6 Paragraph Elements Content

6.1 Basic Text Content

6.1.1 General

The paragraph elements `<text:p>` and `<text:h>` and their descendant elements contain the text content of any document. The *character content* of a paragraph consists of the character data of the paragraph element and the character data of its descendant elements concatenated in document order, with the following exceptions:

 Character data contained in the following elements or their descendant elements are not included in the character content of a paragraph:

- `<text:note-body>`. 6.3.4.

- `<text:ruby-text>`. 7.5.19

- `<office:annotation>`. 14.1

- Drawing shape and frame elements defined in sections 10.3, 10.4, 10.5 and 10.6.

Note: This should significantly ease transformations into other formats, since transformations need not interpret most of the descendant elements of a paragraph element to obtain the character content of a paragraph.

The phrase "document order" when used with reference to an OpenDocument document instance is defined by Section 5.0 Data Model [XPath].

6.1.2 White Space Characters

Consumers shall collapse white space characters that occur in

- a `<text:p>` or `<text:h>` element (so called paragraph elements), and

- in their descendant elements, if the OpenDocument schema permits the inclusion of character data for the element itself and all its ancestor elements up to the paragraph element.

Collapsing white space characters is defined by the following algorithm:

1) The following [UNICODE] characters are replaced by a " " (U+0020, SPACE) character:

    - HORIZONTAL TABULATION (U+0009)

    - CARRIAGE RETURN (U+000D)

    - LINE FEED (U+000A)

2) The character data of the paragraph element and of all descendant elements for which the OpenDocument schema permits the inclusion of character data for the element itself and all its ancestor elements up to the paragraph element, is concatenated in document order.

3) Leading " " (U+0020, SPACE) characters at the start of the resulting text and trailing SPACE characters at the end of the resulting text are removed.

4) Sequences of " " (U+0020, SPACE) characters are replaced by a single " " (U+0020, SPACE) character.

6.1.3 <text:s>

The `<text:s>` element is used to represent the [UNICODE] character " " (U+0020, SPACE).

This element shall be used to represent the second and all following " " (U+0020, SPACE) characters in a sequence of " " (U+0020, SPACE) characters.

Note: It is not an error if the character preceding the element is not a white space character, but it is good practice to use this element only for the second and all following " " (U+0020, SPACE) characters in a sequence.

The `<text:s>` element is usable within the following elements: `<text:a>` 6.1.8, `<text:h>` 5.1.2, `<text:meta>` 6.1.9, `<text:meta-field>` 7.5.19, `<text:p>` 5.1.3, `<text:ruby-base>` 6.4.2 and `<text:span>` 6.1.7.

The `<text:s>` element has the following attribute: `text:c` 19.763.

The `<text:s>` element has no child elements.

6.1.4 <text:tab>

The `<text:tab>` element represents the [UNICODE] tab character (HORIZONTAL TABULATION, U+0009). A `<text:tab>` element specifies that content immediately following it should begin at the next tab stop.

The `<text:tab>` element is usable within the following elements: `<text:a>` 6.1.8, `<text:h>` 5.1.2, `<text:meta>` 6.1.9, `<text:meta-field>` 7.5.19, `<text:p>` 5.1.3, `<text:ruby-base>` 6.4.2 and `<text:span>` 6.1.7.

The `<text:tab>` element has the following attribute: `text:tab-ref` 19.877.

The `<text:tab>` element has no child elements.

6.1.5 <text:line-break>

The `<text:line-break>` element represents a line break.

The `<text:line-break>` element is usable within the following elements: `<text:a>` 6.1.8, `<text:h>` 5.1.2, `<text:meta>` 6.1.9, `<text:meta-field>` 7.5.19, `<text:p>` 5.1.3, `<text:ruby-base>` 6.4.2 and `<text:span>` 6.1.7.

The `<text:line-break>` element has no attributes.

The `<text:line-break>` element has no child elements.

6.1.6 Soft Hyphens, Hyphens, and Non-breaking Spaces

Soft hyphens, hyphens, and non-breaking blanks are represented by [UNICODE] characters.

Table 8 - Unicode characters

The [UNICODE] character...	*Represents...*