INTERNATIONAL STANDARD

ISO/IEC 8613-6

Second edition
1994-12-15

Information technology — Open Document Architecture (ODA) and Interchange Format: Character content architectures

Technologies de l'information — Architecture des documents ouverts (ODA) et format d'échange: Architecture des contenus de caractères

Reference number
ISO/IEC 8613-6:1994(E)

Contents

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 8613-6 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology,* Subcommittee SC 18, *Document processing and related communication,* in collaboration with ITU-T. The identical text is published as ITU-T Rec. T.416.

This second edition cancels and replaces the first edition (ISO 8613-6:1989), which has been technically revised.

ISO/IEC 8613 consists of the following parts, under the general title *Information technology — Open Document Architecture (ODA) and Interchange Format*:

- *Part 1: Introduction and general principles*
- *Part 2: Document structures*
- *Part 3: Abstract interface for the manipulation of ODA documents*
- *Part 4: Document profile*
- *Part 5: Open Document Interchange Format*
- *Part 6: Character content architectures*
- *Part 7: Raster graphics content architectures*
- *Part 8: Geometric graphics content architectures*
- *Part 9: Audio content architectures*
- *Part 10: Formal specifications*
- *Part 11: Tabular structures and tabular layout*
- *Part 12: Identification of document fragments*
- *Part 13: Spreadsheet*
- *Part 14: Temporal relationships and non-linear structures*

Annex A forms an integral part of this part of ISO/IEC 8613. Annexes B, C and D are for information only.

Introduction

This ITU-T Recommendation | International Standard was prepared as a joint publication by ITU-T Study Group VII and ISO/IEC Joint Technical Committee 1.

At present, the ITU-T Set of Recommendations in the T.410 series | International Standard ISO 8613 consists of:

- Introduction and general principles;
- Document structures;
- Document profile;
- Open document interchange formats;
- Character content architectures;
- Raster graphics content architectures;
- Geometric graphics content architectures;
- Formal specification of the Open Document Architecture (FODA)

  (The formal specification is applicable to ISO/IEC 8613 only).

Further Recommendations | International Standards may be added to this series of ITU-T Recommendations | International Standards.

Development of this set of ITU-T Recommendations | International Standards was originally in parallel with ECMA-101 standard: *Open document architecture*.

This series of ITU-T Recommendations | International Standards is a new edition of the CCITT T.410-Series of Recommendations (1998) and ISO 8613 (1989).

Significant technical changes are the inclusion of the following amendments as agreed by ITU-T and ISO/IEC:

- Alternative Representation;
- Annex on use of MHS/MOTIS;
- Colour;
- Conformance Testing Annex;
- Document Application Profile, Proforma and Notation;
- Security;
- Streams;
- Styles;
- Tiled Raster Graphics.

In addition, a number of technical corrigenda have been applied to this ITU-T Recommendation | International Standard.

This ITU-T Recommendation | International Standard contains four annexes:

- Annex A (integral): SGML representation of character content-specific attributes for ODL;
- Annex B (non-integral): Summary of content architecture classes Summary of content architecture classes;
- Annex C (non-integral): Coded representations of control functions;
- Annex D (non-integral): Summary of object identifiers.

This page intentionally left blank

INTERNATIONAL STANDARD

ITU-T RECOMMENDATION

INFORMATION TECHNOLOGY – OPEN DOCUMENT ARCHITECTURE (ODA) AND INTERCHANGE FORMAT: CHARACTER CONTENT ARCHITECTURES

1 Scope

The purpose of the ITU-T Rec. T.410-Series | ISO/IEC 8613 is to facilitate the interchange of documents.

In the context of these Recommendations | International Standards, documents are considered to be items such as memoranda, letters, invoices, forms and reports, which may include pictures and tabular material. The content elements used within the documents may include graphic characters, raster graphic elements and geometric graphic elements, all potentially within one document.

NOTE – These Recommendations | International Standards are designed to allow for extensions, including hypermedia features, spreadsheets and additional types of content such as audio and video.

