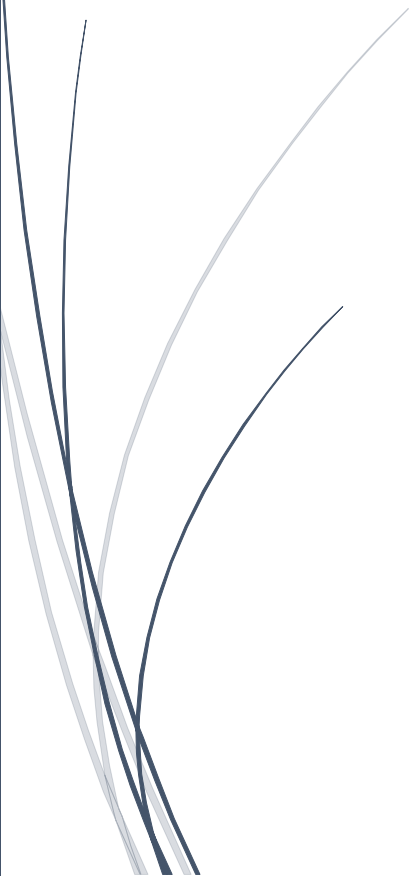


Teaching students how to single-source documents with Pandoc and Markdown

A proposal for the Technical Writer Program at Algonquin
College



Jesse Laprade
ALGONQUIN COLLEGE
2018-11-20

Contents

Executive summary	2
Glossary.....	3
Introduction	3
Problem: Students are not provided with sufficient experience in single-sourcing documents.....	3
Solution: Teach a module on single-sourcing documents with Pandoc and Markdown.....	4
What is Markdown?.....	4
What is Pandoc?	5
Financial cost and time cost of adopting Pandoc and Markdown.....	6
Evaluation of Pandoc and Markdown against Microsoft file formats	7
Required criteria	7
Additional benefits.....	7
Limitations of Pandoc and Markdown.....	8
Implementation of Pandoc and Markdown.....	8
Phase I – Installing Pandoc on Algonquin College’s remotely accessible virtual machine	8
Phase II – Teaching Markdown	8
Phase III – Teaching Pandoc.....	8
Final recommendations	8
References	10

Executive summary

This proposal suggests that the teaching of single-sourcing documents to the Technical Writer Program students at Algonquin College will provide them with enough experience to understand how single-sourcing works at a basic level. A module on the teaching of single-sourcing documents must be equal or superior to using Microsoft file formats on the following two criteria:

- **Stable** – The conversion tool must be a mature, well-developed tool that will not crash easily, and the markup language must have already-made, and accessible documentation and standard(s)
- **Easily learned** – The syntax of Markdown is easily learned within several minutes and due to the nature of its syntax it additionally enforces the user to style their documents as they write in order to be formatted after being converted to other file formats and platforms.

Pandoc and Markdown, a conversion tool and a markup language, meet all of these criteria. Pandoc is a conversion tool, which is used to convert several file formats to and from one another. It can create slideshows, wikis, websites, Microsoft formats, PDFs, and many more. Markdown is a markup language that is simple visually and functionally. The user can easily format their document as they type it, which saves time and encourages the user to be critical while formatting, thus encouraging the planning of document structure.

Pandoc and Markdown are free, and the only costs involved are the instructor who teaches two 2-hour lectures (see Table 1 below):

TABLE 1: COST BREAKDOWN OF IMPLEMENTING PANDOC AND MARKDOWN

Resources	Cost	Time
Markdown	\$ 0.00	One 2-hour lecture
Pandoc	\$ 0.00	One 2-hour lecture
Instructor	Salary: \$75,000 Weeks worked per year: 52 Hours worked per week: 40 $(75,000/52) / 40$ = \$ 36.06 per hour \$ 36.06 x 4 hours of lecture time	
Total	\$ 144.23	Two 4-hour lectures

The cost of implementing this short module will not affect Algonquin College's budget as the instructor is already being paid to teach for the time being.

Glossary

Platform – A digital location in which a document can be presented, such as websites, slideshows, PDFs, etc.

