11.3.2 NBH - No break here

A control function which indicates a point where no line break must occur when text is formatted (see § 12.2.1.3.2). Note - The graphic character NBSP (No Break SPace) should be taken to be equivalent to SPACE followed by NBH. PTX - Parallel texts

11.3.3 PTX - Parallel texts

A control function which delimits passages of text which are interchanged one after the other, but intended to be presented in parallel with one another (see § 5.2.5).

The only parameter values permitted are:

0 1 3

The default value of the parameter is 0.

PTX with parameter value 1 is the opening delimiter of the first (principal) of two passages of text intended to be presented in parallel with one another.

PTX with parameter value 3 is the closing delimiter of the first passage of text and the opening delimiter of the second (supplementary) passage intended to be presented in parallel with the first.

PTX with parameter value 0 indicates the end of the supplementary passage of text.

Note - Japanese Ruby permits the specification of exactly one supplementary passage of text.

11.4 Delimiters

11.4.1 SOS - Start of string

A control function that acts as the opening delimiter of a string of graphic characters and/or control functions that is marked to facilitate its removal by a subsequent content layout process. The string is closed by the terminating delimiter control function ST (String Terminator).

A string thus delimited may contain occurrences of graphic characters and control functions, in particular CR (Carriage Return), LF (Line Feed) and HYPHEN, introduced as a result of a formatting process (see § 12.2.1.3). 11.4.2 ST - String terminator

A control function that acts as the terminating delimiter of a string opened by the delimiter control function SOS (Start Of String).

11.5 SP - Space

A character with properties of both a graphic character and a logical control function.

As a control function, SP is significant to the content layout process. It acts as a word delimiter and indicates a potential line break point except when it is immediately followed by another SP or by an occurrence of the control function NBH (No Break Here) (see § 11.3.2).

As a graphic character, SP causes the active position to be advanced without a graphic symbol to be imaged. However, any graphic renditions that are in effect e.g. underlining, also apply to SP.

Any SP(s) that precede a line terminator, and follow the last graphic character of a line, are ignored by the imaging process.

12 Content layout process

This section describes a content layout process for basic logical objects associated with content architectures of type character.

Its purpose is to aid understanding of the semantics of the presentation attributes and control functions by describing the required results of such a process. However, it is not intended to specify any process that might be carried out in a particular implementation to achieve these results.

12.1 Introduction

12.1.1 Purpose

The content layout process describes a process of formatting and laying out character content into an allocated area. This area is referred to as the available area and is determined by the document layout process defined in Recommendation T.412.

The purpose of the content layout process is to convert content associated with basic logical components into content associated with basic layout objects. This might imply a transformation of the content from one form to another.

The content layout process results in the creation of a basic layout object(s) into which the content is to be positioned. The dimensions of each basic layout object is returned to the document layout process which determines the precise position of that basic layout object within the available area.

12.1.2 Available area

The content layout process is constrained by the available area. The maximum line length is constrained by the dimension of the available area in the direction of the character path.

During the layout of the content associated with a basic logical object into a basic layout object, the following cases can occur:

The formatted or formatted processable content fits into the available area.

- The formatted or formatted processable content does not fit into the dimensions of the available area in the direction of line progression. In this case, an additional or a new available area is required depending upon any constraints imposed by the document layout process.

- The formatted content does not fit into the dimension of the available area in the direction of the character path. In this case, a large available are is required.

Note - This case is most likely to occur when laying out formatted content associated with a basic logical object. 12.1.3 Presentation attributes

The content layout process takes into account the presentation attributes applying to the basic logical object with which the content is associated. It also takes into account any control functions that are embedded in the content.

The presentation attributes applying to the content layout process can be specified in the generic layout structure and presentation styles. The values of these presentation attributes are determined according to the defaulting rules specified in Recommendation T.412.

