PharmaSUG China 2019 - Paper DV-079

Extract X and Y coordinates from Image by GROOVY Procedure

KuenHung Lin, Pharmaceutical Co., Ltd, Taipei, Taiwan

ABSTRACT

When creating graphics, the basic material is point. As far as we know, two points are able to create a line and three lines form a plane. Hence, if we could control X and Y coordinates of points, we could draw more complex figures. In order to accomplish this purpose, we would print out the image and manually calculate the coordinates with grid. In the meantime, we also could try several mathematical functions to achieve the goal. Regardless of manual or mathematical calculation would take so much time. This paper provides automation approach that using PROC GROOVY to easily extract X and Y coordinates of points from image to draw more vivid figures efficiently.

INTRODUCTION

Nowadays, SAS/GRAPH statistical graphics procedures (SG procedures) use the Graph Template Language (GTL) to create the most common graphs. However, requests of graphs are ever-changing in pharmaceutical and biotechnology industries. Sometimes, we must face some graphs can't be directly created by the SG procedures. Fortunately, the SG procedures provide a mechanism for adding shapes, images, and other annotations approach to graph outputs. X and Y coordinates play an important role when we implement annotation task.

Whether we manually guess or try several mathematical functions to estimate coordinates of a point, we would take so much time. Under the circumstances, programmers always need to attempt to figure difficulties out and use appropriate SAS procedures to efficiently obtain the coordinates.

The rest of this paper is organized as follows. First, we brief what is PROC GROOVY and what it can do. Second, we present the automation approach which will dramatically reduce your time to obtain the coordinates by PROC GROOVY. Besides, we apply an image to illustrate the proposed method. Finally, some concluding remarks are given at the end of this paper.

OVERVIEW PROC GROOVY

Groovy is a dynamic language which runs on the Java Virtual Machine (JVM), the syntax is Java-like, and most of the Java code can also work in Groovy. PROC GROOVY enables to execute Groovy code with SAS code on the JVM. To deal with image in Portable Document Format (PDF) document, we would like to apply the Apache PDFBox® library to demonstrate an example in this paper. It is an open source Java PDF library and it allows to create new PDF documents, manipulation of existing documents and the ability to extract content from documents. In other words, we can combine SAS and Groovy code for Apache PDFBox® library to flexibly develop more automation approach.

AUTOMATION APPROACH FOR EXTRACTING COORDINATES FROM IMAGE

The method which we present in this paper can be divided into two steps. The first step is as pre-process, we prepare required documents for Apache PDFBox® library. The second step is as automation approach, we would like to invoke the information from the first step to draw graph and present more details in following sections.

STEP1: PRE-PROCESS

At the beginning, you have to insert blank table into Microsoft Word. The numbers of rows and columns which are similar with dots per inch (DPI) will affect resolution of our graph. It depends on demand for painter. And then, we can insert image into the file. According to the table as grid, we can easily add text into the cell to mark point which we would like to draw in final graph. Furthermore, we also can distinguish multiple areas by different text. Figure 1 shows the demonstration of the process. For the next step, we save the marked file as PDF document.

Figure 1. Demonstration of pre-process

STEP2: AUTOMATION APPROACH

Using PROC GROOVY to extract coordinates of texts which we marked in the first step as Figure 2.

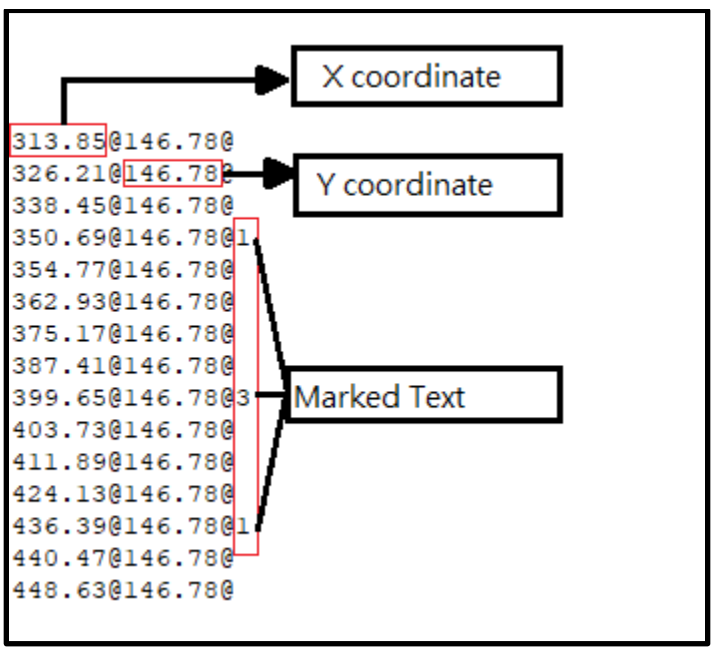

Figure 2. Sample of extracting information

After that, managing above file to a SAS data set to apply the SG procedures to draw the graph. Figure 3 shows the example data set.

	PGM	X	Y	Z	X_BODY	Y_BODY	X_ORANGE	Y_ORANGE
16	350.69@136.45999@1	350.69	136.45999	1	350.69	136.45999		
17	436.39@136.45999@1	436.39	136.45999	1	436.39	136.45999		
18	350.69@146.78@1	350.69	146.78	1	350.69	146.78		
19	399.65@146.78@3	399.65	146.78	3			399.65	151.78
20	436.39@146.78@1	436.39	146.78	1	436.39	146.78		
21	350.69@156.98001@1	350.69	156.98001	1	350.69	156.98001		
22	436.39@156.98001@1	436.39	156.98001	1	436.39	156.98001		
23	362.93@167.29999@1	362.93	167.29999	1	362.93	167.29999		
24	424.13@167.29999@1	424.13	167.29999	1	424.13	167.29999		
25	350.69@177.5@1	350.69	177.5	1	350.69	177.5		
26	362.93@177.5@1	362.93	177.5	1	362.93	177.5		

Figure 3. Sample of SAS data set for drawing

THE RESULT OF PROPOSED METHOD

In this section, Figure 4 presents the result of proposed method based on the image from Figure 1. The graph is created by SGPLOT procedure.

Figure 4. Result of proposed method

CONCLUSION

The motivation of the paper is to create animation by SAS. As far as we know, the animation is composed of many pictures. In order to efficiently draw outputs, the approach presented in this article was designed to create figure based on existing image in PDF document. Thus, we can reduce time to draw animated figure. However, the resolution of picture is a challenge to the approach. If we want to obtain high DPI picture, we need more cells to mark the text of image. For the purpose, we must find more appropriate applications instead of Microsoft Word. Although, we can choose other more suitable software for creating animation. The main idea of the article is finding other way to break the limitation of the SG procedures.

REFERENCES

Yan, Yin-Jhen and Chang, Kyle. 2018. “No Solution to Auto-generating acrf.pdf? Try to Use GROOVY Procedure!” *Proceedings of the PharmaSUG China 2018*, paper AD-71. Available at http://www.pharmasug.org/proceedings/china2018/AD/Pharmasug-China-2018-AD71.pdf.

Apache PDFBox®. Available at https://pdfbox.apache.org/.

ACKNOWLEDGMENTS

The author would like to thank my friend Winkle Lu for valuable comments that help improve the manuscript, and all participant in competition of drawing for fun that encourage me to find more efficient way for creating animation via SAS.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: KuenHung Lin
E-mail: liinhater@gmail.com