SENTIMENTAL ANALYSIS

A multimedia data processing system and algorithms are proposed to analyze the sentiments of users from large amounts of unstructured text data generated by SNSs. The proposed method is composed of a parallel HDFS system based on the Hadoop ecosystem and on four MapReduce functions. In addition, it uses five types of data dictionary for sentiment analysis. The proposed method stably processes data loading according to the increase in the number of data items. The system load is distributed to each node by parallel processing. When the proposed sentiment analysis functions have processed the data effectively, the system load is not concentrated on a single node but is evenly distributed among all nodes. The results of sentiment analysis with the proposed system are very close to the results of sentiment analysis with a manual process

PROCEDURES:
Data Gathering:
The data collection method of the proposed system was processed through Twitter and Topsy for performance analysis. Topsy analyzes the activity of users in SNS services such as Google Plus and Twitter. Topsy provides data by analyzing about 500 million pieces of data per day. After the acquisition of the historical data, Twitter4j has been used to collect data for continuous incremental data.

Extraction and Processing of Necessary Data:
The data collected from Twitter and Topsy contain a lot of unnecessary data. Therefore, only the necessary data needs to be extracted from the collected data. Table 4 shows an example of the type of data to be extracted from the raw data. And then, in the next step, the extracted data is processed to match the format of the extracted data from Twitter and Topsy

Data Loading and Sentiment Analysis:
The extracted data are stored in the proposed HDFS in parallel. Then, the stored data undergo sentiment analysis via the MapReduce functions. MapReduce functions consist of Reduce and Map. Figure 4 shows the general functionality of Map and Reduce. After the sentences to be analyzed are separated into data with their value and key, they are sorted based on the key value, then the sentences are sorted by their value. Finally the data are reduced on the key, and the final results are obtained