12.1.4 Character content architecture classes

The content layout process is described for basic logical objects associated with content that conforms to any of the three character content architecture classes (see § 4.1) as follows:

- processable form character content in which the content layout process provides for formatting of the content. The content layout process results in the output of content in formatted or formatted processable form depending upon the desired form of document;

- formatted processable form character content in which the content layout process provides for reformatting of the content. This involves an initialization process which must be carried out on the content before the content layout process can be applied to that content. The content layout process results in the output of content in formatted or formatted processable form depending upon the desired form of document;

- formatted form character content in which the content layout process has no effect on the content itself but still determines the dimensions of the basic layout object into which that content is to be positioned.

12.1.5 Use of delimiters

When formatted processable form content is created as a result of the content layout process, all shared control functions and graphic characters inserted into the content as a result of the content layout process are enclosed between the delimiter control functions SOS (Start Of String) and ST (String Terminator).

12.1.6 Layout of the content

For each of the three character content architecture classes, three cases of laying out the content of basic logical objects into basic layout objects are possible:

- single basic logical object to single basic layout object: the content of a single basic logical object can be laid out into a single basic layout object and is the only content associated with this basic layout object;

- single basic logical object to multiple basic layout object: the content of a single basic logical object is split among two or more basic layout objects, i.e. the content portions associated with two or more basic layout objects are derived from a single basic logical object;

- multiple basic logical object to single layout object: the content of two or more basic logical objects is laid out into a single basic layout object, i.e. the content portions of two or more basic logical objects are associated with a single basic layout object.

Multiple basic logical object to multiple basic layout object is also possible, but not described explicitly since this is a combination of the last two cases above.

12.1.7 Layout sequence

In all cases, the same sequence of steps for laying out content associated with a basic logical object is executed as follows:

- initialization;
- determination of initial point;
- formatting of the content;
- identification of content portions;
- determination of basic layout object dimensions;
- determination of the value of the presentation attribute "initial offset".

This also results in the creation of a basic layout object.

Although the sequence of steps is the same in all cases, the action performed at individual steps may vary.

- 12.2 Content layout process for processable content
- 12.2.1 Single basic logical object to single basic layout object
- 12.2.1.1 Initialization

In the case of processable form content, no initialization of the content is necessary.

12.2.1.2 Determination of initial point

The location of the initial point depends upon:

- the presentation attributes "character path" and "line progression" (determining the start edge and top edge of the basic layout object);

- the presentation attribute "kerning offset" (specifying the location of the start edge of the positioning area relative to the start edge of the basic layout object);

- the presentation attribute "indentation" (specifying the distance between the initial point and the start edge of the positioning area);

- the invocation of a font by the presentation attribute "graphic rendition" and the presentation attribute "character orientation" (specifying the minimum backward extent of the first line box);

- the presence of control functions PLU (Partial Line Up), PTX (Parallel Texts), VPB (Line Position Backward) and font invocation by SGR (Select Graphic Rendition) in the first line of characters to be imaged (modifying the backward extent of the first line box).

The position of the initial point relative to the start edge and top edge of the positioning area is determined such that:

- its distance from the top edge of the positioning area is equal to the backward extent of the first line box;

- its distance from the start edge of the positioning area is equal to the value specified by the presentation attribute "indentation".

The position of the initial point relative to the upper left corner of the basic layout object can be determined only after the dimension of the basic layout object has been determined (see § 12.2.1.5). This value is assigned to the presentation attribute "initial offset" and should always be specified explicitly in order to achieve the desired result by the content imaging process.

12.2.1.3 Formatting of the content

Formatting of the content involves:

- the positioning of character images within a line box (see § 5.2);
- the determination of line breaks;
- the positioning of line boxes within the basic layout (see \S 5.3).

It may involve the insertion of control functions and the assignment of presentation attribute values to the basic layout object.

As a general rule, the available area is utilized as much as possible in the direction of the character path.

12.2.1.3.1 Position of character images within a line box

There are two sets of operations that have an effect on the positioning of character images within a line box. One of these sets is related to the presentation attribute "formatting indicator", the other is not.

The operations related to the attribute "formatting indicator" are:

- alignment (presentation attribute "alignment");
- tabulation (presentation attribute "line layout table" and the control function STAB);
- first line offset (presentation attribute "first line offset");
- itemization (presentation attribute "itemization");
- pairwise kerning (presentation attribute "pairwise kerning").

