

Section 1: Obtaining and Understanding Text

Now that you are done with the videos of section 1, let's assess your learning. Here, are a few questions, followed by 4 options, out of which 1 is the correct option. Select the right option and validate your learning! The answers are provided in a separate sheet

Q1. How many file formats can Apache Tika parse?

- a. Only MS Office files
- b. Only web pages
- c. Over 1000 formats
- d. Only data exchanges: XML and JSON

Q2. In sentiment analysis, what is the advantage of language processing over word scoring?

- a. Word scoring is effective
- b. Words like 'not' flip related scores
- c. Language processing interprets metaphores
- d. Language processing understands more words