

Politechnika Warszawska
Wydział Elektroniki i Technik Informacyjnych
Instytut Informatyki

Academic year 2016/2017

Praca dyplomowa magisterska

Dennis Durairaj

Creating PDF Documents from Web Applications

Opiekun pracy:
Tytuł Imię i Nazwisko

Ocena

.....

Podpis Przewodniczącego
Komisji Egzaminu Dyplomowego



Specjalno: Informatyka –
Inżynieria oprogramowania
i systemy informacyjne

Data urodzenia: 1 stycznia 1980 r.

Data rozpoczęcia studiów: 1 października 2002 r.

Wzrost

Nazywam się

.....
podpis studenta

Egzamin dyplomowy

Zoży egzamin dyplomowy w dn.

Z wynikiem

Ogólny wynik studiów

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	2
1.1. PDF - Portable Document Format	2
1.2. Web Communication Methodologies	3
1.2.1. An Overview	5
1.3. Current methods to convert data to PDF	5
2. Creating PDF documents in front-end	6
2.1. Frontend technologies	6
2.2. HTML5, CSS and JavaScript	6
2.3. Built-in browser PDF converter	7

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 technology 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 1990s.

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

1.2. Web Communication Methodologies

When facing frontend development, we start with the browser and the capabilities it offers. But with backend development, the field is much wider.

Firstly, we need to think about the language, with a plethora of server-side language available today, it is sometimes a problem to choose one with each having its own advantages. The language you choose will determine the operating system to install in the server. A web framework is a set of tools that can be used to help ease the process of setting up the server, since it provides functions and methods for the developer to work with and instead of creating his own from scratch. Hence it solves many web development problems and provides a good structure to begin with. This accelerates a lot of the initial setup. We shall touch on some backend frameworks in the following sections.

Languages and Frameworks

a. PHP

PHP is maybe the most popular language for web development. Its pre-installed in almost all hosting services. It has a syntax very similar to C and Java, so coming from these languages is a plus in familiarity (its my case).

It started as a procedural language, making a transition to object orientation in version 4, and finally being a true object oriented language in version 5. Version 7 bring more features to the language, and makes great improvements to speed and memory consumption.

Facebook is built with PHP, although they wrote some libraries and compilers to optimize the speed.

The classic and popular frameworks for PHP are Zend Framework and Symfony. I have been using symfony 1.1, symfony 1.4 and Symfony2 for several projects, it was a very good tool.

Nowadays there are many good alternatives, such as Yii and Laravel.

b. Python

Python is a language that uses a simpler syntax than PHP. Its designed to have a very readable code, and for that reason is very recommended to learn programming.

Its well tested, Google chose it to develop their services, and thats a good thing.

I havent used this language for any web application, but I used it to develop a tetris-clone game using the library Pygame.

The most popular framework for Python is Django.

c. Ruby

Ruby is designed to be a fun language. As the slogan says: a programmers best friend. It has a focus on simplicity and productivity with an elegant syntax.

In Ruby everything is an object, and that's interesting because it encourages to the programmer to think this way when developing. But, in my opinion, the most amazing thing about Ruby is the community. There is a huge amount of libraries (called gems) that you can use in your projects, making development very fast. The popular gems (which are many and varied) are well maintained and constantly improved. Twitter was built using Ruby, although now it's rewritten in Java.

The most popular framework for Ruby is, by far, Ruby on Rails, but for very small projects there is a popular micro-framework: Sinatra. Micro-frameworks are designed for small applications, having few files and being easier to maintain than (mis)using a full framework.

d. **ASP .NET**

ASP.NET is an open-source server-side web application framework designed for web development to produce dynamic web pages. It was developed by Microsoft to allow programmers to build dynamic web sites, web applications and web services.