Document – A digitally formatted computer file that has been created by a user

File format – A file suffix which is appended to either modify the internal data of a file, and/or identify a file type

Introduction

This proposal suggests that the single-sourcing of documents should be taught to the students in the Technical Writer Program at Algonquin College.

A majority of students in the Technical Writer Program at Algonquin College use either Microsoft Windows or macOS, which both have different Office suites available and both use different file formats. At Algonquin College, both Microsoft Windows and macOS users are advised to download a licensed version of Microsoft Office from the Algonquin College website to use software, such as Microsoft Word, PowerPoint, and Project. Microsoft products rely on Microsoft-specific file formats, and students both write and submit documents using these products and file formats. Every week, guest speakers have been coming in to talk about their workplace environment, and several of them have mentioned that single-sourcing documents is important when writing technical documents.

The Technical Writer Program at Algonquin College should teach students how to single-source their documents in plain text using Pandoc and Markdown. This proposal will show that this is an efficient and time-saving method of writing, formatting and converting documents between operating systems.

This proposal will first examine the inefficiencies of using Microsoft file formats to single-source documents, and the Technical Writer Program's use of these file formats that have made Microsoft Word the preferred solution to date. Second, it will explain the recommended solution and evaluate it against the use of Microsoft file formats. Finally, it will discuss the implementation of Pandoc and Markdown for the Technical Writer Program, which will save time and aid students in writing portable, formatted documents.

Problem: Students are not provided with sufficient experience in single-sourcing documents

After several interview questions with the Technical Writer Program coordinator, Jim Catton, there seems to be an interest in getting students more involved with single-sourcing documents earlier on, and exposing students to more platforms regarding document publishing. Currently, students create documents in either Microsoft Word, if the requirement is to submit an assignment in .docx format, or a text editor of the student's choice, if the requirement is to submit an assignment in .pdf format. Students become accustomed to using Microsoft file formats to create documents, which isn't always the situation in workplaces. Although working in Microsoft Word is efficient to work in, it doesn't allow for the exploration or learning of different platforms where documents may be published.

A portable markup language and conversion tool will not only allow for a more universal method of publishing documents and converting them to different platforms (single-sourcing), but it will also decrease the amount of time spent formatting documents, troubleshooting solutions to formatting problems, and figuring out how to use document formats on operating systems that do not have the proper software for editing Microsoft-specific formats.

A portable markup language and conversion tool must have the following features to be considered:

- Time-saving – This markup language must save time when writing, creating and publishing documents to other platforms, and save time when converting to other file formats
- Cost-effective – This conversion tool must not affect the budget of the Technical Writing Program or Algonquin College
- Lightweight – The conversion tool must not be resource heavy and must only leave a small memory footprint on the user’s computer or Algonquin College’s virtual machine to avoid unnecessary delays or inconveniences for the students
- Stable – The conversion tool must be a mature, well-developed tool that will not crash easily, and the markup language must have already-made, and accessible documentation and standard(s)

Additionally, a portable markup language and conversion tool will:

- Single-source documents with little effort
- Be easily learned
- Enforce document styling in students

Solution: Teach a module on single-sourcing documents with Pandoc and Markdown

A teaching module on Pandoc and Markdown will provide students with the experience in single-sourcing documents.

What is Markdown?

Markdown is a markup language that uses plain text [1] syntax to format documents. It is designed to be an “easy-to-read and easy-to-write plain text format” that avoids the look of having been formatted with tags or formatting instructions [2] and kept the writer focused on writing rather than stopping to pause and click to format. When saved, because it is plain text, it can be opened in any text editor.

Although it was originally designed to be converted to HTML or XHTML, it’s base features, such as titles, authors, dates, headers, bullet/numbered lists, links, citations, references, etc. were portable characteristics to other platforms such as wikis, XML, graphs, mathematical expressions, slideshows (including Microsoft’s PowerPoint), LaTeX, Emacs org, Microsoft Word, UNIX man pages, and many more. This caused an influx of people to create software that utilized the Markdown standard by creating multitudes of tools, that can convert it to multiple formats, including “What you see, is what you get” (WYSIWYG) software, some of which can be used live in a browser without any installation.

