

Simply Written

by David Turner

Introduction

Simply Written is aimed at being a digital book authoring and publishing platform, giving authors more control over distribution than ever before. In this document I want to cover the history of this project, from the initial idea through to where things currently stand, as well as take a look at where things are going in the future as well.

Background

Every web based service I work on has a starting point firmly rooted in solving a problem. Quite often it relates to some form of issue I have encountered but sometimes the idea comes from the problems others have. In this case it is the latter.

I Like Reading

Ever since I was a child I have enjoyed books. Sitting down and losing myself in a world someone else can craft with just words is a truly enjoyable experience, one which is as true today as when I was younger.

I Like Digital

I work in an entirely digital world, and I own a lot of devices through which I am able to consume information. My computer, my iPad, my iPhone and my kindle are all capable of reading books.

It's very convenient for me to be able to bring along a collection of books on the devices I use every day. It's also significantly lighter to have them on such devices than carrying the printed versions.

I Have an Author Friend

I have a friend, Tami Olsen, who has released two books herself. Both were available in both printed and digital formats. Both books were really enjoyable experiences in terms of reading when printed. This sadly wasn't the case when it came to the digital versions of the same content. Why?

Most Publishers Produce Poor Digital Books

Sadly the formats that are commonly used when sending books off for print aren't so suitable for conversion to other digital formats. Older books might not be saved in a format that easily translates to other formats. This, obviously, causes a problem.

The Problem

The problem is that content made available in two different formats, print versus digital, results in two entirely different experiences. This is often fine, but when one results in an enjoyable experience and the other leaves you wanting a refund, you have something that needs to be resolved.

Poor Quality Digital Books

The obvious problem is that the quality of digital versions of books is not on the same level as that of the printed edition of the same book. Different is fine when it comes to the content in a book, but worse is never a good thing.

What is the Cause?

It's easy to point fingers and assign blame. It's harder to find the cause of the problem. You could blame the publisher. You could blame the author, they did after all write the book. You could blame the format. Working with a format that causes issues isn't pleasant. From the time I have spent looking into resolving this issue the problem lies with process.

What Can We Do About It?

Having a process that works well for a single format is fine... if you are working in that single format. This is not the case if you are trying to use the same process to output multiple formats. This approach causes problems that need to be dealt with.

The Solution

A problem with process is one that is easily fixed. You change the process. In order to be able to make that change you first need to get an understanding of how things need to change, and how you can get from the current process to the one that will resolve the current issues.

Doing It Right

The current process in place takes a single, very inflexible, format and tries to make it more flexible. The PDF format is a useful format for *presenting* information but it does not serve as a good starting point for *storing* information. This poses a question "How do we store this information?" If the PDF format isn't the best format for storing the information then we need to identify a better way to store this information.

There are lots of ways that information can be stored. There are multiple formats. Microsoft Word files are a common starting point for a lot of content. As are Apple Pages files. Unfortunately neither of these formats are particularly easy to work with outside of software that is tailor made

to work with them. This causes issues with authoring content in multiple formats, and is a similar problem to what currently exists.

Let us take a different approach to solving this problem. What is the best format for content to be stored in so that it can be used to create content in as wide a range of content as necessary? In the digital content world there are two key formats for digital content:

- ePub (for iPad, Android and pretty much every other digital book reader)
- mobi (for Kindle devices)

The ePub format is an open format, which makes it easy to work towards. It is based largely on xHTML which is, rather conveniently, text based. This means that content in xHTML format can be opened in pretty much any content editing software. It also means that it can be, quite easily, converted from one format to another.

The mobi format is not nearly as open as it owned by the makers of Kindle, Amazon. Fortunately there are communities dedicated to reverse engineering technology involved in pretty much everything, including the mobi file format. As it turns out the mobi file format is very similar to that of xHTML, with a few extra elements added in for the benefit of Kindle software.

This means that we have two file formats that can be based on identical content in a format that is entirely text. The only remaining format is the PDF format which is largely used for print. It is possible to take content stored in xHTML and convert it into a PDF file.

This gives us a common language that covers all of the formats needed but should also be rather future proof as the underlying language has been build to survive the test of time. So we have an end result that meets with our needs. How do we get there?

Take Control of Content

The first step is to have a way of controlling the content that goes into books. The approach of write a book, create a PDF of the book, and ship for printing no longer fits with the formats needed to produce both printed and digital books. If we want to work with xHTML content that will work in any format we need then the content needs to be xHTML itself. This leads to the question of “How?”

How Do We Do This?

The question of “How?” can be a difficult one to answer and requires a better understanding of what is needed for each of the formats.

