UNIVERSITY RESULT LEDGER SHEET ANALYSIS

Nageshkumar Mhetre, Sandeep Posham, Vitthal Narayankar, Sadguru Katakdhond

Computer Science & Engineering

N. B. Navale Sinhgad College Of Engineering, Solapur

Abstract

The traditionally universities declares the results in the form of PDF. When colleges get those PDF's, then the professors or the teachers has to take each record manually from the PDF for each student.
Sometimes it takes 10-15 days to get the records from the PDF into an Excel. Generally there are thousands of students in a college and hundreds of students in a particular department. So it becomes difficult for the staff to analyze those data.

Introduction

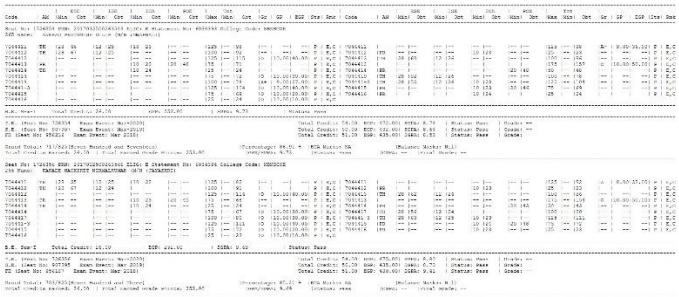
The main purpose of this project is to provide an easy way that can help staff and students to get the data from PDF file which contains result of students. It becomes very difficult to scan the data and fetch the required data from that PDF because it contains a huge data that contains results of all the students who are studying in various colleges under the university.

So we have to do manual analysis on that PDF file which is not an easy task and it is very time consuming. That's why we are going to read the PDF file and extract required data from that file and convert it into excel file, so that the analysis becomes easier and faster than the previous.

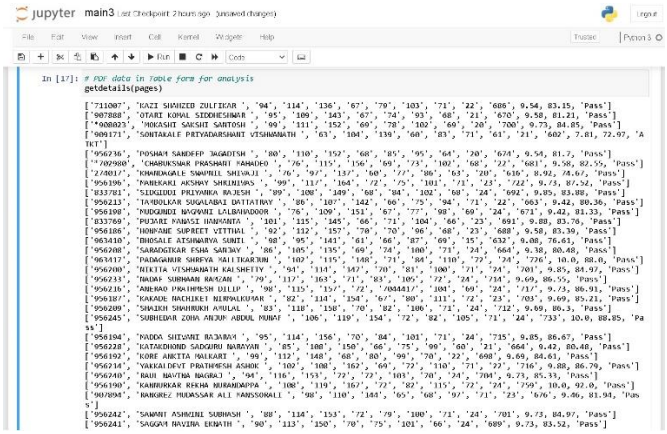
In this project we're going to take input as a PDF file, process that file and extract all the required data like student's name, seat number, college name, department name, etc. Now this extracted data, we put that all data into an text file first and then convert that text file into an Excel file. In Excel we're going show our output in the tabular form.

So that the user (staff/student) can easily work on that analyzed data in Excel file and they can compare the result of each student. So they'll know about the weakness and strength of students and they treat them accordingly.

Input PDF format:

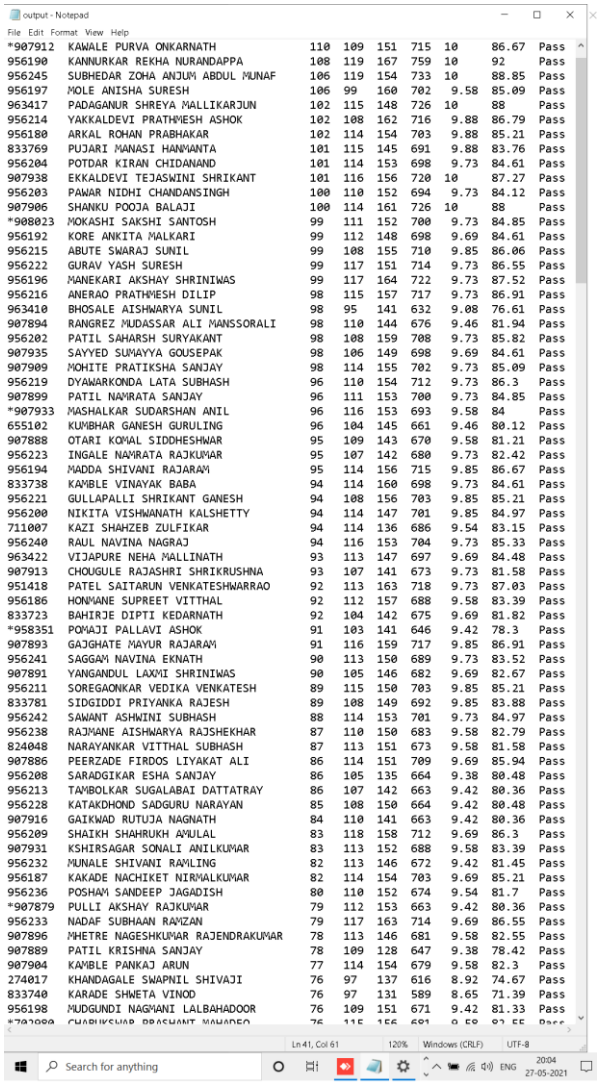


Extracted data from PDF file:



Final Output:

Data saved inside file



output - Notepad

File Edit Format View Help

*907912	KAWALE PURVA ONKARNATH	110	109	151	715	10	86.67	Pass
956190	KANNURKAR REKHA NURANDAPPA	108	119	167	759	10	92	Pass
956245	SUBHEDAR ZOHA ANJUM ABDUL MUNAF	106	119	154	733	10	88.85	Pass
956197	MOLE ANISHA SURESH	106	99	160	702	9.58	85.09	Pass
963417	PADAGANUR SHREYA MALLIKARJUN	102	115	148	726	10	88	Pass
956214	YAKKALDEVI PRATHMESH ASHOK	102	108	162	716	9.88	86.79	Pass
956180	ARKAL ROHAN PRABHAKAR	102	114	154	703	9.88	85.21	Pass
833769	PUJARI MANASI HANMANTA	101	115	145	691	9.88	83.76	Pass
956204	POTDAR KIRAN CHIDANAND	101	114	153	698	9.73	84.61	Pass
907938	EKKALDEVI TEJASWINI SHRIKANT	101	116	156	720	10	87.27	Pass
956203	PAWAR NIDHI CHANDANSINGH	100	110	152	694	9.73	84.12	Pass
907906	SHANKU POOJA BALAJI	100	114	161	726	10	88	Pass
*908023	MOKASHI SAKSHI SANTOSH	99	111	152	700	9.73	84.85	Pass
956192	KORE ANKITA MALKARI	99	112	148	698	9.69	84.61	Pass
956215	ABUTE SWARAJ SUNIL	99	108	155	710	9.85	86.06	Pass
956222	GURAV YASH SURESH	99	117	151	714	9.73	86.55	Pass
956196	MANEKARI AKSHAY SHRINIWAS	99	117	164	722	9.73	87.52	Pass
956216	ANERAO PRATHMESH DILIP	98	115	157	717	9.73	86.91	Pass
963410	BHOSALE AISHWARYA SUNIL	98	95	141	632	9.08	76.61	Pass
907894	RANGREZ MUDASSAR ALI MANSSORALI	98	110	144	676	9.46	81.94	Pass
956202	PATIL SAHARSH SURYAKANT	98	108	159	708	9.73	85.82	Pass
907935	SAYYED SUMAYYA GOUSEPAK	98	106	149	698	9.69	84.61	Pass
907909	MOHITE PRATIKSHA SANJAY	98	114	155	702	9.73	85.09	Pass
956219	DYAWARKONDA LATA SUBHASH	96	110	154	712	9.73	86.3	Pass
907899	PATIL NAMRATA SANJAY	96	111	153	700	9.73	84.85	Pass
*907933	MASHALKAR SUDARSHAN ANIL	96	116	153	693	9.58	84	Pass
655102	KUMBHAR GANESH GURULING	96	104	145	661	9.46	80.12	Pass
907888	OTARI KOMAL SIDDHESHWAR	95	109	143	670	9.58	81.21	Pass
956223	INGALE NAMRATA RAJKUMAR	95	107	142	680	9.73	82.42	Pass
956194	MADDA SHIVANI RAJARAM	95	114	156	715	9.85	86.67	Pass
833738	KAMBLE VINAYAK BABA	94	114	160	698	9.73	84.61	Pass
956221	GULLAPALLI SHRIKANT GANESH	94	108	156	703	9.85	85.21	Pass
956200	NIKITA VISHWANATH KALSHETTY	94	114	147	701	9.85	84.97	Pass
711007	KAZI SHAHZEB ZULFIKAR	94	114	136	686	9.54	83.15	Pass
956240	RAUL NAVINA NAGRAJ	94	116	153	704	9.73	85.33	Pass
963422	VIJAPURE NEHA MALLINATH	93	113	147	697	9.69	84.48	Pass
907913	CHOUGULE RAJASHRI SHRIKRUSHNA	93	107	141	673	9.73	81.58	Pass
951418	PATEL SAITARUN VENKATESHWARRAO	92	113	163	718	9.73	87.03	Pass
956186	HONMANE SUPREET VITTHAL	92	112	157	688	9.58	83.39	Pass
833723	BAHIRJE DIPTI KEDARNATH	92	104	142	675	9.69	81.82	Pass
*958351	POMAJI PALLAVI ASHOK	91	103	141	646	9.42	78.3	Pass
907893	GAJGHATE MAYUR RAJARAM	91	116	159	717	9.85	86.91	Pass
956241	SAGGAM NAVINA EKNATH	90	113	150	689	9.73	83.52	Pass
907891	YANGANDUL LAXMI SHRINIWAS	90	105	146	682	9.69	82.67	Pass
956211	SOREGAONKAR VEDIKA VENKATESH	89	115	150	703	9.85	85.21	Pass
833781	SIDGIDDI PRIYANKA RAJESH	89	108	149	692	9.85	83.88	Pass
956242	SAWANT ASHWINI SUBHASH	88	114	153	701	9.73	84.97	Pass
956238	RAJMANE AISHWARYA RAJSHEKHAR	87	110	150	683	9.58	82.79	Pass
824048	NARAYANKAR VITTHAL SUBHASH	87	113	151	673	9.58	81.58	Pass
907886	PEERZADE FIRDOS LIYAKAT ALI	86	114	151	709	9.69	85.94	Pass
956208	SARADGIKAR ESHA SANJAY	86	105	135	664	9.38	80.48	Pass
956213	TAMBOLKAR SUGALABAI DATTATRAY	86	107	142	663	9.42	80.36	Pass
956228	KATAKDHOND SADGURU NARAYAN	85	108	150	664	9.42	80.48	Pass
907916	GAIKWAD RUTUJA NAGNATH	84	110	141	663	9.42	80.36	Pass
956209	SHAIKH SHAHRUKH AMULAL	83	118	158	712	9.69	86.3	Pass
907931	KSHIRSAGAR SONALI ANILKUMAR	83	113	152	688	9.58	83.39	Pass
956232	MUNALE SHIVANI RAMLING	82	113	146	672	9.42	81.45	Pass
956187	KAKADE NACHIKET NIRMALKUMAR	82	114	154	703	9.69	85.21	Pass
956236	POSHAM SANDEEP JAGADISH	80	110	152	674	9.54	81.7	Pass
*907879	PULLI AKSHAY RAJKUMAR	79	112	153	663	9.42	80.36	Pass
956233	NADAF SUBHAAN RAMZAN	79	117	163	714	9.69	86.55	Pass
907896	MHETRE NAGESHKUMAR RAJENDRAKUMAR	78	113	146	681	9.58	82.55	Pass
907889	PATIL KRISHNA SANJAY	78	109	128	647	9.38	78.42	Pass
907904	KAMBLE PANKAJ ARUN	77	114	154	679	9.58	82.3	Pass
274017	KHANDAGALE SWAPNIL SHIVAJI	76	97	137	616	8.92	74.67	Pass
833740	KARADE SHWETA VINOD	76	97	131	589	8.65	71.39	Pass
956198	MUDGUNDI NAGMANI LALBAHADOOR	76	109	151	671	9.42	81.33	Pass
*702080	CHABUKSWAR PRASHANT MAHADEO	76	115	156	691	9.58	82.55	Pass

