

**A**  
**Mini Project**  
**on**  
**“PDFLY- AN ONLINE PDF UTILITY”**

**SUBMITTED BY:**  
**SHRADDHA SINGH(1908410100061)**

**SUBMITTED TO:**  
**DR. ANURAG SEWAK**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**  
**RAJKIYA ENGINEERING COLLEGE, SONBHADRA**

# INDEX

<b>S. No.</b>	<b>TITLE</b>	<b>Page No.</b>
<b>1.</b>	<b>Abstract</b>	<b>4</b>
<b>2.</b>	<b>Literature Survey</b>	<b>5</b>
<b>3.</b>	<b>Problem Statement</b>	<b>6</b>
<b>4.</b>	<b>Technology</b>	<b>7</b>
<b>5.</b>	<b>Methodology</b>	<b>10</b>
<b>6.</b>	<b>Scope</b>	<b>11</b>
<b>7.</b>	<b>Application</b>	<b>12</b>
<b>8.</b>	<b>References</b>	<b>12</b>

## **ACKNOWLEDGEMENT**

It is a great pleasure for me to express my gratitude to DIRECTOR GEETAM SINGH TOMAR SIR for providing such an impeccable learning environment with all resources needed. I would like to also mention and acknowledge HOD DR. AMOD KUMAR TIWARI SIR for his leadership and constant motivation provided in successful completion of my academic semester. I deeply thank him for providing the efficient faculty and facilities to make our ideas into reality.

I express my sincere thanks to our project coordinator (Ass. Prof.) Dr. ANURAG SEWAK SIR for his guidance , encouragement, appreciation and intellectual zeal which motivated me to venture this project successfully. He has been a constant guide and monitored the project over a period of time . We express our deep gratitude and affection to our parents who stood behind us in all our endeavours. Finally, I am pleased to acknowledge the indebtedness to all those who devoted themselves directly or indirectly to make this project report success.

## **ABSTRACT:**

Primarily , this project deals with pdf conversion and pdf manipulation at single click. Since pdfs are still hard to manipulate, this project tries to ease the way of achieving the same. As free pdf converters are very less in the online market , people still struggle to do manipulations like replacing, merging, etc.

This project took motivation from this hardship of pdf manipulation and implemented it using API and Node.js. This project had many challenges , first of all , finding an API which is free and easy to implement and provide efficient solutions. Second, system incompatibility, since APIs work on dependencies of their time, as dependencies get updated, incompatibility occurs and many features become inapplicable.

Since everything is going paperless and online, flexible pdf is a requirement of today. This project will assist students, working professionals and businesses in their day to day tasks like documentation . This project has great future scope if made properly under complete business prospective.

## LITERATURE SURVEY:

**Definition:** PDFs have been in the market for about 28 years, it was created by the adobe systems and remained a proprietary format until it was released as an open standard in 2008.

PDF was developed for the purpose of sharing documents consisting inline images, text formattings among computer users of different platforms who may not have access to mutually-compatible application software. A utility that provides aid to working professionals as well as students for free or at efficient costs for PDF manipulations and conversions is very much appreciated.

Thus coming up with an online PDF Utility that tackles all these problems was the main idea behind this project.

Resources that we have referenced to research while coming up with the idea were:

- Node Package Manager Official website npmjs: <https://www.npmjs.com/>
- convertAPI Official Documentation: <https://www.convertapi.com/>
- PDFTron API Official Documentation:  
<https://www.pdftron.com/documentation/nodejs/get-started/integration/linu>  
[x](#)
- Stack Overflow: <https://stackoverflow.com/>

This online PDF utility uses the express server in the backend and ejs and css for the front end. It uses the PDFTRON API in the backend and serves the above mentioned purpose efficiently. It serves the purpose of PDF manipulation and conversions and serves the files in pdf format. It has features such as page extraction, page removal, thumbnail generation, optimization and water mark. It can convert a number of file formats into PDFs such as docx, pptx, png, jpg, svg, txt, etc.

## **PROBLEM STATEMENT:**

- PDFs are pretty brittle and hard to manipulate and their manipulation and conversion are hard to implement. This is a need of each and every student and professionals who work with documentations.
- There are very few good PDF Utilities that provide cost efficient services that makes them unsuitable for community usage. Some of them provide basic services for free but may charge an amount for complex manipulations, hence they become unsuitable for students who may need them for only a single time usage.
- All types of pdf manipulation are unavailable in one place . You can not find services like removing pdf pages, thumbnail generation at one place online. We try to get everything in one place.

## TECHNOLOGY:

Technology mainly used is web technology which includes :

### EXPRESS JS:-



Express is a minimal and flexible Node.js web application framework that provides a robust set of features for web and mobile applications. Express provides a thin layer of fundamental web application features, without obscuring Node.js features. Express is a fast, assertive, essential and moderate web framework of Node. It provides a robust set of features to develop web and mobile applications.

