DOCUMENT APPLICATION PROFILE PM1 FOR THE INTERCHANGE OF PROCESSABLE FORM DOCUMENTS

CONTENTS

0 Introduction

The purpose of this Recommendation is to specify a document application profile referred to as PM1 that will support the transfer of documents containing character coded content only between word processors. This profile is defined in accordance with the T.410 Series of Recommendations.

This Recommendation contains two main sections. Section 5 defines the features that are supported by PM1 in terms of features commonly found on word processors as perceived by users. Section 6 then formally defines the document application profiles in accordance with Recommendation T.411. That is, it defines the document architecture and content architectures levels, and the corresponding allowable attributes and attribute values, that pertain to these profiles.

When using this Recommendation to encode a document, it is intended that the features in the document are

represented in terms of the features described in section 5 which can then be encoded in accordance with section 6.

This Recommendation does not define a precise mapping between the features in a particular document and the document architecture and content architecture features defined in section 6. Although this mapping will be obvious for most documents, in some cases a feature in a document may not have a precise equivalent in this profile. In this case, a feature in a particular document may have to be approximated, if possible, by a relate feature specified in section 5 of this profile.

The definition of these mappings is outside the scope of this Recommendation and is more appropriately defined in Recommendations specifying the characteristics of terminal equipment and the service aspects.

This Recommendation is intended to provide a means of encoding documents that can be used in any telematic service. It is independent of the means used to create, process, reproduce or transfer documents. These aspects may be specified in other Recommendations that make use of this Recommendation.

This version of PM1 provides for the representation and encoding of documents in which the text is laid out and read from left to right and from top to bottom on a page. That is, it provides for documents that contain latin based languages. It is intended to extend this Recommendation to provide for documents in which the text is written from top to bottom and from left to right on a page.

1 Scope

1.1 This Recommendation defines a document application profile conforming to the T.410 Series of Recommendations. This profile is referred to as PM1.

Its purpose is to specify interchange formats suitable for the transfer of documents between word processors. The profile caters for the transfer of documents such as memoranda, letters and reports that contain character only.

Documents can be transferred in either of the following forms:

- processable form, which facilitates the revision of documents by a recipient;

- formatted form, which facilitates the reproduction of documents as intended by a recipient;

- formatted processable form, which facilitate the reproduction of documents by a recipient as intended by an originator and facilitates the revision of documents by a recipient;

1.2 The features which can be interchanged using this application profile fall into the following categories:

a) page format features - these concern how the layout of each page of a document will appear when reproduced;

b) character content - these concern the character sets and control functions that make up the document content;

c) character content layout and imaging features - these concern how the document content will appear within the page of the reproduced document;

d) document management features - these concern the information associated with the document that relates to the document as a whole, such as its title, history and creation date; this information can be used in applications such as filing and retrieval.

1.3 It is assumed that when negotiation is performed by the service using this document application profile, that all non-basic features are subject to negotiations.

2 Field of application

2.1 The document application profile defined in this Recommendation is designed to be independent of the means to create or transfer the encoded documents.

2.2 This Recommendation defines a document application profile that may be used by any telematic service.

3 References

T.410 Series of Recommendations: "Open document architecture (ODA) and interchange format".

4 Definitions used in attribute tables

4.1 Definitions of terms

The terms defined in Recommendation T.411 are applicable to this Recommendation.

4.2 Notation used in attribute tables

The notation used in attribute applicability tables in this Recommendation is as follows:

The applicability of attributes for components is denoted by .../...; this represents: object class descriptions/object descriptions.

The symbol ... is then replaced by:

M mandatory attribute
NM non-mandatory attribute
D defaultable attribute
-- attribute is not applicable
(--- is equivalent to --/--).

In the tables defining allowable attribute values, the word 'any' means that any value is allowed subject to that value being a permissible value specified by the T.410 Series of Recommendations. A dash '-' in an attribute value table indicates that it is not applicable to specify a value for that entry. For example, it is not applicable to specify a default value for a non-mandatory attribute.

The presence of attributes in layout styles and presentation styles is denoted by the symbols:

O the attribute must always be present

X the attribute may be present

- the attribute is always absent

5 Characteristics supported by this document application profile

5.1 Overview

This section summarizes the processable and layout features which are supported by the document application profile defined by this Recommendation, in terms which are known by users of current word processors.

