

Indian Institute of Technology Kharagpur

Open Soft Problem Statement

February 23, 2016

Plots to Tables

The input is a set of (multiple) scanned pdf pages, each containing (one or more) figures which are two-dimensional plots of experimental results. A scanned page can have a single or multiple plots which may or may not be embedded in text. Each plot has the x and y axis with their labels and unit measurements marked in the plot (linear scale). Inside each figure are one or more plots, each with a different colour depicting a certain plotted entity (E_i) and their labels given separately within the plot as a caption. (The example enclosed shows such an input set.)

You are required to read a set of scanned pages as input where each page has one or more plots embedded in text and convert them to a set of two-dimensional tables, one table per plot, where each row of the table has the following values - the x-axis value, y-axis value and values for each E_i or a dash (-) in case there is no value for that E_i . Each Table should have the first row as the name labels for the x-axis, y-axis and various E_i values. It should cover x and y axis values from the minimum to maximum range with one tenth of the minimum marked unit in the plot as granularity. The Table as a whole will have a caption as per the caption of the figure.

The output will be a set of pdf pages which contain the name of the participant as the first page followed by a sequence of results having the first input page followed by the set of tables corresponding to the figures in that page (one table per page) followed by the next input page and the tables of that page, etc.

An example set of input pages is enclosed.