It can be used to design single-page, multi-page and hybrid web applications.

- It allows setting up middlewares to respond to HTTP Requests.
- It defines a routing table which is used to perform different actions based on HTTP method and URL.
- It dynamically renders HTML Pages based on passing arguments to templates.

### EJS:-

EJS is a simple templating language that lets you generate HTML markup with plain JavaScript. No religiousness about how to organize things. No reinvention of iteration and control-flow. It's just plain JavaScript.



EJS or Embedded Javascript Templating is a templating engine used by Node.js. Template engine helps to create an HTML template with minimal code. Also, can inject data into an HTML template at the client side and produce the final HTML. EJS is a simple templating language which is used to generate HTML markup with plain JavaScript. It also helps to embed JavaScript to HTML pages

#### **NPM:-**

- npm is the package manager for Node.js. It was created in 2009 as an open source project to help JavaScript developers easily share packaged modules of code.
- The npm Registry is a public collection of packages of open-source code for Node.js, front-end web apps, mobile apps, robots, routers, and countless other needs of the JavaScript community.
- npm is the command line client that allows developers to install and publish those packages.

#### **CSS:-**



Cascading Style Sheets (CSS) is a stylesheet language used to describe the presentation of a document written in HTML or XML (including XML dialects such as SVG, MathML or XHTML). CSS describes how elements should be rendered on screen, on paper, in speech, or on other media. Also , Bootstrap libraries are used to design web pages.

#### **PDFTRON:**





PDFTRON API is used to bring accurate PDF viewing, annotating, editing, creation, and generation to any web, mobile, desktop or server framework or application. We used a demo version of this api and full document and data lifecycle by deploying on our infrastructure.

We implemented it in our Node Js framework and used its features of demo/developer free version of API such as page extraction, page removal, watermark, pdf conversion. etc

## **METHODOLOGY:**

Since this project is built on a local server and is not hosted for public use, Back end or server side programs were built using expressjs . Many third party packages are used to avoid lengthy and extensive coding like FLASK, MULTER, etc

Firstly, the Web page was defined with tags using EJS and basic packages were imported to use them in our JS file. Then we used Multer to upload our file onto a local server by using get and post requests and we defined the path of the directory where it has to be saved.

After a file is uploaded and saved , we call PDFTron web API, specifically its one part known as PDFNET (which was imported earlier) using Asynchronous JS. We use await function to avoid crashes and delay , and for also smooth functioning of function calls.

After API calls are made, files from local server gets uploaded on API cloud or storage and processed accordingly and it sends back using output defined by us on local server, from there we download the output file on our local system.

We also implemented try and catch blocks in our program to handle errors.

Following things are implemented in the program :

- ***Conversion of docx, png, jpg, jpeg ,etc into pdf format***
- ***Remove a page from pdf***
- ***Create watermark in pdf***
- ***Create thumbnail in pdf***
- ***Extract text from pdf***
- ***Optimize a pdf***

## **APPLICATION:**

This project has a great deal of application from school/college to businesses, nowadays everything is going paperless so pdfs are today's digital paper.

- PDF manipulation and conversion tools are in great demand online for single time use especially at free of cost.
- We can convert any file in docx, jpg, jpeg format into pdf format efficiently and effectively.
- We can also optimize the size of the pdf (Program removes redundant info and spaces).
- We can remove the page from the pdf by specifying the page number.
- We can also add a watermark onto our pdf .
- We can extract text from pdf and save it in txt format so that text can be used in further manipulation.
- One can create thumbnails of pdf to use to publish the first page of pdf in png format.
- There can also many more possibilities of pdf manipulation which are left to implement yet like merging pdfs, pdf to other formats, etc.

## **FUTURE SCOPE:**

There have been several platforms who provide pdf conversion and manipulations like online but they lack better interface and services and of course they aren't free after a certain extent.

After covid , when the online education market has boomed , online documentation got more into people's life from lower class students to large corporate businesses , from university research work to small businesses, these types of online document manipulation and conversion has great demand in the market. According to needs, some might need one time use and some might need everyday. It can become one's business prospective if implemented effectively.

Several APIs are present to implement this , one needs to purchase them or build themselves one , it can assist millions users. Quality conversion , data integrity , data security are still challenges to tackle but it has been a growing market and continues to be.

## **REFERENCES:**

This project took references from the following resources:

- Node Package Manager Official website npm js: <https://www.npmjs.com/>
- JAVAPoint <https://www.javatpoint.com/expressjs-tutorial>
- PDFTron API Official Documentation:  
<https://www.pdftron.com/documentation/nodejs/get-started/integration/linu>  
[x](#)
- Stack Overflow: <https://stackoverflow.com/>