# SMART WAY OF FILE CONVERSION USING PYTHON

## Ms. Snehal S. Somawanshi[1], Shrikant R. Kanawade[2], Rupesh A. Autade[3], Harshal V. Javare[4]

*[1,2,3,4]Sir Visvesvaraya Institute of Technology, Nashik*

-----------------------------------------------------------------------***---------------------------------------------------------------------

**ABSTRACT**- Now a days we get multiple documents which are necessary in our day to day life, but converting them is from one to format to another is very difficult at this day, you can't get appropriate result as you want. There are many online tool available for this kind of work, they convert pdf to word, word to pdf, ppt to pdf and many more format. Problem with this is that when you convert pdf to word by using this tools we see changes of format, line spacing, table, etc. In our system we are making conversions more accurate than existing tools. For solving the drawback of online tools, we are going to make a multiple file converter which will convert all type of files without making changes to the existing data which is present in the.

Keywords: PDF to Word Converter, document data prevention, multiple files conversions, line Spacing

## I. INTRODUCTION

In most of the digital libraries and websites the artefacts are available in Portable Document Format (PDF). PDF allows users to exchange documents in an easy and reliable way across platforms. PDF is good for distributing documents that need to be printed. Most of the users read online these PDF documents in their daily activities and face problems while reading multicolumn PDF documents. Portable Document Format (PDF) is a distribution format that retains a document's true for-matting across platforms. It was developed by Adobe Corporation to allow users to exchange and view their electronic documents in an easy and reliable way, independent of the environment in which these documents were created. Originally graphic artists, designers and publishers for producing colour page proofs, mostly used PDF. With its evolving technology, however, today PDF is used for virtually any data that needs to be exchanged among applications and users. PDF is emerging as a powerful format for reproducing documents over the web. A PDF document consists of a collection objects that together describe the appearance of one or more pages, possibly accompanied by additional interactive elements and higher-level application data. A PDF contains objects making up a PDF document along with associated structural information all represented as a single self-contained sequence of bytes. PDF files provides online view of the printed page and printed exactly as the author intended using an acrobat PDF file reader. Being a high-level, interpreted language with a relatively easy syntax, Python is perfect even for those who donate have prior programming experience. Popular Python libraries are well intel-grated and provide the solution to handle unstructured data sources like Pdf and could be used to make it more sensible and useful. PDF is one of the most important and widely used digital media. Used to present and exchange documents. PDFs contain useful information, links and buttons, form ends, audio, video, and business logic.1

## II. LITERATURE REVIEW

1. 2015 Imran Bole, Prashant Kokane, Piraji Hiding new words in a PDF document through software Photographer we generate two font files with the same character codes. The secrete key is stored in the _restyle and it contains the font shapes formed from the new words.

2. 2012 Aanchal Rawat, Srishtie Dixit The research and application of the creation PDF document based on the iTextSharp PDF (Portable Document For-mat) has undoubtedly become the new standard of electronic text industry for its high basis, full-featured and advantages to all other popular text format.
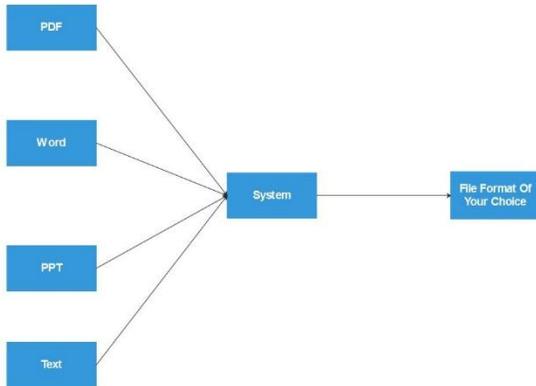
3. 2017 V. Sagar Reddy, Dr. L. Padma Sree, V. Naveen Kumar PDF document watermarking algorithm based on discarded page object in this paper presents a PDF document watermarking algorithm based on PDF document structure, where a watermarking signal is hided in the discarded page object.

4.2013 Taren Harsha Algorithm for com-paring two divergent Printouts of the same PDF Document Portable document format (PDF) is rapidly gaining acceptance in the global marketplace as the standard _le format for electronic Document distribution and print-ing. Here is an algorithm which is designed for a tool for BMP _le comparison.