Look At The Formats

In the context of digital book content we have already identified that both ePub and mobi files are the two formats used in digital books. To answer the question of “How?” we need to get a better understanding of these formats and how we can make them.

ePub

The ePub format is the easiest format to get to grips with, as it is completely open which makes research easier to understand the intricacies involved with creating a book in this format. In particular the work of Paul Salvette detailing how to create books in this format served as a very useful starting point for my work.

An ePub file consists of a series of files that make up and manage the book, compressed and stored in a single file. A significant amount of this is the content written by the author, but there are parts which deal with information about the book, so that applications which can read the content are able to inform the user of the contents.

Mobi

By contrast the mobi format is quite difficult to get much of an understanding at all about how the format works. As it is a proprietary format owned by Amazon it is a lot more difficult to get an understanding for how things work. Fortunately Amazon provide a tool, called KindleGen, which is able to create the required files based largely upon the same content you would use to create an ePub file.

It is not quite that easy sadly. Whilst KindleGen will, quite happily, produce mobi files from the same content as ePub files, there are some consistency issues. This is mostly an issue of duplicate content but it can be frustrating to encounter such issues with a platform that has no documentation. Fortunately some trial and error was able to identify that some of the data that is used by ePub book readers caused the issue with mobi files. Removing it from the mobi version resulted in the issue disappearing.

Building Books

By taking the time to develop an understanding of what makes up both ePub files and mobi files I have been able to better appreciate the intricacies of the formats. Going forward I wanted to be able to create books on the fly, so taking the time to work out what causes issues and how to avoid them will be beneficial further into development. For now I want to get a better understanding of the process of actually creating books from content.

Start Small

Nothing complex appeared fully formed. A wall is built from many bricks. A band is made up of many members. A system is made up of many components. Starting small allows you to create the pieces of something larger. As I want to be able to create digital books I think that it is important to develop something that is capable of creating the books. I can look into content management later.

Semi-Manual Creation

There are tools available for creating digital books, but they all require a lot of effort on the part of the person using them. I know that with the completed system that I will be able to collect a lot of the information that these tools require. The aspect of these tools that I need to get working is taking the information and producing a finished digital book.

Focus on a Format (ePub)

To start with I want to focus on a single format. There is no point in working on multiple formats and dividing my attention. It would result in less being achieved in the same amount of time. Fortunately, in this instance the choice is easy. As mobi files are generated from almost identical information to those of ePub files, it is an easy choice to work towards getting ePub files working.

The process of preparing the content for the creation of an ePub file is rather easy. Actually converting the files into a finished ePub file is much more problematic. This is the result of a single file, called a mimetype file. This is a small file that lets an ePub file know that it's an ePub file and, of all the files that make up a book, it is the only file that cannot be compressed. This proved problematic but was eventually I was able to work around it.

Work on Other Formats (mobi)

From the foundation of ePub generation it was very easy to take the same content that was used there and repurpose it for mobi file generation. With this foundation it was very easy to start generating this second format.

Crafting Content

Having worked out a system for generating books the next stage is to develop a system for authoring content. This is the second part of the equation. Once both are working individually I can bring them together to create digital books from user generated content.

Build a System

It is important to ensure that I have a system that can handle authoring content. To create this I will be leveraging my development skills in order to quickly build and iterate a platform for Simply Written.

PHP

PHP is one of several server-side languages that be used for creating web based systems. When we started working on Simply Written it was also the language I was most comfortable writing, which made it an obvious choice for the work I wanted to do. Having the understanding of PHP that I

have allowed me to quickly get some basic systems in place to allow us to test and iterate ideas rapidly. The value of this can not be over stated.

Frameworks

Frameworks provide a set of tools that make working in a certain language easier. They are designed to do most of the heavy lifting so that the people working on a project can do exactly that, work on the project. With Simply Written I used the Wee_ Framework, based off of my own work over the past year. It's focused on giving only the bare minimum of assistance to get things working. So a limited set of tools that let you get work done. As a starting point for any project, I couldn't ask for anything more.

Book Control

As a part of authoring content it is important to provide control over the various aspects that make up a book. By providing the author with this level of control it ensures that the finished books will be tailored to the person writing them.

Books

This starts with giving the user control over the various aspects of the book itself. This goes beyond the Title and Author of the book, and delves into the various details that help identify a book, as well as some information about where it can be purchased once it is published.

Chapters

A book is nothing without it's chapters. With chapters of a book comes content. By separating this section of content we are able to ensure that generated content will be of the highest possibly quality regardless of the format generated.

Bringing It Together

With both halves of Simply Written developed independently of one another is is time to bring them together to create a single piece of software.

