

---

## **Pengembangan Versi Awal BahasaNet: Sebuah WordNet untuk Bahasa Indonesia**

**Staff** : Ruli Manurung, Mirna Adriani, Desmond Darma Putra,  
Abdul Arfan, Eliza Margaretha and Franky  
**Students** : -  
**Sponsors** : RUUI 2007  
**Email contact** : maruli@cs.ui.ac.id, mirna@cs.ui.ac.id

A WordNet is a useful lexical resource where specific senses of words are clustered together into synonym sets, and semantic relationships between these sets are specified. The original Princeton WordNet is widely used in various research fields, ranging from information retrieval, computational linguistics, to artificial intelligence. Currently, there are WordNets for over 40 languages. The research presented is a project to create an Indonesian WordNet using a slightly modified version of the expand model approach, i.e. by mapping existing WordNet entries to Indonesian word sense definitions. More specifically, Princeton WordNet synsets are mapped to their analogue sense definitions in the Kamus Besar Bahasa Indonesia (KBBI) widely acknowledged to be the definitive Indonesian lexicon.

The mapping process is carried out in two ways: firstly, a manual annotation of the Global WordNet Common Base Concepts to KBBI senses is facilitated by a web-based application. Secondly, experiments are conducted in computing the mapping automatically using Latent Semantic Analysis (LSA). A semantic model is derived from a small English-Indonesian parallel corpus. Given a particular word or concept, the semantic model is then used to identify its neighbours in a high-dimensional semantic space. Results from various experiments indicate that, whilst far from successful due to the small training data collection, LSA is able to discern some level implicit semantic knowledge from the training data. In both manual and automatic mapping instances, measurements of inter-annotator reliability are shown using the Fleiss Kappa statistic.

Both the web-based annotation tool and the resulting Indonesian WordNet are available online at <http://bahasa.cs.ui.ac.id/wordnet> and <http://bahasa.cs.ui.ac.id/iwn> respectively.