The logical and layout views of a document may be described in the same interchange format, in order to cope with the needs of different office automation applications (word processors, mail services, printing services, filing services, etc.).

Only character content may be used within the document.

5.2 Logical characteristics

From the logical point of view, the document content is divided into portions referred to as "paragraphs". Three types of "paragraphs" are distinguished, namely paragraphs corresponding to header, footer and body text. These types of paragraph are intended to be reproduced in the header, footer and body areas respectively of each page of the document, as

described in § 5.3.

"Paragraphs" corresponding to body text are arranged into groups, which mays contain any number of "paragraphs".

The division of the content into "paragraphs" provides the means to specify different layout and presentation requirements for individual or groups of successive "paragraphs".

The grouping of "paragraphs" allows different parts of the content of a document to be laid out in different sets of pages which have different layout format (as described in § 5.3).

The header and footer text also consists of a group of one or more "paragraphs". This allows different layout and presentation characteristics to be specified for different parts of the header and footer text.

Also, a document may consist of any number of such groups of header and footer "paragraphs". This allows different layout and presentation characteristics to be change within the document, as well as the layout and presentation of that content.

It is not guaranteed that the semantics of "paragraphs" and groups of "paragraphs" are the same for the originator and recipient.

5.3 Layout characteristics

5.3.1 The document layout structure

From the layout point of view, the document consists of one or more page sets. This allows sets of pages having different layout characteristics to be distinguished.

Each page set consists of a sequence of one or more pages, in accordance with one of the following formats:

a) a single page;

b) a sequence of two or more pages, all of which have the same layout characteristics;

c) a sequence of pages which are intended to be laid out alternatively on the 'recto' and 'verso' (see Note 1) side of a presentation medium; the layout characteristics of the 'recto' and 'verso' pages may be identical or different;

d) an initial page followed by a sequence of one of more pages such that the layout characteristics of the initial page is different from that of the subsequent pages; (Note - The initial page may have the same layout characteristics as the subsequent pages but may have different header and/or footer text);

e) an initial page followed by a sequence of recto-verso pages as described in c); the layout characteristics of the initial page may be (but is not necessarily) different from that of the 'recto/verso' pages.

The area made available within each nominal page (see Note 2) for the reproduction of the document content is called the text area. The text area has the same general characteristics for every page in the document and may consist of three independent and non-overlapping areas.

These consist of a header area lying at the top of the text area that is reserved for header text, a footer area lying at the bottom reserved for footer text and a body area lying between the header and footer areas that is reserved for body text. Either or both the areas reserved for header and footer text may not be present on each of the pages within a particular page set; however, each page of the document must have an area reserved for body text.

Note 1 - A 'recto' page is one that is imaged on the side of a sheet that is to be read first. A 'verso' page is imaged on the side of a sheet that is to be read second (see Recommendation T. T.412).

Note 2 - A nominal page is the ideal size of the presentation medium on which the document is reproduced, e.g the sheet of paper on which the content is to be imaged (see Recommendation T.412).

5.3.2 Page layout characteristics

5.3.2.1 The text area

The text area is the area made available for the positioning and display of the document content. It consists of three independent and non overlapping areas, as shown in Figures 1/T.502 and 2/T.502), namely:

- the header area (optional);

- the body area;
- the footer area (optional).

Text may be laid out only within these three areas. The header and footer areas may or may not be present within the text area; the body area must always be present.

Each text area is intended to be reproduced within a nominal page; the following nominal page sizes in both portrait and landscape orientations are supported:

- basic nominal pages: ISO A4 and NAL (North American Letter);
- non-basic nominal pages: ISO A3.

The sizes of these nominal pages are defined in Recommendation T.412.

FIGURE 1/T.502

Illustration of the text area (portrait orientation)

5.3.2.2 Size of the text area

The text area is specified in terms of its length (vertical dimension) and width (horizontal dimension).

This document application profile allows the size of the text area to be specified as equal to or smaller than the common assured reproduction area of ISO A4 and NAL (North American Letter) paper sizes. Larger areas, up to the size of the nominal paper size of ISO A3 may also be specified, but this is a non-basic feature.