Building Books + Crafting Content = Simply Written

Writing content with no way of outputting it means nothing. The same is true of generating a book from static content. Together, however, they provide the very core of Simply Written. These two elements provide one central location to users to author and generate digital books that they can then sell.

By focusing on web technologies this also has another benefit. With everything being focused on

the web, their content is securely stored online. This means that authors can work on their content from any device that they want. It also prevents potential loss of content in the event that they experience issues with their own hardware.

Where We Are

That brings us up to date on what Simply Written has gone through, in bits and pieces. I am very happy with the progress that I have made thus far with Simply Written. It is currently at a stage where it can be used to create content that can be downloaded in both mobi and ePub formats. This content is currently limited to text based content but to have achieved as much as I have in such a short space of time makes me optimistic for this changing in the near future.

In addition to this I am opening up a public beta of Simply Written for people to get to grips with it, and to get a wider base of users testing it for me. You can register by visiting <http://davidturner.name/recommended/simply-written>.

Where We Are Going

I have covered the the idea behind Simply Written, and documented the process, in general terms, behind generating the file formats currently supported by the system. Covering what has been and what is does not cover the full scope of my work. Possibly the most important part of Simply Written's development is where things are going.

New Format (PDF)

Throughout this document I have referred to PDF files as being used for print. They are also used generally as a format for documents. Whilst working on Simply Written I originally focused primarily on ePub and mobi formats, but there is a market for PDF versions of books as well.

By developing for this format as well people will have an easier time creating content that truly bridges across digital and print, as authors can use the PDF formatted file for print. In addition to this the PDF file can also be used on most digital book readers as well as on computers. This is highly beneficial as neither ePub nor mobi files can be opened by default on most computers.

Further Development (Image Support)

In addition to additional formats for content to be generated in, I also need to develop the feature set of the site. This is entirely focused, in the immediate future, on image support for books. Providing visuals to accompany the text, and adding book covers, are of great importance in many forms of written content. I do not want to exclude them.

It needs to be handled correctly, so it is not something I will develop in a rush. When image support is added I want to ensure that I will not need to completely overhaul the feature, as this would cause issues going forward.

Design

Simply Written, to date, has been very heavily focused on the development side of things. This is another area that will require a great amount of attention. The underlying architecture of Simply Written has been developed to be of a high quality. The design of the site needs to match that.

Branching Out

Simply Written is aimed squarely at authoring digital books. There are other areas that similar approaches could be applied. One of these is journalism. Investigative journalist Lyra McKee has been working with me to secure funding to develop a similar service focused on providing similar tools to journalists.

Pricing/Promotion

Pricing and promotion are two very important areas to consider when it comes to developing a web based service. You don't want to be providing a service at a loss after all. It is also important that you have people using it.

These are also areas that I have only vague plans. With pricing I am currently of two minds. The first pricing plan I have considered is to charge a monthly fee, with possible discounts for yearly payments. I am somewhat wary of this approach as people are not always writing, and subscription fees may push people away from Simply Written

The alternative I am considering is to charge per publish. This might be a better approach as it allows people to author content and, once they are ready to sell their work, they can generate the finished books. This approach may result in an income that may not pay for the product however. It may be that neither is a good fit. This is why I have yet to settle on a pricing structure.

Promotion is also another area that could be troublesome. I am fortunate in that I know authors of different types of content, both educational and non-educational. These authors know other people that produce similar content and I hope to be able to leverage this word of mouth promotion. I do not believe that this will be enough to ensure that Simply Written develops a large user base so it is, again, another area that needs further thought.

References

Paul Salvette, 21/08/2011, Paul Salvette in Bangkok: EPUB Template - eBook Formatting [Paul Salvette in Bangkok], [online]. Available: <http://www.paulsalvette.com/2011/08/epub-template-ebook-formatting.html> [19/11/2011].

undefined, 03/01/2011, Downloading a .mobi file from a website [Kindle Boards - Index], [online]. Available: <http://www.kindleboards.com/index.php/topic,47503.o.html> [19/11/2011].

Michael Butler, 16/11/2011, How to Specify a Cover on KindleGen | eHow.com [eHow | How to Videos, Articles & More - Discover the expert in you. | eHow.com], [online]. Available: http://www.ehow.com/how_12164382_specify-cover-kindlegen.html [22/11/2011].

Araby, 29/12/2010, Kindle book cover size | Webbish Books [Webbish Books], [online]. Available: <http://webbishbooks.com/2010/12/kindle-book-cover-size/> [19/11/2011].

George Benthien, 16/08/2011, Example OPF File [Numerical Mathematics Tutorials], [online]. Available: http://gbenthien.net/Kindle/opf_example.html [19/11/2011].