

Getting and Cleaning Data Course Project - Codebook

This codebook describes the variables, data and transformations performed to clean up the data.

Content

These initial datasets are the basis of the transformation of the data:

- X_train.txt - Training set
- y_train.txt - Training labels
- subject_train.txt - Subject performing the experiment (1-30), training set
- x_test.txt - Test set
- y_test.txt - Test labels
- subject_test.txt - Subject performing the experiment (1-30), test set
- activity_labels.txt - Labels for the activities carried out (1 WALKING, 2 WALKING_UPSTAIRS, 3 WALKING_DOWNSTAIRS, 4 SITTING, 5 STANDING, 6 LAYING)
- features.txt - List of features

Process

First, all data is read. As a second step, the columns of the datasets are renamed, for better clarity but also to have common identifiers, such as "subjectID" or "activityID". The labeling of the x_train and x_test comes out of the features.txt. The datasets from the test data (x_test, y_test, subject_test) and the training data (x_train, y_train, subject_train) are merged, resulting in allData. The next step is to extract only the variables containing mean and standard deviation for each of the measurements. After this, the outcoming data is merged with the activity_labels in order to have the activities labeled properly, resulting in allDataAct. Finally, a new dataset is created with the average of each variable for each activity and each subject (meanData.txt).