Abstract

There are millions of applications uploaded by the

developers on the daily basis. Without any check and balance

millions of users download these applications. Theses duplicated

applications damage the users trust on Google play store and can grab the confidential information of user. There is no more

information provided by developers on the front end of the

application that can define the legitimacy of the application. In

this project, by using a scrapy framework build a Google play

store dataset with all categories of apps. Scraping at least 1.1 M+ applications of each category of in free and respectively in

paid applications by using scrapy web crawling, cumulatively

scrape the 1.1 M+ paid applications and free applications of all

categories. The categories of these applications

use respectively are Games,Lifestyle,Business,Books,Travel,Food and drinks and 26+ other categories. On each

application on Google play store, scrape maximum 26 attributes,

but use four attributes for analysis in this paper that is Installs,

Advertisements support, InApplicationPurcahses and Ratings. In

this paper, visualizing the InAppPurchase rate of free and paid

applications, Percentage of the advertisement support in free and

paid applications, Ratings of free and paid application with

histogram, Installs of free and paid application with a histogram

with all categories of application. To check the

relationship in between attributes also, visualize . This visualization is more helpful for game developers

in the development phases, also for the users of the game’s

application for the selection of the game that they want to play.