

How many words are in an average A4 page?

Luc Angevare

April 2022

1 Introduction

1.1 Information page

How many words are in an average A4 page?

An analysis and simulation on the amount of words that fit on an average page

Author: L. Angevare
E-mail-address: lucangevare@gmail.com

Education: Corlaer College, Atheneum
Nijkerk, April 5, 2022

1.2 Preface

Many websites state that a good rule of thumb is that 500 words make an A4 page, however, my experience said that this number was more around the 400 words. This made me interested enough to program a simulation to test the biggest repository of English words (which consists of more words than the Oxford English Dictionary) to see how many words would fit on the average A4 page.

This paper aims to find the true number of words a page can fit, so students have a true answer on about how many words they would need to write to fit the assigned amount of words. This can be an important and convenient statistic for preparing for an oral exam or research paper, when a teacher assigns one of these with a word count, students know about how many pages would correspond to that.

I'm Luc Angevare, and this research was written during my education of Atheneum, as well as my experience with programming in various languages including but not limited to Python (the language I have decided to use in this research); this gives me the professional backing to be able to write the simulations to finish this research. For my methods of research, see the section Method, for my experience in programming and more information on me, visit my portfolio.

1.3 Table of Contents

Contents

1	Introduction	2
1.1	Information page	2
1.2	Preface	3
1.3	Table of Contents	4
1.4	Abstract	5
2	The research	6
2.1	Literature	6
2.2	Method	6
2.3	Discussion	7

1.4 Abstract

Various literary works and sources (see literature) state that on average 500 words fill an A4 page, this study however, has shown a different outcome. Instead of finding 500 words per average page, this study claims to find 392 words on average. The amount of characters found per page was closer to the amount the sources had stated. This amount was close to 3600 characters, as opposed to the amount of 3000 according to the cited sources. This means that per usable cm^2 1.12 words could be expected, and 10.29 characters. This, in turn, meant that on average every word in the English alphabet consists of on average 9.18, providing a stark contrast to the 4 characters that various sources provide.

2 The research

2.1 Literature

Some sources, even including a website called wordcounter.net which is devoted to and named after a way to count how many words corresponds to how many pages, said about 500 words makes an a4 page (interestingly when actually using the tool, the minimal amount of words to fill a page is 428, which means the company or people behind the website didn't setup their metadata correctly). The same goes for a website named yourdictionary.com, which includes a very neat table saying that one single-spaced page counts about 500 words, or a wordcounter.io blog and many other websites made and named just to count words. Another website named anycount.com backs all of these up, and goes a step further to say that 3000 characters fill a page. None of these, however, name any sources or research to show what their method was.

It seems they all get their data from one another and the first website just got their source from empirical data, opening up a random research paper, seeing how many words filled a page and assumed this was an average. This research goes to show the actual number of words a page fills, as well as some other linguistic statistics disproving the previous websites' statistics.

2.2 Method

The method for this program was more of a brute-force simulation than any calculation whatsoever, the program exists of exactly 50 lines of code and is in no way optimized nor efficient. The way the program works is by first saving the dictionary of theoretically 290.000 english words to memory (takes a lot of memory, but it means we don't have to look up and save a predefined amount of words every iteration, meaning we save time at the cost of memory). After this we run a for loop during the entire dictionary, in which we run another for loop for a predefined amount of words (between 300 and 700, pre-research shows that the minimal amount of words has always been above 300 and maximum under 700, so it was safe to start at 300 and end at 700).

In this nested for loop, the program opens the Word document so everything could be set up in a different thread if need be by the docx module while not being in the way. Then we get the array of words between the dictionary iteration and the dictionary iteration plus the amount of words determined by the second for loop. This array we implode with a space, so all the words would be put on a page with a space in between, as this is how sentences are usually written. We then save that document, use pandoc to convert it to a PDF because I was unable to read the amount of words off a docx document (I attempted opening it as .zip and reading the xml file, but this always displayed 1 page, whether it was 4 or 5 in reality, so I found that this was the best replacement method. It did turn out to be somewhat lossy because the margins created by LaTeX were much bigger than the ones created on default by Word, but I did correct for this in the actual calculations) and read the amount of words and pages off that. We then compare the amount of words we had expected to the amount of words we actually found to be sure that they were the same and nothing had

gone wrong there, and then continue.

We then add all found values to the CSV file. I chose CSV for this because Excel has a limit to the amount of entries it can have, and CSV was the fastest and easiest way to enter the data otherwise, as it is just a comma-delimited text document in principle. If we found the amount of pages to be 2, we break the second for loop to go back to the first one and iterate it by the amount of words that were used, so we get to use all of the words the dictionary had.

2.3 Discussion

The biggest point up for discussion is not one that impacts the research done in this study, the way I counted the amount of pages was by having to convert the Word document to a pdf using pandoc, the findings were impacted by this, but this was easy to correct for. The only thing that wasn't easy to correct for, was the amount of time lost doing this. Each run of the second for loop took about half of a second, and most of this was lost due to the conversion of document formats. This means that the whole research took on minimum $(290.000 \div 2) \div 60^2 = 40$ hours (unrounded 40 hours and 17 minutes). Doing this any other way would have spared at least $\frac{2}{3}$ of this time, but there were not really any other ways to do this in my research.