Alignment and tabulation are mutually exclusive but itemization and first line offset may be applied in combination with either of them.

These operations may be performed by either the content layout process or the content imaging process. However, they can be performed by the content layout process only if the presentation attribute "formatting indicator" and the control functions HPB (Character Position Backward), HPR (Character Position Relative), SACS (Set Additional Character Separation), SRCS (Set Reduced Character Separation) and SSW (Set SPACE Width) are available in the content architecture level concerned.

The result of performing these operations by the content layout process is the insertion of the above control functions in the content.

If the content layout process has performed all of the specified operations for a basic layout object and inserted all necessary control functions, the value of the presentation attribute "reformatting indicator" is set to 'yes', otherwise it is set to 'no'.

Whether or not the content layout process inserts these control functions, it always determines the allocation of characters to each line box and the dimensions of each line box.

Note - If the value of the presentation attribute "formatting indicator" is 'no' or if a font substitution has been made, then the content imaging process should perform these operations.

The other set of operations which are not related to the presentation attribute "formatting indicator" are:

- character ordering (control function SRS);
- parallel annotation (control function PTX);
- graphic character composition (control function GCC).

12.2.1.3.1.1 Pairwise kerning

If the presentation attribute "pairwise kerning" specifies 'yes' and the content layout process is capable of performing this function and the font used provides the necessary information, then certain combinations of character images are positioned closer to (or further apart from) each other than determined by their position and escapement points.

When pairwise kerning is performed by the content layout process, it will result in the insertion of control functions HPB (Character Position Backward) or HPR (Character Position Relative) between the two characters involved.

12.2.1.3.1.2 First line offset

The characters associated with the first line can be controlled to be laid out differently from the rest of the lines in this object.

The presentation attribute "first line offset" specifies if the first line has an overhang or indentation relative to the line home position.

When first line offset is performed by the content layout process, it will result in the insertion of a control function HPB (Character Position Backward) or HPR (Character Position Relative).

12.2.1.3.1.3 Itemization

The first line of a basic layout object may contain an item identifier. The position of the item identifier is controlled by the presentation attribute "itemization".

When itemization is performed by the content layout process, it will result in the insertion of a control function HPB (Character Position Backward) or HPR (Character Position Relative) before the text of the item identifier. 12.2.1.3.1.4 Alignment

None of the alignments except 'start-aligned' can be performed until the dimensions of the basic layout object have been determined (see § 12.2.1.4).

When alignment is performed by the content layout process, the line length for alignment is determined to be:

- for the first line, the distance between the line home position and the end edge of the positioning area minus the value of the presentation attribute "first line offset";

- for all other lines, the distance between the line home position and the end edge of the positioning area. The various values of the presentation attribute "alignment" are treated as follows:

- 'start-aligned' does not result in the insertion of any control functions;

- 'end-aligned' and 'centred' result in the insertion of a control function HPR (Character Position Relative) either before the first graphic character of each line or after the CR (Carriage Return) delimiting the item identifier if the presentation attribute "itemization" specifies a value other than 'no itemization';

- 'justified' results in the insertion of zero, one or more control functions SSW (Set SPACE Width), SACS (Set Additional Character Separation) and/or SRCS (Set Reduced Character Separation) in each line which ends with a line terminator inserted by the content layout process. The precise usage of SACS SRCS and SSW is implementation dependent.

Note - The presentation attribute "Alignment" does not apply to the item identifier.

When the presentation attribute "alignment" has the value 'justified', irrespective of whether the alignment is performed by the content layout process or the content imaging process, the control function JFY (No Justify) is inserted at the beginning of the last line of a character sequence in order to avoid justification of this line by the content imaging process.

12.2.1.3.1.5 Tabulation

When tabulation is performed by the content layout process, it results in the insertion of a control function HPR (Character Position Relative) or HPB (Character Position Backward) between each occurrence of the control function STAB (Selective Tabulation) and the first graphic character following it.

12.2.1.3.1.6 Parallel annotation

The occurrence of the control function PTX (Parallel Texts) in the content specifies that a string of characters is to be laid out as a parallel annotation to another specified string of characters. The result of the content layout process is as described in § 5.2.5.