When the nominal page is in portrait orientation, the positions of the header and footer areas are as illustrated in Figure 1/T.502.

When the nominal page is in landscape orientation, the positions of the header and footer areas are as illustrated in Figure 2/T.502.

The size of the text area supported by this document application profile is the common assured reproduction area of ISO A4 and NAL. Larger sizes are supported as a non-basic feature.

FIGURE 2/T.502

Illustration of the text area (landscape orientation)

5.3.2.3 Text area offset

The text area offset is the distance between the positions of the left and top edges of the text area and the left and top edges of the nominal page respectively (see Figures 1/T.502 and 2/T.502).

The value of this offset may vary for alternate pages in order to provide for the reproduction of pages in 'recto/verso' form.

5.3.2.4 Header area

The header area lies between the top edge of the text area and the top of the body area and is the area made available for header text. The dimensions and position of this area must be such that it does not extend beyond the edge of the text area in any direction or overlap the body area.

5.3.2.5 Body area

The body area is the area intended for the reproduction of the document content, apart from any header or footer content. It lies between the bottom of the header area and the top of the footer area. The dimensions and position of this area must be such that it does not extend beyond the edge of the text area in any direction or overlap either the header or footer areas.

5.3.2.6 Footer area

The footer area is an area which lies between the bottom of the body area and the bottom of the text area and is the area made available for footer text. The dimensions and position of this area must be such that it does not extend beyond the edge of the text area in any direction or overlap the body area.

5.4 Document layout features

This section defines the features associated with the presentation of text within the text area. Unless otherwise indicated, these properties can be changed anywhere in the document.

5.4.1 Layout of the document content

The successive paragraphs in a document can be laid out in the header, body and footer areas in a direction of 270 degrees relative to the positive horizontal direction of the page coordinate system (as defined in Recommendation T.412). This is illustrated in Figure 3/T.502.

FIGURE 3/T.502

Layout of the document content

5.4.2 Left and right margins

The left and right margins are the distances, or offsets, between a portion of the document content and an edge of the particular area in which that content is positioned. The margins specify the extents between which text is allowed to be positioned. Margins can be independently specified for the content in the header, footer and body areas and also they may vary throughout the document.

The left margin position is the first character position that is available on each line of text. This position is specified relative to the left edge of the area in which that content is positioned.

The right margin position is the maximum extent of each line of text. This position is also specified relative to the left edge of the area in which that content is positioned.

There is no restriction on the positions of these margins, provided that neither is set so that it exceeds the width of the area in which the text is positioned. Also, the position of the right margin must be equal to or to the right of the position of the left margin.

If the left and right margins are not explicitly specified then they are set to coincide with the position of the left and right edges, respectively, of the area in which the content is positioned.

5.4.3 Separation

This feature specifies the number of blank lines to be placed between one paragraph and the next text if the two paragraphs are on the same page.

If a value for the separation is not explicitly specified, then the next paragraph will be laid out directly on the line below the last line of the previous paragraph in accordance with the line spacing specified.

5.4.4 Page breaks

When the content associated with a section of a document is laid out, as many lines of text as possible will be placed in the body area of the current page before a new page is generated.

Because of this, page breaks can occur at inconvenient points within the text and hence a number of methods are provided to control the points at which page breaks can occur.

5.4.4.1 Unconditional page breaks

This feature specifies that an unconditional page break is required immediately. This is, the subsequent text must be displayed on the next page.

5.4.4.2 Conditional page breaks - widows and orphans

a) Widows and orphans

Widows and orphans control where page breaks may occur within the body of a paragraph.

The orphan size specifies the minimum number of lines of text in a paragraph that must be placed on the current page when a paragraph is split over two pages. If this minimum number cannot be accommodated, then the whole paragraph is to be placed on the next page.

The widow size specifies the minimum number of lines that must be allocated to the second page when a paragraph is split over tow pages. If, during the layout process, the number of lines of text on the second page is less than the value specified, then lines must be moved from the bottom of the first page to the top of the second page until the value is satisfied.

b) Indivisibility and association of paragraphs

This feature determines whether a single paragraph or a group of two or more paragraphs is allowed to be split over more than one page when the document content is laid out. This can be used, for example, to ensure that a section title is placed on the same page as the following text.

