Abstract

Citizen's happiness can be a parameter for determine the attainment of a Nation. Value of the attainment can be used as a parameter for evaluating and developing public policy. In the smaller scope, this thing can be applied to a city. In line with these objectives, government of Bandung city develop a system that can measure their citizen's happiness index. In measuring citizen's happiness index, would need data from their citizen itself. Government of Bandung city has done happiness index measures performed by Badan Pusat Statistik (BPS), which captured data using a random sampling survey of heads of households or spouses to represent 151 villages in Bandung. Data retrieval in the survey requires great effort and has expensive cost. Efforts to do other than to use survey data collection, namely by taking the data in the form of public opinion from social media.

In this research, created a system that can process data in the form of opinions coming from social media Twitter. Twitter selected because Twitter is one of social media that is very much used by the people of Bandung in expressing their opinions. Data coming from tweets on Twitter are classified in accordance with the needs of the Bandung citizen's happiness index measure. Ontology that contains knowledge as well as the terms that are designed to filter out tweets that are relevant to the citizen's happiness index are used for classifying tweets. Ontology development is done by the bottom-up paradigm to address the mismatch between the appearance of the term in a tweet with the term of ontology. The classified tweets by ontology were analysed using Weka classifier to get the polarity of the tweet. Every tweets that already has a class based on ontology and sentiment analysis used in measuring Bandung citizen's happiness index. Evaluation to measure the results of the system using F1 measure values and accuracy.

In the evaluation of the system found that the completeness of the ontology affect the performance of the classification results analysed from the value of accuracy and F1 measure value on two different verions of ontology completeness. In addition, with these characteristics of tweet data, the amount of training data affect the performance of the Weka classifier in classifying sentiment of tweets. The greater the amount of data the grerater value of accuracy and F1 measure value in the classifying sentiment of tweets. From the result of happiness index measures performed by system, obtained value of happiness index about 99,9%. The result is not separated from the performance of the classifier in classifying sentiment of tweets within 55,3% accuracy and 40% F1 measure value.

Keywords: happiness index, retrieve data, ontology, twitter, sentiment analysis