If the output of the content layout process is in formatted form, the positioning of the parallel annotation is achieved by the removal of all occurrences of PTX and the insertion of the appropriate control functions HPR (Character Position Relative), HPB (Character Position Backward), VPR (Line Position Relative) and VPB (Line Position Backward).

If the output of the content layout process is in formatted processable form, the positioning of the parallel annotation is achieved by inserting the control functions HPR (Character Position Relative), HPB (Character Position Backward), VPR (Line Position Relative) and VPB (Line Position Backward) and enclosing them by the delimiter control functions SOS (Start Of String) and ST (String Terminator).

12.2.1.3.1.7 Character ordering

The occurrence of the control function SRS (Start Reverse String) in the content of a basic logical object controls the direction of imaging of the interchanged characters. The result of the content layout process is as described in § 5.2.4.

12.2.1.3.1.8 Graphic character composition

The control function GCC (Graphic Character Composition) is used to combine two or more graphic characters into a single symbol. The width of these graphic characters may be less than the sum of the widths of the images of the component characters.

12.2.1.3.2 Insertion of line breaks

The formatting process may cause the insertion of additional line breaks into the content. Existing hard line terminators (CR/LF combinations) that are already in the content remain in the content but must be taken into account during the formatting process.

If the output is in formatted processable form, the inserted line breaks consist of soft line terminators represented by the control function sequence SOS CR LF ST. Optionally, characters inserted by an implementation dependent hyphenation algorithm are included in the SOS-ST string. No characters are deleted from the content.

If the output is in formatted form, the inserted line breaks consist of hard line terminators represented by the control function sequence CR LF. All occurrences of the control functions BPH (Break Permitted Here) and NBH (No Break Here) are deleted.

It is the intention of the formatting process that the number of characters between the inserted line breaks is the maximum possible for each line. the exact algorithm for inserting line breaks is implementation dependent and is not defined in the T.410 Series. However, the line break algorithm shall conform to the following constraints:

- a soft line break may by inserted:
- after a SP which is not immediately followed by another SP or the control function NBH (No Break

Here);

- after the control function BPH (Break Permitted Here);
- at a point determined by an implementation or language dependent algorithm;
- a line break is not permitted:
 - when a subscribed rendition is active;
 - within a string with reversed presentation direction;
 - within a string of parallel annotation.

12.2.1.3.3 Positioning of line boxes

The first line box is positioned with its line home position at the initial point as described in § 12.2.1.2.

Each line box is positioned with its line home position on the line from the initial point in the direction of line progression.

When proportional line spacing is not to be performed, the distance between the line home positions of two successive line boxes is equal to the current line spacing. The initial value of the current line spacing is the value of the presentation attribute "line spacing". This value may be altered by occurrences of the control function SVS (Select Line Spacing) and SLS (Set Line Spacing).

When proportional line spacing is to be performed by the content layout process, the distance between the line home positions of two successive line boxes is evaluated by an implementation dependent algorithm. If the evaluated distance differs from the current value of line spacing, the control function SLS (Set Line Spacing) is inserted before the line terminator of the first line. The evaluated distance is inserted as the parameter of this control function and also becomes the current line spacing.

If the output of the content layout process is in formatted processable form, those occurrences of the control function SLS (Set Line Spacing) inserted by the content layout process are enclosed by the delimiters SOS (Start Of String) and ST (String Terminator).

12.2.1.4 Identification of content portions

filled.

The content layout process shall also provide a value for the attribute "Content portion identifier - layout" for each content portion associated with the layout structure.

12.2.1.5 Determination of basic layout object dimensions

The formatting action continues until all the available content has been allocated or all the available area has been

The content layout process attempts to allocate the minimum portion of the available area sufficient to hold all of the

available content. Hence, the dimensions of the basic layout object can only be determined once formatting is complete. The mapping to horizontal and vertical dimensions of the basic layout object is dependent upon the character path as

follows:

| 1 | |
|---------|---|
| Charact | er path Dimension in direction Dimension in direction orthogonal of character path to character path |
| Γ | |
| | 0, 180 Horizontal dimension Vertical dimension |
| Γ | |
| 9 | 0, 270 Vertical dimension Horizontal dimension |
| | |

12.2.2 Single basic logical object to multiple basic layout objects

W

If the formatted content does not fit into the available area in the direction of line progression then an additional available area has to be obtained from the document layout process e.g. in the case of a page boundary. In this case, the content of a single basic logical object is allocated to more than one basic layout object. The original content portion is split into several content portions, each corresponding to a different basic layout object.