Markdown uses a minimal syntax to format documents, avoiding traditional tagging and nesting of content. A brief sample of what Markdown would look like, if it was converted to Microsoft Word can be seen below (see Table 2).

TABLE 2: MARKDOWN BEFORE AND AFTER BEING CONVERTED

Markdown syntax	Formatted result
# This is heading 1	This is heading 1
## This is heading 2	This is heading 2
### This is heading 3	This is heading 3
* This is a bullet * This is a second bullet * This is a third bullet	<ul style="list-style-type: none"> • This is a bullet • This is a second bullet • This is a third bullet
This is a [link](google.ca) in a sentence.	This is a link in a sentence.
# How to plant a garden	How to plant a garden
This is a [link](how-to-plant-a-garden) to the header above.	This is a link to the header above.
This is <i>italic</i> and this is bold .	This is <i>italic</i> and this is bold .

What is Pandoc?

Pandoc is a conversion tool for converting documents to and from dozens of different formats. It is a widely used tool, often for writers and scholars for publishing documents. Its most supported format is an extended version of Markdown, but several other versions of Markdown are also supported [3].

Pandoc is used as a background tool in many programs to output saved files into different formats. It is mainly used as a command-line tool as it is quick and efficient to bring up something as lightweight as a terminal such as Microsoft Windows cmd.exe, or a Bash shell in macOS and GNU/Linux operating systems.

Pandoc doesn't have a graphical user interface, but there are graphical programs that integrate it with it. Instead, it uses command line arguments that use simple options for outputting a format. The user simply types "-o" to specify that the following text will be the filename with the preferred file format to be created (see Figure 1 below).

```

Terminal
jesse@home:~$ ls
bak@ bin/ blog@ dev@ doc@ technical-writing-proposal.txt
jesse@home:~$ pandoc technical-writing-proposal.txt -o technical-writing-proposal.pdf
jesse@home:~$ ls
bak@ bin/ blog@ dev@ doc@ technical-writing-proposal.pdf technical-writing-proposal.txt
jesse@home:~$

```

FIGURE 1: A SCREENSHOT OF A PANDOC COMMAND BEING USED IN A TERMINAL TO CONVERT A MARKDOWN FILE INTO A FORMATTED PDF

Financial cost and time cost of adopting Pandoc and Markdown

Pandoc and Markdown are both cost-effective solutions. The financial cost and time cost breakdowns are as follows (see Table 3 below):

TABLE 3: BREAKDOWN OF TOTAL COSTS INVOLVED WITH ADOPTING PANDOC AND MARKDOWN IN THE CLASSROOM

Resources	Cost	Time
Markdown	\$ 0.00	One 2-hour lecture
Pandoc	\$ 0.00	One 2-hour lecture
Instructor	Salary: \$75,000 Weeks worked per year: 52 Hours worked per week: 40 $(75,000/52) / 40$ = \$ 36.06 per hour \$ 36.06 x 4 hours of lecture time	
Total	\$ 144.23	Two 4-hour lectures

These costs indicate that Pandoc and Markdown themselves are very cost-effective as they are free, and the only cost, which would already be included regardless of what was being taught, would be the instructor's salary broken down into an hourly wage and then multiplied by the total amount of lecture hours it would take to teach Pandoc and Markdown.

Evaluation of Pandoc and Markdown against Microsoft file formats

For Pandoc and Markdown to be considered, it must be equal or superior to Microsoft file formats on the two criteria (Stable, easily learned), and should meet the last three criteria (easily single-sourced, easily learned, enforce document styling). Pandoc and Markdown meet all the criteria of the ideal solution (see Table 4 below).

TABLE 4: EVALUATING PANDOC AND MARKDOWN AGAINST MICROSOFT FILE FORMATS

Criteria	Pandoc and Markdown	Microsoft file formats
Time-saving	X	
Cost-effective	X	
Lightweight	X	
Stable	X	X
Easily single-sourced	X	
Easily learned	X	X
Enforce document styling	X	

Required criteria

Pandoc and Markdown are designed to provide a time-saving solution to document formatting and single-sourcing documents. It is cost-effective in that it is completely free, and always will be due to its licensing. Pandoc, being a command line tool, is lightweight and does not require a powerful computer to use, as well as having the common reputation of other command line programs of being extremely stable. Documents are styled in any chosen markup language, many prefer Markdown, which can be converted to many different file formats to provide for an easily single-sourced document. The syntax of Markdown is easily learned within several minutes and due to the nature of its syntax it additionally enforces the user to style their documents as they write in order to be formatted after being converted to other file formats and platforms.

