he source code behind docx2txt is derived from code in the docx package, which can also be used to scrape Word Documents. docx is a powerful library for manipulating and creating Word Documents, but can also (with some restrictions) read in text from Word files.

In the example below, we open a connection to our sample word file using the docx.Document method. Here we just input the name of the file we want to connect to. Then, we can scrape the text from each paragraph in the file using a list comprehension in conjunction with doc.paragraphs. This will include scraping separate lines defined in the Word Document for listed items. Unlike docx2txt, docx, cannot scrape images from Word Documents. Also, docx will not scrape out hyperlinks and text in tables defined in the Word Document. Opaaaa fela