

You can use [python-docx2txt](#) library to read text from Microsoft Word documents. It is an improvement over [python-docx](#) library as it can, in addition, extract text from links, headers and footers. It can even extract images.