There are two changes with respect to the description in § 12.2.1:

- the presentation attributes of the second and subsequent basic layout object are given values corresponding to the status of rendition (line spacing, character spacing, graphic renditions etc.) at the end of the previous basic layout object;

- the allocation of content to basic layout objects is to be performed such that the presentation attributes "orphan size" and "widow size" are fulfilled.

12.2.3 Multiple basic logical objects to single basic layout object

When concatenation results in the content associated with a more than one basic logical object being laid out in a single basic layout object (see § 7.5), it may be necessary for the content layout process to insert control functions at the beginning of the second and subsequent basic logical objects so that the values of certain presentation attributes associated with those basic logical objects are applied. These control functions are:

- SHS (Select Character Spacing) or SCS (Set character Spacing) for "character spacing";
- Designation and/or invocation sequences for "graphic character sets";
- IGS (Identify Graphic Subrepertoire) for "graphic character subrepertoire";
- SGR (Select Graphic Rendition) for "graphic rendition";
- SVS (Select Line Spacing) or SLS (Set Line Spacing) for "line spacing".

If the output of the content layout process is in formatted processable form, then the control functions inserted by the content layout process are enclosed by the delimiters SOS (Start Of String) and ST (String Terminator).

The presentation attribute "proportional line spacing" specified for the second or subsequent basic logical objects is interpreted by the content layout process as described in § 12.2.1.

The other presentation attributes specified for the second and subsequent basic logical objects are ignored (see § 7.5).

12.3 Content layout process for formatted processable content

For formatted processable form content, the initialization step of the content layout process shall:

- remove all SOS-ST control strings from the content;
- remove all layout control functions (BS, HPB, HPR, JFY, SACS, SRCS and SSW) from the content (see

§ 11.2);

_

- combine all content portions associated with the same basic logical object into a single content portion in order to prevent unnecessary fragmentation of the document content which could otherwise occur;

delete the content portion attribute "content layout identifier - layout", if present.

After initialization, the content is in processable form. The remaining steps of the content layout process are as described for processable content (see § 12.2).

12.4 Content layout process for formatted content

For formatted content, the content layout process still has to determine the dimensions of the basic layout object to be allocated.

The same steps of the content layout process are used as for processable form content. In this case, however, formatting only involves:

- the positioning of character images within a line box as described in §§ 5.2 and 12.2.1.3.1;

- the positioning of line boxes within a basic layout object as described in § 5.3.

13 Content imaging process

This section describes a content imaging process for basic layout objects associated with content architecture of type character.

Its purpose is to aid understanding of the semantics of the presentation attributes and control functions by describing the required results of such a process. However, it is not intended to specify any process that might be carried out in a particular implementation to achieve these results.

13.1 Introduction

The content imaging process is only concerned with the logical structures, the presentation styles and the content of basic layout components conforming to this Recommendation.

All logical control functions, if any, are ignored.

The content imaging process is only applicable to the formatted and formatted processable form character content architecture classes.

13.2 Content imaging process for formatted content

This section describes how the various shared and layout presentation attributes and shared and layout control functions influence the image of the contents.

Most shared presentation attributes and shared control functions serve the purpose of positioning and orienting character images along reference lines and for positioning and orienting these reference lines within the basic layout object.

Thus, the effect of most shared presentation attributes and shared control functions have already been described in the content layout process.

Most layout presentation attributes and control functions are related to positioning and, thus, have already been described as the result of the content layout process.

The following sub-sections provide additional information relating to the content imaging process.

13.2.1 Determination of initial point

The active position for imaging is set on the initial point within the basic layout object. This information is derived from the presentation attribute "initial offset".

