# *Abstract*

*Currently the data in the form of text very often found. The human language text data is unstructured or semi-structured so that the required data conversion process unstructured text into a structured text representation. In the conversion process, the weighting term is one part that is usually done.*

*There has been no comprehensive comparative study of term weighting for-corpus luteum in Indonesian language. Though Indonesian is the national language of Indonesia, and is one of the languages used by many people.*

*Therefore, in this final project conducted a comprehensive comparative study of the methods for the corpus-term weighting Indonesian corpus. The study is limited to the application for classification, data mining an important job. Doing research the effect of the weighting term elections affect the effectiveness of the classification. There are two sides why we declare as a comprehensive study. First, the method that will be examined is the whole method that we have encountered, namely TF, IDF, ITF, TF-IDF, RF, TF-RF, OR, TF-OR, IG, TF-IG, NGL, GSS, QF, IQF, ICF , VRF, QF-IQF-ICF, CHI-SQUARE, TF-X2, BINARY, and MI. Second, the Indonesian corpus will be used vary. Of the corpus that use formal writing such as news articles, non-formal such as email, which is relatively long as well as short news articles such as tweets.*

*In this study indicates that the overall method of NGL has a performance with an f-measure better than methods other for a corpus of Indonesian, especially for the classification process, and note that the value performance obtained from any weighting method relies on data and classifier used.*

***Keywords:*** *text preprocessing, term, term weighting.*