ANDRI SETYAWAN. Automatic Subject Classification Based on DDC System for Library Document. Supervised by FIRMAN ARDIANSYAH.

Subject classification for library document with Dewey Decimal Classification (DDC) system is difficult to perform manually. The goal of this research is to make an application that automatically do perform subject classification for library document using similarity method. We use Natural Language ToolKit (NLTK) with Wordnet module to find similarity between keyword and DDC class. DDC is a hierarchy classification. We use Lin Similarity to find similarity between two words, with Brown corpus for Information Content (IC) of Wordnet. Wordnet can find similarity for only noun and verb, so we do not process other kinds of word. We use 30 documents combination of theses and dissertations in Bogor Agriculture University. We use 3 different methods to decide the relevant class of DDC which is similar to a document keyword. The first method is maximum-maximum method, the second one is maximum-average method, and the third method is maximum-minimum method. The first method results in 6 documents having the same main class, 2 documents having the same division class, and 0 document having the same section class. The second method results in 5 documents having the same main class, 1 document having the same division class, and 0 document having the same section class. The third method results in 3 documents having the same main class, 2 documents having the same division class, and 0 document having the same section class.

Keyword: Dewey Decimal Classification, Automatic Subject Classification, Lin Similarity, Natural Language ToolKit