If the specified paragraph or paragraphs must be displayed within one page, then it may be necessary to cause a page break to occur if the current page has insufficient space to accommodate the specified paragraphs. In the absence of the specification of this feature, no restriction is placed on placing of successive paragraphs on successive pages.

5.4.4.3 Sheet breaks

This feature provides the ability to specify that the following text is to begin on a recto or on a verso page, irrespective of the type of page on which the immediately preceding text is laid out.

When a document is reproduced on paper, this may cause the generation of a new sheet of paper. This may occur anywhere in the document content.

5.5 Content layout and imaging characteristics

5.5.1 Character repertoires

The basic character repertoire that can be used in the subrepertoire of ISO 6937═2, corresponding to Recommendation T.61 (including non-spacing underline).

The coding of the character repertoire is that defined in Recommendation T.61 (or ISO 6937═2).

Any other registered graphic set can be used and are regarded as non-basic features (i.e their use must be indicated in the document profile).

5.5.2 Line spacing

This feature specifies the distance between successive lines of text.

The basic values are:

- 3 lines per 25.4 mm;
- 4 lines per 25.4 mm;

- 6 lines per 25.4 mm;
- 12 lines per 25.4 mm.

The following is a non-basic value:

- 8 lines per 25.4 mm.

The default is 6 lines per 25.4 mm.

5.5.3 Character spacing

This feature specifies the distance between successive characters on a line of text.

The basic value is:

- 10 characters per 25.4 mm.

The non-basic values are:

- 6 characters per 25.4 mm;
- 12 characters per 25.4 mm;
- 15 characters per 25.4 mm.

The default is 10 characters per 25.4 mm.

5.5.4 Character path and line progression

The character path is the direction of progression of successive characters along a line of characters. The lines progression is the direction of successive lines of text relative to the characters path direction.

The basic values are:

- character path: 0 degrees;
- line progression: 270 degrees.

There are no non-basic values.

5.5.5 Emphasis

This feature concerns the imaging of the graphic characters on the presentation medium.

The following basic modes of emphasis may be used:

- normal rendition;
- normal intensity;
- increased intensity (bold);
- italicized;
- not italicized;
- underlined;
- not underlined.

The following mode is non-basic:

- crossed out.

If a mode of emphasis is specified then it remains in effect until changed into a mutually exclusive mode or by the specification of 'normal rendition' (see below). Mutually exclusive modes are normal/increased intensity, italicized/not italicized and underlined/not underlined. One mode from each mutually exclusive set may be in operation at any point in the document content.

Normal rendition cancels the effect of all methods of emphasis that are currently in operation and specifies that the text should be displayed in accordance with the default rendition parameters set for the presentation device. Thus, if it is required to ensure that the content is not underlined, then the appropriate parameter value must be explicitly specified.

5.5.6 Tabulation

Tabulation stop positions can be specified at any character position along the character path. Each stop is specified by means of the following:

a) the tabulation position relative to the left margin position;

b) an optional alignment qualifier that specifies the type of alignment to be used at the designated tabulation position. The type can be as follows (see Figure 4/T.502):

i) start aligned - the first character is placed at the tabulation stop position;

ii) end aligned - the last character is placed at the tabulation position;

iii) centred - the character string is centred around the tabulation stop position;

iv) aligned on - the first character of a specified group of characters is placed at the tabulation stop position.

Only one set of tabulation stops can be specified to be applicable for a particular paragraph. No limit is placed on the number of tabulation stops that can be specified within a given set.

FIGURE 4/T.502

Examples of tabulations

5.5.7 Alignment

This feature specifies whether the text is to be aligned only at the left of each line of text, aligned only at the right of each line, centred or both left and right aligned (see Figure 5/T.502). If this feature is not specified, then the paragraph is assumed to be left aligned only.

Note - The value 'left aligned' means that the first character on each line is positioned at the indentation position. 'Right aligned' means that the content of each line is adjusted in position such that the last character on each line is placed adjacent to the margin position in the direction of the character path.

FIGURE 5/T.502

Examples of the use of 'alignment'

5.5.8 Indentation

Indentation is the distance between the first character on a line of text and the position of the margin position in the direction opposite to the direction of the character path.