## III. BLOCK DIAGRAM

We are making a web based application or android application, in which will a user login and registration will be done, we are using python as programming language which will help us to get more accurate result compare to other online tool available in market, In our system we are provide multiple options of _le to be converted, which will help user to convert multiple files. User registration is

done restyle, after registration user will get unique id and password. So user has enter his/her id and password to login to the system.



## IV. PyCharm Software

PyCharm is an integrated development environment (IDE) used in computer programming specially for the Python language. It is developed by the Czech company JetBrains.[6] Itprovides code analysis, a graphical debugger, an integrated unit tester,integration with version control systems (VCSes), and supports web development with Django as well as Data Science with Anaconda.[7] PyCharm is cross-platform, with Windows, macOS and Linux versions. The Community Edition is released under the Apache License,[8] and there is also Professional Edition with extra features ^a released under a proprietary license.

## V. PyCharm Feature:

Coding assistance and analysis, with code completion, syntax and error highlighting, linter integration, and quick xes

Project and code navigation: specialized project views, le structure views and quick jumping between les, classes, methods and usages

Python refactoring: includes rename, extract method, introduce variable, introduce con-stant, pull up, push down and others

Support for web frameworks: Django, web2py and Flask Integrated Python debugger

Integrated unit testing, with line-by-line code coverage

Google App Engine Python development

Version control integration: unied user interface for Mercurial, Git, Subversion, Perforce and CVS with change lists and merge

## VI. Xampp software

XAMPP is a free and open-source cross-platform web server solution stack package developed by Apache Friends,[2] consisting mainly of the Apache HTTP Server, Maria DB database, and interpreters for scripts written in the PHP and Perl programming languages.[3][4] Since most actual web server deployments use the same components as XAMPP, it makes transitioning from a local test server to a live server possible. XAMPP's ease of deployment means a WAMP or LAMP stack can be installed quickly and simply on an operating system by a developer. With the advantage a number of common add-in applications such as WordPress and Joomla! Can also be installed with similar ease using Bitnami.

## VII. Advantages

- The main objective is to make PDF content more usable and readable. Facilitation and satisfaction of users of PDF with enhanced usability.
- It will give its users more control over document^as formatting for reading/viewing pur-poses.:
- Easy navigation and searching of the data when made available in WORD format. Color of visited links will be changed, and fast access to the required content.
- Achieving standard web formats in PDF les, providing standardized view of PDF content in web pages.
- Reducing the time and e ort required to read online the PDF documents

## VIII. CONCLUSION

PDF is the preferred format for publishing formatted documents. PDF les are becoming more and more widespread on web. It is good for printing document. It provides online view of printed page. It is not very usable for online reading and presentation. It is not a standard web page format. It causes severe usability problems on web [Nielsen 2003, Nielsen 2001, Sajka 2002]. To overcome these problems, PDF to word Converter is developed for online reading and presenting contents in word. It is developed to enhance the usability of PDF documents and making PDF contents more usable for online reading of scienc literature. Connection, and GPS enabled android phone to utilize the app.

## IX. REFERENCES

[1] Ahmad, PDF to HTML Converter, April 2006.

[2] P. J. Lynch, S. Horton, Web Style Guide, [online] Available: http://www.webstyleguide.com/.

[3] J. Nielsen, "Avoid PDF for On-Screen Reading", Alert box: Current Issues in Web Usability, June 2001, [online] Available: http://www.useit.com/alertbox/.

[4] J. Nielsen, "PDF: for Human Consumption", Alert box: Current Issues in Web Usability, July 2003, [online] Available: http://www.useit.com/alertbox/.

[5] K. Had jar, M. Riga Monti, D. Lalanne, R. Ingold, "Xed: a new tool for extracting hidden structures from electronic documents", DIAL '04. First Int'l Conference on Document Image Analysis for Libraries, pp. 212-224, 2004.

[6] T. Hassan, R. Baumgartner, "Table recognition and understanding from PDF files", ICDAR 2007. Ninth Int'l Canf on Document Analysis and Recognition, pp. 1143-1147, 2007.

[7] E. Oro, M. Ruolo, "PDF-TREX: An approach for recognizing and extracting tables from pdf documents", 10th Intel Conf. on Document Analysis and Recognition, pp. 906-910, 200