Abstract

Now, people can learn science with no effort. However, the science that exists in the world is too many, and to read the whole paper of science requires a amount of time. Because of that, we required a system that can provide the important information. This problem can be solve with classification text. Process that happen in classification text are preprocessing, extraction feature, fiture selection, an classification. Preprocessing is one of process that can affect classification text process, where in thish process, that happen transformation from raw data to data that can be classification. Preprocessing usually do by manual. In this final task, writer did comparison between preprocessing with manual system and preprocessing wih automation system. The data in this comparison are 30 scientific papers with pdf form and found on internet. Automation preprocessing done by PDF conversation. And then, split the sentences, an last cititation detection. This research done with 2 groups of features in which the first feature group only used features obtained by applying RegEx and getting more than 97%, while the second feature group applied ReEx plus N-gram and yielded a 93% percentage. The results of this study resulted in an average percentage of system accuracy over 95%.

keyword: classification text, cititation sentence, preprocessing.