Indentation acts as temporary alteration in the position of the offset in the direction opposite to the direction of the character path. When text is formatted, it is intended to be laid out between the indentation position and the right margin position (see Figure 3/T.502 for an example).

5.5.9 First line format

This feature specifies how the first line of a paragraph is to be laid out and provides for the itemization of paragraphs.

It allows the first character in the paragraph to be positioned at some points along the character path relative to to indentation position (as specified in § 5.5.8). This point may be in the direction of the character path or in the direction opposite to the direction of the character path relative to the indentation position.

In addition, this feature provides for the specification of an item identifier on the first line. The item identifier is a string of characters that precedes and is separated from the remaining characters that form the first line. The control function CR (carriage return) is used as the separator.

The features provided correspond to examples 10.1 to 10.4 shown in Figure 10/T.416.

5.5.10 Page numbering

This feature provides a means of indicating the number of each page of document.

The page number can be reset at the beginning of each page set. Also, the page number can be represented as a numeric character string, an alphabetic character string (lower or upper case) or as a roman numeral (lower or upper case).

These features allow, for example, the method of numbering the introduction or annexes of a document to be different than the method of page numbering used in the body of the document.

The page number can be used within a string of characters that is to laid out in the header or footer area. By this means, each page of a document can be automatically numbered when the document is laid out. These page numbers cannot be referenced in the body text.

An example of page numbering is "Page X" which consists of two concatenated character strings. The first is the literal character string 'Page' And this is concatenated to a string function denoted by 'X'. When 'X' is evaluated in a particular instance it may, for example, return the character string 'iv', the roman numeral (lower case) for the number '4'.

5.6 Document management features

A document profile is associated with every document to provide information about the document as a whole.

The features specified by the document profile are listed below. A definition of the information contained in these features is given in the corresponding attribute definitions in Recommendation T.414.

Presence of document constituents:

- generic layout structure;
- specific layout structure;
- generic logical structure;
- specific logical structure;
- layout styles;
- presentation styles.

Document characteristics:

- document application profile;
- document application profile default;
- document architecture class;
- content architecture class;
- interchange format class;
- ODA version date.

Non-basic document characteristics:

- profile character sets;
- comments character sets;
- page dimensions;
- medium types;
- layout path;
- coding attributes;
- presentation attributes;

Document management attributes:

- document reference.

Any other of the document management attributes defined in Recommendation T.414 may be specified.

The attributes that constitute "presence of document constituents" must be present when applicable (e.g if the document contains a specific layout structure then this must be indicated by this appropriate attribute).

The document characteristics attributes listed above are all mandatory.

The appropriate non-basic document characteristics attribute must be used when a non-basic feature is used within a document. The use all other feature listed above are non-mandatory.

6 Specification of the document application profiles

This section contains the technical specification for the document application profile PM1.

The notation used in the tables of attributes contained in this section is described in § 4.2. The allowable values of expressions are defined using the notation defined in Annex A of Recommendation T.412.

The unit scaling factor (see Recommendation T.412) used throughout PM1 is (1,1). Because of this, all dimensions and positions are specified in BMUs.

6.1 Summary of the technical specification

6.1.1 Overview

PM1 allows documents to be represented in the following forms:

- processable for, which facilitates the revision of a document by a recipient;

- formatted form, which facilitates the reproduction of a document as intended by the originator;

- formatted processable form, which facilitates the reproduction of a document as intented by the originator or facilitates the revision of a document.

6.1.2 Specification of constituents

This paragraph specifies the required and optional constituents used for the representation of documents that conform to PM1. Also, it specified the content architectures that may be present in these documents.

Constituents specified as 'required' must occur in any document that conforms to PM1. Constituents listed as 'optional' may or may not be present in the document depending upon the requirements of the particular document. The document profile indicated which constituents are present in the document.

6.1.2.1 Formatted form documents

6.1.2.1.1 Required constituents:

- a document profile as specified in § 6.5;

- layout object descriptions representing a specific layout structure as defined in § 6.3.2.

6.1.2.1.2 Optional constituents:

- layout object class description representing a 'partial' generic layout structure, as defined in § 6.3.1.2;

- presentation styles, as defined in § 6.4.4.2.