Additional benefits

Because the user styles their document as they write, when using Markdown, the user spends more time writing, and less time formatting – the formatting is left for Pandoc to handle. This keeps the user's thought process focused on the idea they want to purvey.

Many examples of Pandoc and Markdown are available on both of their websites, and while Pandoc is a command line tool, which the average user is not used it, it is very intuitive.

Many users who have recorded their time writing, editing, and formatting report that they spend on average 1/3 of their time formatting when using document processors such as Microsoft Word, LibreOffice/OpenOffice, and Google Docs. Pandoc eliminates this time completely, by embedding the formatting syntax in the user's text. This means users spend more time creating quality text, rather than spending time on quality formatting, which is handled by Pandoc.

Markdown documents can be effortlessly exported to webpages, slideshows, manual formats, Microsoft Word, PowerPoint, PDFs, math expressions, and many more.

Limitations of Pandoc and Markdown

Pandoc and Markdown may be deterring due to the nature of their interface, or lack thereof. If the user is not using a Markdown editor and just using a plain text editor that does not provide syntax highlighting or a live preview of what is being typed, the user could be unmotivated to learn if not provided with the right resources, software-recommendation-wise.

Implementation of Pandoc and Markdown

Pandoc and Markdown should be introduced in three phases to avoid confusion in students. The first phase will be completed by Algonquin College before the program, and the last two phases will consist of two 2-hour lectures that involve students and an instructor.

Phase I – Installing Pandoc on Algonquin College’s remotely accessible virtual machine

In phase I, Pandoc will be installed on the Algonquin College’s remotely accessible virtual machine. This will allow students to access it as a tool in the same manner that Framemaker and MadCap Flare are accessible as remote software. This way, students have the option of installing and using it locally for free on their personal computer, or using it remotely. This will avoid any in-class issues with students trying and waiting to install Pandoc, and save time for the instructor and program.

Phase II – Teaching Markdown

In Phase II, an instructor will teach the basics of Markdown, followed by a hybrid hour of advanced Markdown syntax after class. Students will be provided with a list of available Markdown editors to use, including online web-based ones, or they can use any text editor of their choice, which could be something as something as Microsoft’s notepad.exe will work. Students will practice creating files in Markdown and view their live preview with any of the Markdown editors that provide the WYSIWYG feature.

Phase III – Teaching Pandoc

In Phase III, an instructor will teach the basics of Pandoc, followed by a hybrid hour of additional Pandoc features, and links to documentation for further Pandoc use. Students will be provided with a list of examples of how to convert different file formats to and from one another. Students can then further explore how these documents look on different platforms and as different file formats.

Final recommendations

This proposal recommends that the Technical Writer Program at Algonquin College adopts a lecture on Pandoc and Markdown to provide students with practice and experience in single-sourcing documents before they begin their co-op or future work, and to provide students with documents that can be showcased on multiple platforms for future portfolio pieces.

Pandoc and Markdown have no additional costs, and would only require two short 2-hour lectures in total.

Students will spend less time focusing on formatting, and more time creating quality content.

After finishing these two short lectures, students will become more critical when formatting documents in other software such as Microsoft Word, LibreOffice, Pages, or Google Docs, etc.

References

- [1] "Plain text - Wikipedia," 2018. [Online]. Available: https://en.wikipedia.org/wiki/Plain_text. [Accessed 2018].
- [2] "Markdown - Wikipedia," 2018. [Online]. Available: <https://en.wikipedia.org/wiki/Markdown>. [Accessed 2018].
- [3] "Pandoc - Wikipedia," 2018. [Online]. Available: <https://en.wikipedia.org/wiki/Pandoc>. [Accessed 2018].