Ln 41, Col 61 120% Windows (CRLF) UTF-8

Search for anything ENG 20:04 27-05-2021

Literature Survey

PDF is a versatile document format, but it can be difficult to read and edit it. To easily extract a table or edit text in a spreadsheet format, you can to convert PDF to editable Excel spreadsheets. There few applications or online PDF to Excel convertors, but these applications convert the whole PDF in Excel. Many of the times PDF has some irrelevant data that also get converted to Excel.

To perform analysis the data must be in appropriate format with required relevant data.

An Excel/CSV file store the data in tabular form that helps analyst to analyze data without any difficulty on other hand these things cannot be performed on PDF. Traditionally data analysis performed on data in Excel/CSV, but there is not a reliable way available to get PDF data into Excel.

Implementation

University declares the result which is in the format of ledger sheet. If staff wants to analyze the result then it'll be difficult because the result is in the pdf format.

So in our project we are making the process of analysis easy.

But we need data in tubular form like Excel/CSV to do analysis, so we're converting that pdf file into Excel/CSV file so the user can analyze result of each student.

We're doing this project using python language and using ide as anaconda jupyter. Python has libraries that helps to convert data from one file

format to another. We're currently using PyPDF2 and Numpy libraries for our project.

PyPDF2 is used for reading the data from pdf file and Numpy is used for working with arrays and lists that we are using for analysis.

At the beginning we take input file which is in the pdf format and we read all the data included in that file using PyPDF2 library.

Then we extract the only data which we required and store it in another file, after that we convert the data in the file in tables and there we do our final step as analysis.

In that we analyze student data such as marks gained by student in each Subject, Students Percentage/SGPA, Overall Marks, etc.

Future Scope

Now our project supports only the specific format of input PDF file.

In future many applications like this can be build which are robust enough to support multiple input formats.

This project is a standalone application till now but we can make it a web application by adding some extra features, So that any of the user can access this product remotely on browser/web.

CONCLUSION:

After successful completion of this project we reached to the following conclusion.

- ✓ While working on this project, we worked with some unstructured data & tried to format the result in a structured tabular manner.
- ✓ We worked with many python libraries like PyPDF2, Numpy, tkinter, pandas, & tabula to achieve our goal and to solve the one of the problem in our institute and simplified the task of user for analysis purpose.

References

- https://pypi.org/
- https://www.anaconda.com

Guidance:

Prof. H. T. Gurme
(CSE)