13.2.2 Choosing character images

The following presentation attributes and control functions determine the character images to be chosen for imaging:

- "graphic character sets" and code extension announcer, designation and invocation control functions;
- "character fonts" (together with the attribute "fonts list" in the document profile);
- "graphic rendition" and SGR (Select Graphic Rendition).

substitute for this font by making use of the font information available in the document profile attribute "fonts list".

13.2.3 Formatting indicator

The presentation attribute "formatting indicator" specifies whether first line offset, itemization, alignment, tabulation and pairwise kerning have already been performed by the content layout process or not.

If not, or if the result from the content layout process has been invalidated by a font substitution, then the content imaging process must perform the task in the same way as described in the content layout process (see § 12.2.1.3.1). 13.3 Content imaging process for formatted processable content

For content in formatted processable form, the only difference from the case of formatted form is that all logical control functions and the delimiter control functions SOS (Start Of String) and ST (String Terminator) are ignored. The effect of shared and layout presentation attributes and shared and layout control functions is as described in § 13.2.

14 Interactions between presentation attributes and control functions

This section contains a summary of the interactions among and between presentation attributes and control functions defined in various places in this Recommendation as follows:

LF is restricted to be used in the following cases:

- 1) at the beginning of the content of a basic layout component;
- 2) immediately following a control function CR (Carriage Return);
- 3) immediately following another LF.
- BPH or CR is not permitted in the following cases:
 - 1) when PLU or PLD is active;
 - 2) after VPR and/or VPB have moved the active position away from the reference line;
 - 3) between the control functions SRS 1 and SRS 0;
 - 4) between the control functions PTX 1 and PTX 0.

- Rendition aspects defined by some presentation attributes can be overridden by control functions embedded in content portions as follows:

W



- When the presentation attribute "line layout table" specifies any tabulation stops, the presentation attribute "alignment" is assumed to have the value "start-aligned".

15 Definition of character content architecture classes

This section defines the three classes of character content architecture as described in § 5, namely:

- a formatted form character content architecture which allows for document content to be presented (e.g. printed or displayed) as intended by the originator. Formatted form can be used in any basic component;

- a processable form character content architecture which allows for document content to be processed (e.g. edited or formatted). Processable form can be used in any basic logical component;

- a formatted processable form character content architecture which allows for document content to be processed and also to be presented as intended by the originator. Formatted processable form can be used in any basic component.

Paragraphs 15.1, 15.2 and 15.3 specify the categories of presentation attributes and control functions that pertain to these content architecture classes. The individual presentation attributes and control functions are summarized in Tables 5 and 6/T.416. The permissible values and the default values of the presentation attributes and the control functions parameter values are defined in §§ 7 and 11 respectively.

TABLE 5/T.416

Summary of presentation attributes



| L | Widow size | | Х | L | |
|---|------------|-------|---|---|--|
| | - | J | | | |

TABLE 6/T.416

Summary of control functions



In order to aid the definition of content architecture levels for use in document application profiles (see Recommendation T.411), the presentation attributes and the control functions that are applicable to each content architecture class are listed in Annex A, together with their permissible values and default values.

15.1 Formatted character content architecture class

The following categories of presentation attributes and control functions pertain to the formatted character content architecture class:

shared presentation attributes (see § 7.1);

-

- layout presentation attributes (see § 7.2);
- shared control functions (see § 11.1);
- layout control functions (see § 11.2).
- 15.2 Processable character content architecture class

The following categories of presentation attributes and control functions pertain to the processable character content architecture class:

- shared presentation attributes (see § 7.1);
- logical presentation attributes (see § 7.3);
- shared control functions (see § 11.1);
- logical control functions (see § 11.3).
- 15.3 Formatted processable character content architecture class

The following categories of presentation attributes and control functions pertain to the formatted processable character content architecture class:

- shared presentation attributes (see § 7.1);
- layout presentation attributes (see § 7.2);
- logical presentation attributes (see § 7.3);
- shared control functions (see § 11.1);
- layout control functions (see § 11.2);
- logical control functions (see § 11.3);
- delimiters (see § 11.4).