In addition to the content types defined in these Recommendations | International Standards, ODA also provides for arbitrary content types to be included in documents.

These Recommendations | International Standards apply to the interchange of documents by means of data communications or the exchange of storage media.

These Recommendations | International Standards provide for the interchange of documents for either or both of the following purposes:

- to allow presentation as intended by the originator;
- to allow processing such as editing and reformatting.

The composition of a document in interchange can take several forms:

- formatted form, allowing presentation of the document;
- processable form, allowing processing of the document;
- formatted processable form, allowing both presentation and processing.

These Recommendations | International Standards also provide for the interchange of ODA information structures used for the processing of interchanged documents.

This ITU-T Recommendation | International Standard:

- defines a character content architecture that can be used in conjunction with the document architecture defined in ITU-T Rec. T.412 | ISO/IEC 8613-2;
- defines the internal structure of content conforming to this character content architecture;
- defines those aspects of rendition applicable to the presentation of character content;
- defines the presentation and content portion attributes applicable to this character content architecture;
- describes a character content layout process which, together with the document processing model described in ITU-T Rec. T.412 | ISO/IEC 8613-2, determines the layout of character content in basic layout objects.

2 Normative references

The following ITU-T Recommendations and International Standards contain provisions which, through references, in this text, constitute provisions of this Recommendation | International Standard. At the time of publication, the editions indicated were valid. All Recommendations and Standards are subject to revision, and parties to agreements based on this Recommendation | International Standard are encouraged to investigate the possibility of applying the most recent edition of the Recommendations and Standards listed below. Members of IEC and ISO maintain registers of currently valid International Standards. The ITU-T Secretariat maintains a list of currently valid ITU-T Recommendations.

2.1 Identical Recommendations | International Standards

- ITU-T Recommendation T.411 (1993) | ISO/IEC 8613-1:1994, *Information technology – Open Document Architecture (ODA) and Interchange Format: Introduction and general principles.*
- ITU-T Recommendation T.412 (1993) | ISO/IEC 8613-2:1994, *Information technology – Open Document Architecture (ODA) and Interchange Format: Document structures.*
- ITU-T Recommendation T.414 (1993) | ISO/IEC 8613-4:1994, *Information technology – Open Document Architecture (ODA) and Interchange Format: Document profile.*
- ITU-T Recommendation T.415 (1993) | ISO/IEC 8613-5:1994, *Information technology – Open Document Architecture (ODA) and Interchange Format: Open Document Interchange Format.*
- ITU-T Recommendation T.417 (1993) | ISO/IEC 8613-7:1994, *Information technology – Open Document Architecture (ODA) and Interchange Format: Raster graphics content architectures.*
- ITU-T Recommendation T.418 (1993) | ISO/IEC 8613-8:1994, *Information technology – Open Document Architecture (ODA) and Interchange Format: Geometric graphics content architectures.*

2.2 Paired Recommendations | International Standards equivalent in technical content

- CCITT Recommendation X.208 (1988), *Specification of Abstract Syntax Notation One (ASN.1)*;

  ISO/IEC 8824:1990, *Information technology – Open Systems Interconnection – Specification of Abstract Syntax Notation One (ASN.1).*

2.3 Additional references

- ISO 2022:1986, *Information processing – ISO 7-bit and 8-bit coded character sets – Code extension techniques.*
- ISO 2375:1985, *Data processing – Procedure for registration of escape sequences.*
- ISO/IEC 6429:1992, *Information technology – Control functions for coded character sets.*
- ISO 6937-1:1983, *Information processing – Coded character sets for text communication – Part 1: General introduction.*
- ISO 6937-2:1983, *Information processing – Coded character sets for text communication – Part 2: Latin alphabetic and non-alphabetic graphic characters.*
- ISO/IEC 7350:1991, *Information technology – Registration of repertoires of graphic characters from ISO 10367.*
- ISO/IEC 8613-10:1991, *Information technology – Text and office systems – Open Document Architecture (ODA) and interchange format – Part 10: Formal specifications.*
- ISO 8879:1986, *Information processing – Text and office systems – Standard Generalized Markup Language (SGML).*
- ISO/IEC 9541-1:1991, *Information technology – Font information interchange: Part 1 – Architecture.*

3 Definitions

For the purposes of this Recommendation | International Standard, the definitions given in ITU-T Rec. T.411 | ISO/IEC 8613-1 apply.

