Politechnika Warszawska
Wydzia Elektroniki i Technik Informacyjnych
Instytut Informatyki

Academic year 2016/2017

Praca dyplomowa magisterska

Dennis Durairaj

# Creating PDF Documents from Web Applications

Opiekun pracy:
Tytu Imi i Nazwisko

Ocena .....

.....
Podpis Przewodniczcego
Komisji Egzaminu Dyplomowego


*Specjalno:* Informatyka –
Inynieria oprogramowania
i systemy informacyjne

*Data urodzenia:* 1 stycznia 1980 r.

*Data rozpoczcia studiw:* 1 padziernika 2002 r.

## yciorys

Nazywam si ....

.....
podpis studenta

## Egzamin dyplomowy

Zoy egzamin dyplomowy w dn. .....

Z wynikiem .....

Oglny wynik studiw .....

Dodatkowe wnioski i uwagi Komisji .....

.....

## Streszczenie

*Praca ta prezentuje ...*



**Sowa kluczowe**: *sowa kluczowe.*



## Abstract

**Title**: *Creating PDF Documents from Web Applications*

*This thesis describes an in-depth research in handling creation of PDF documents through various web technologies in both the client-side as well as server-side.*



**Key words**: *key words.*

# Contents

# 1. Introduction

## 1.1. PDF - Portable Document Format

PDF was developed in the early 1990s as a way to share documents with people over different locations, which included formatting text as well as embedding inline images. Before the rise of the Internet and HTML, PDF was used mainly for publishing workflows.

### Why Use PDF?

A well structured PDF will maintain the original text format, images, as well as the keep perfect layout of the document. PDF was mostly used by graphic designers and publishers for producing color page documents and designs. A PDF file can be shared, viewed, and printed by anyone using the freely available PDF reading softwares without depending on the type of the operating system, the original design application or fonts. However, with the changes happening in technology, nowadays PDF is used for any type data to be shared between users or applications. It is an open source file format specification and PDF is freely available to people who want to create tools for developing, viewing or manipulating PDF documents.

### Why is PDF important?

Today from a user's stand-point, it is becoming increasingly easier to create PDF files as the process is become as simple as printing a document. To put it in other words, anything that can be done with a piece of paper can be done with a PDF. Offset printed documents are using the technoligy of PDF more frequently. Adding to mainstream usage is the fact that a large number of applications allow users to save, upload or download a document as a PDF, and you can also find a variety of PDF conversion softwares tools freely available. With the power to embed metadata in a PDF document, along with the use of password protection options and electronic signatures, PDF is also being used as a standard for data archiving. It may have not been the perfect solution in the beginning but with years of dedication and effort by the development team at Adobe, today a large number of people are turning to PDF as a solution for something no one thought of in the early 1900s.

### Technical details about PDF

A PDF file is a 7-bit ASCII file, with the exception for certain elements that may have content in binary format. A PDF file begins with a header that contains a magic number and the version of the format. The format is a part of the COS ("Carousel" Object Structure) format. A COS tree file consists main of objects, of which there are eight types:

— Boolean values, representing true or false

— Numbers

— Strings, enclosed within parentheses ((...)), may contain 8-bit characters.

— Names, starting with a forward slash (/)

— Arrays, ordered collections of objects enclosed within square brackets ([...])

— Dictionaries, collections of objects indexed by Names enclosed within double pointy brackets

— Streams, usually containing large amounts of data, which can be compressed and binary

— The null object

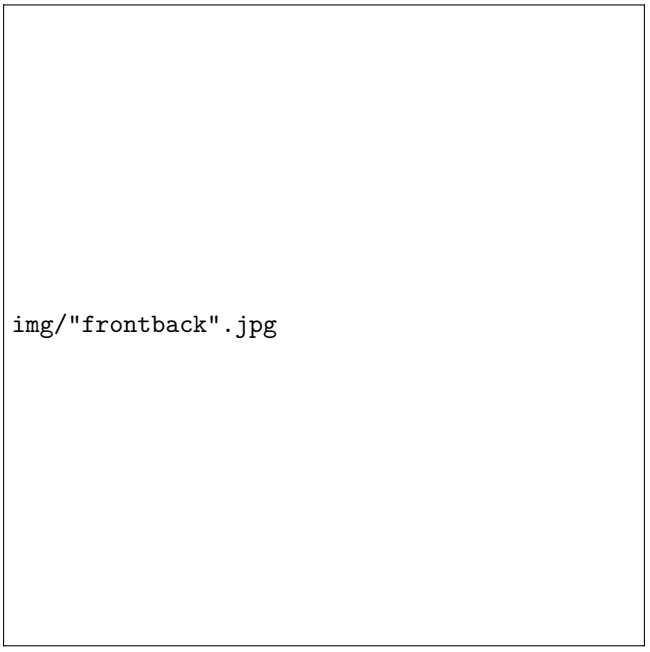


Figure 1.1. IDEF0

## 1.2. Web Communication Methodologies

### 1.2.1. An Overview

## 1.3. Current methods to convert data to PDF

# 2. Creating PDF documents in front-end

## 2.1. Frontend technologies

Front-end web development (also known as client-side development) is the practice of producing websites or Web Applications using HTML, CSS and JavaScript so that an end user can see and interact with them. The challenges that come with front end development is that the tools and techniques used to create the front end of a website change constantly and so the developer needs to constantly be aware of how the field is developing.

## 2.2. HTML5, CSS and JavaScript

a. **What is HTML5?**

HTML, or HyperText Markup Language, is the core element of the Internet. Its the language used to describe how a webpage should be structured. However, HTML on its own is bland because it can only display static pages; so in order to meet the increasing demand for more impressive web features, HTML with plugins like CSS, Flash, Java, Silverlight, etc. create the modern day Internet that we utilise in our daily life.

However, it has become something of a mess since different browsers have implemented these features in their own ways. With the advent of HTML5, it is meant to solve HTMLs big problems for a cleaner and more efficient web.

b. **Cascading Stylesheets**

CSS is a styling language that defines layout and design of HTML documents. For example, CSS covers margins, lines, width, background images, fonts, colours, height, advanced positioning and many other things. HTML can be used (or misused) to add layout to webpages. However, CSS offers additional options and is more accurate and sophisticated. CSS is completely supported by all browsers today.

In order to use HTML and CSS together, you use HTML to describe the body of the document and CSS to specify the document's layout, visual appearance, style,etc. not its content.

c. **JavaScript - The language of the Web**

JavaScript is a powerful client-side scripting language which has recently found its way into server-side scripting as well thanks to node.js which is a JavaScript based server-side programming framework. JavaScript is used mainly for improving the interaction of a client with the webpage. In other words, you can make your webpage more interactive and user-friendly, with the help of JavaScript. Today, from the browser to the server, JavaScript proves to be one of the most popular and versatile languages powering the modern web.

Client-side JavaScript makes use of the core language by providing special objects to control a browser using its Document Object Model (DOM). For instance, client-side extensions allow front-end developers to respond to user events such as clicks, scrolls, page navigation etc.

JavaScript has made its way to the server-side fairly recently. It extends the core language by supplying objects used to run JavaScript on a server. For instance, extensions on the server-side allow an application to provide passing of information from one invocation to another of the application, communicate with a database, or perform file manipulations on a server.

## 2.3. Built-in browser PDF converter

Modern browsers have the capability of saving webpages as a PDF file. This can come in hand in situations where one would like to save certain details. However, this is a read-only method of creating PDF files from webpages and involves no special application or code to achieve this feature. This method fails to save form details which requires more complex procedures since we need to handle validation to produce a valid form/document for the user.