6.1.2.1.3 Content architecture

- the formatted character content architecture defined in § 6.4;
- the formatted processable character content architecture defined in § 6.4.

6.1.2.2 Processable form documents

6.1.2.2.1 Required constituents:

- a document profile as defined in § 6.5;
- logical object class descriptions representing a 'complete' generic logical structure, as defined in § 6.2.1;
- logical object descriptions representing a specific logical structure, as defined in § 6.2.2;
- layout object class description representing a 'complete' generic layout structure defined in § 6.3.1.1;
- layout styles as defined in § 6.2.4.

6.1.2.2.2 Optional constituents:

- presentation styles as defined in § 6.4.4.2.

6.1.2.2.3 Content architecture

- the processable form content architecture, as defined in § 6.4;
- the formatted processable form character content architecture, as defined in § 6.4.

6.1.2.3 Formatted processable form documents

6.1.2.3.1 Required constituents

- a document profile as defined in § 6.5;
- logical object class descriptions representing a 'complete' generic logical structure, as defined in § 6.2.1;
- logical object descriptions representing a specific logical structure, as defined in § 6.2.2;
- layout object class descriptions representing a 'complete' generic layout structure, as defined in § 6.3.1.1;
- layout object descriptions representing a specific layout structure, as defined in § 6.3.2;
- layout styles as defined in § 6.2.4.

6.1.2.3.2 Optional constituents

- presentation styles, as defined in § 6.4.4.2.

6.1.2.3.3 Content architectures

- the formatted character content architecture level defined in § 6.4;
- the processable content architecture level defined in § 6.4;
- the formatted processable character content architecture level defined in § 6.4.

Note 1 - The formatted character content architecture level may only be contained in content portions referenced by basic layout objects only.

Note 2 - The processable form content architecture can only be used in content portions associated with generic logical objects.

6.1.3 Interchange format

The interchange format class "A" is to be used in this application profile, as defined in Recommendation T.415.

6.1.4 Object identifiers

The ASN.1 object identifier value to be used to designate the document application profile PM1 is:

{0 0 20 502 0}

6.2 Logical structures

6.2.1 The generic logical structure

the generic logical structure is shown in Figure 6/T.502. It consists of two parts, namely a "body" part, which defines the allowable specific logical structures that may be used to represent the document, and the "header and footer" part, which specifies the header and footer text that may be used in the document.

The "body" part consists of:

- a single document logical root class;
- a single composite logical object class;
- a single basic logical object class.

Content portions and the attribute "content generators" may not be associated with the basic logical object class.

The "header and footer" part is optional and, if present, contains one or more of either or both of the following:

- a composite logical object class (called "header root") consisting of a sequence of one or more subordinate basic logical object classes named "header text";

- a composite logical object class (called "footer root") consisting of a sequence of one or more subordinate basic logical object classes named "footer text".

In each case, the basic logical object class must reference a single content portion or must contain the attribute "content generator".

Also, each basic logical object class of the types "header text" and "footer text" may be referenced by one or more composite logical object classes of the type "header root" or "footer root".

Note - Each logical object class of the type "header root" or "footer root" is referenced by an attribute "logical source" applied to a header or footer frame respectively that is defined in the generic layout structure. This causes the content associated with, for example, the header root to be laid out in each instance of the header frame that is generated during the document layout process.

FIGURE 6/T.502

Illustration of the "header and footer" part

6.2.2 The specific logical structure

The specific logical structure is controlled by the "body" part of the generic logical structure as defined in § 6.2.1.

As shown in Figure 6/T.502, the document logical root class specifies that the document logical root consists of a sequence of one or more composite logical objects.

Each of these composite logical objects consists of a sequence of one or more basic logical object referred to as "paragraphs".

Each paragraph may reference one or more content portions.

An example of a specific logical structure is illustrated in Figure 7/T.502.

FIGURE 7/T.502

Illustration of a specific logical structure

6.2.3 Attributes of logical components

6.2.3.1 Attributes applicable to logical component descriptions

Table 1/T.502 defines the attributes applicable to logical object classes for the "body" part of the generic logical structure and corresponding logical objects in the specific logical structure. Table 2/T.502 defines the attributes applicable to logical object classes in the "header and footer" part of the generic logical structure.