This document shows nested tables. On 2019-10-06 we discovered that an 80 kb DOCX file with nested tables was giving rise to 27 Mb of extracted text – clearly wrong.

Test this with:

crate\_test\_extract\_text --width 300 FILENAME > output.txt

There is whitespace at the end of this line. (CRATE may remove it.)

Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped. Here is a very long line that may be word-wrapped.

We start with a basic table.

|  |  |
| --- | --- |
| Row 1 col 1 | Row 1 col 2 |
| Row 2 col 1 | Row 2 col 2 |

Here’s a basic table with large content.

|  |  |
| --- | --- |
| Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 – Row 1 col 1 | Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 – Row 1 col 2 |
| Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 – Row 2 col 1 | Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 – Row 2 col 2 |

We move on to a singly nested table (two levels of tables):

|  |  |
| --- | --- |
| Level 1, row 1 col 1 | Level 1, row 1 col 2 |
| Level 1, row 2 col 1 | Level 1, row 2 col 2   |  |  | | --- | --- | | Level 2, row 1 col 1 | Level 2, row 1 col 2 | | Level 2, row 2 col 1 | Level 2, row 2 col 2 | |

Doubly nested (three levels):

|  |  |
| --- | --- |
| Level 1, row 1 col 1 | Level 1, row 1 col 2 |
| Level 1, row 2 col 1 | Level 1, row 2 col 2   |  |  | | --- | --- | | Level 2, row 1 col 1 | Level 2, row 1 col 2 | | Level 2, row 2 col 1 | Level 2, row 2 col 2   |  |  | | --- | --- | | Level 3, row 1 col 1 | Level 3, row 1 col 2 | | Level 3, row 2 col 1 | Level 3, row 2 col 2 | | |

All OK so far.

Now a large one:

|  |  |
| --- | --- |
| Level 1, row 1 col 1 – Level 1, row 1 col 1 – Level 1, row 1 col 1 – Level 1, row 1 col 1 – Level 1, row 1 col 1 – Level 1, row 1 col 1 | Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 – Level 1, row 1 col 2 |
| Level 1, row 2 col 1 – Level 1, row 2 col 1 – Level 1, row 2 col 1 – Level 1, row 2 col 1 – Level 1, row 2 col 1 – Level 1, row 2 col 1 – Level 1, row 2 col 1 – Level 1, row 2 col 1 – Level 1, row 2 col 1 | Level 1, row 2 col 2 – Level 1, row 2 col 2 – Level 1, row 2 col 2 – Level 1, row 2 col 2 – Level 1, row 2 col 2 – Level 1, row 2 col 2 – Level 1, row 2 col 2 – Level 1, row 2 col 2 – Level 1, row 2 col 2 – Level 1, row 2 col 2 – Level 1, row 2 col 2 – Level 1, row 2 col 2   |  |  | | --- | --- | | Level 2, row 1 col 1 – Level 2, row 1 col 1 – Level 2, row 1 col 1 – Level 2, row 1 col 1 – Level 2, row 1 col 1 | Level 2, row 1 col 2 – Level 2, row 1 col 2 – Level 2, row 1 col 2 – Level 2, row 1 col 2 – Level 2, row 1 col 2 – Level 2, row 1 col 2 – Level 2, row 1 col 2 – Level 2, row 1 col 2 – Level 2, row 1 col 2 – Level 2, row 1 col 2 – Level 2, row 1 col 2 – Level 2, row 1 col 2 – Level 2, row 1 col 2 | | Level 2, row 2 col 1 – Level 2, row 2 col 1 – Level 2, row 2 col 1 – Level 2, row 2 col 1 – Level 2, row 2 col 1 – Level 2, row 2 col 1 | Level 2, row 2 col 2 – Level 2, row 2 col 2 – Level 2, row 2 col 2 – Level 2, row 2 col 2 – Level 2, row 2 col 2 – Level 2, row 2 col 2 – Level 2, row 2 col 2 – Level 2, row 2 col 2   |  |  | | --- | --- | | Level 3, row 1 col 1 – Level 3, row 1 col 1 – Level 3, row 1 col 1 – Level 3, row 1 col 1 | Level 3, row 1 col 2 – Level 3, row 1 col 2 – Level 3, row 1 col 2 – Level 3, row 1 col 2 – Level 3, row 1 col 2 | | Level 3, row 2 col 1 – Level 3, row 2 col 1 – Level 3, row 2 col 1 – Level 3, row 2 col 1 – Level 3, row 2 col 1 | Level 3, row 2 col 2 – Level 3, row 2 col 2 – Level 3, row 2 col 2 – Level 3, row 2 col 2 – Level 3, row 2 col 2 | | |