It was first released in January 2002 with version 1.0 of the .NET Framework, and is the successor to Microsoft's Active Server Pages (ASP) technology. ASP.NET is built on the Common Language Runtime (CLR), allowing programmers to write ASP.NET code using any supported .NET language. The ASP.NET SOAP extension framework allows ASP.NET components to process SOAP messages.

ASP.NET's successor is ASP.NET Core. It is a re-implementation of ASP.NET as a modular web framework, together with other frameworks like Entity Framework. The new framework uses the new open-source .NET Compiler Platform (codename "Roslyn") and is cross platform. ASP.NET MVC, ASP.NET Web API, and ASP.NET Web Pages (a platform using only Razor pages) have merged into a unified MVC 6.

e. **NodeJS**

Node.js is an open-source, cross-platform JavaScript runtime environment for developing a diverse variety of tools and applications. Although Node.js is not a JavaScript framework,[4] many of its basic modules are written in JavaScript, and developers can write new modules in JavaScript. The runtime environment interprets JavaScript using Google's V8 JavaScript engine.

Node.js has an event-driven architecture capable of asynchronous I/O. These design choices aim to optimize throughput and scalability in Web applications with many input/output operations, as well as for real-time Web applications (e.g., real-time communication programs and browser games).

The Node.js distributed development project, governed by the Node.js Foundation,[6] is facilitated by the Linux Foundation's Collaborative Projects program.

1.2.1. An Overview

to 0.8 — X[l] — X[c] — X[r] —

item 11 item 12 item 13

item 21 item 22 item 23

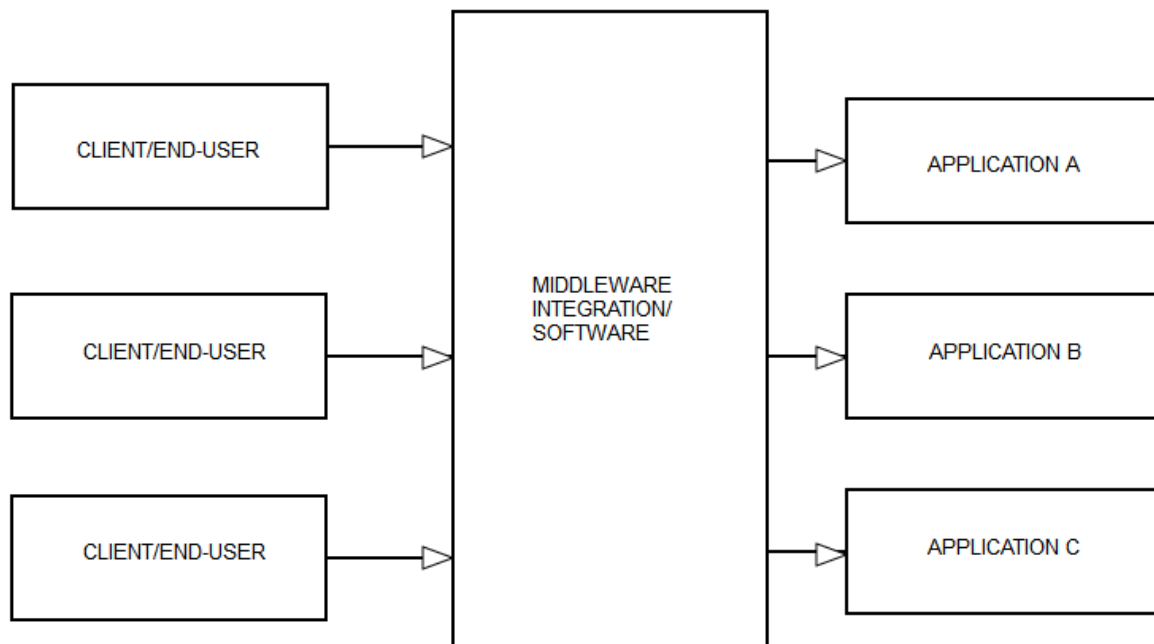


Figure 1.1. IDEF0

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.