordnet Inggris adalah untuk meningkatkan kata-kata yang dapat digunakan untuk membantu penelitian dan pengembangan wordnets bahasa lain dengan model peran menggunakan wordnets bahasa Inggris yang lebih akurat. Dan penggunaan metode ROCK ada peningkatan akurasi pada hasil pengembangan wordnet bahasa Inggris dibandingkan dengan metode sebelumnya, yaitu menggunakan *hierarchical clustering*. Hasil dari penelitian ini menghasilkan peningkatan akurasi sehingga metode ROCK adalah salah satu metode yang baik digunakan dalam pengembangan wordnet bahasa Inggris.

Kata Kunci: WordNet, Robust Clustering Using Links, F-measure, Gold Standard.

Abstract

English WordNet is an important synonym set to present the similarity of meanings between words. Synonym Set is built using Oxford Thesaurus which is accessed through lexico.com, which is a part of the lexical database that will be used. After using the extraction process through Oxford Thesaurus it will produce a synonym set with the same meaning between words. The difference between WordNet and ordinary dictionaries is that the word is interconnected with other words. One method employed for this approach is Robust Clustering Using Links method, which is similarity values and synonym sets that have been created to be used to build a lexical database. Therefore the main purpose of the development of the English WordNet is to produce an accurate synonym set using clustering techniques. The evaluation calculation will use the F-measure method and will use the gold standard for the calculation method. With the ROCK method, there is an increase in accuracy output from dataset input. Building the English wordnet is to improve words that can be used to help research and development of other language wordnets with role models using more accurate English wordnets. And the use of ROCK method there is an increase in the accuracy upon results of the development of English wordnet compared to the previous method, which is using hierarchical clustering. The outcome of this study resulted in improved accuracy so that the ROCK method is one of the good methods used in the development of the English wordnet.

Keywords: WordNet, Robust Clustering Using Links, F-measure, Gold Standard.