4 Abbreviations

BPH	Break Permitted Here
BS	Backspace
CR	Carriage Return
GCC	Graphic Character Composition
HPB	Character Position Backward
HPR	Character Position Relative
IGS	Identify Graphic Subrepertoire
JFY	No Justify
LF	Line Feed
NBH	No Break Here
PLD	Partial Line Up
PLU	Partial Line Down
PTX	Parallel Texts
SACS	Set Additional Character Separation
SCS	Set Character Spacing
SGR	Select Graphic Rendition
SHS	Select Character Spacing
SLS	Set Line Spacing
SOOS	Start Of Original String
SOS	Start Of String
SRCS	Set Reduced Character Separation
SRS	Start Reverse String
ST	String Terminator
SSW	Set Space Width
STAB	Selective Tabulation
SUB	Substitute Character
SVS	Select Line Spacing
VPB	Line Position Backward
VPR	Line Position Relative

5 Conventions

For the purpose of this Recommendation | International Standard, the conventions given in ITU-T Rec. T.411 | ISO/IEC 8613-1 apply.

6 General principles

6.1 Content architecture classes

This part of ITU-T Rec. T.410-Series | ISO/IEC 8613 defines three classes of character content architecture:

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

- A character content architecture for processable content which allows for document content to be processed (e.g. edited or formatted). Processable content can be used in any basic logical component.
- A character content architecture for formatted processable content which allows for document content to be processed and also to be presented as intended by the originator. Formatted processable content can be used in any basic component.

6.2 Content

The content of a basic component that conforms to a character content architecture is a character string. This character string is formed by concatenating the character strings in the content portions of the basic component.

The content character string consists of a combination of graphic characters, control functions and space characters.

6.3 Presentation attributes

Presentation attributes are applicable to basic logical and layout components. They contain information that specifies the initial conditions relating to the layout, the imaging and the selection of graphic characters of the content of these basic components. Some of these conditions can be changed by control functions contained within the content.

Presentation attributes are classified as follows:

- Logical presentation attributes which can be associated with processable and formatted processable character content. These attributes take effect during the content layout process but are ignored during the content imaging process.
- Layout presentation attributes which can be associated with formatted and formatted processable character content. These attributes take effect during the content imaging process. They are generated either by a content layout process or by a process that creates or edits the formatted or formatted processable content.
- Shared presentation attributes which can be associated with all character content architecture classes.

These attributes take effect during either or both of the content layout and imaging processes.

All presentation attributes in this Specification are defaultable.

6.4 Control functions

Control functions with zero or more parameters may specify information relating to the layout or imaging of subsequent graphic characters. A control function can also be used to extend or replace the set of graphic characters being used. The scope of all control functions is limited to the basic component in which they occur and, in the case of basic logical components, any basic logical component concatenated to this component (see 14.2.3).

Classification of control functions is similar to that of presentation attributes:

- Logical control functions which can be used in processable and formatted processable character content.

  These control functions take effect during the content layout process but are ignored during the content imaging process.
- Layout control functions which can be used in formatted and formatted processable character content.

  These control functions take effect during the content imaging process. They are generated by the content layout process. Alternatively, they may be inserted by a process (not described in ITU-T Rec. T.410-Series | ISO/IEC 8613) that creates or edits the formatted or formatted processable content.
- Shared control functions which can be used in all character content architecture classes. These control functions take effect during either or both of the content layout and imaging processes.

In addition, formatted processable content may contain control functions known as delimiters. These delimiters are used to indicate a string of one or more graphic characters and/or control functions that have been inserted or deleted as the result of a content layout process (see clause 14). The inserted graphic characters and/or control functions take effect only during the content imaging process. The delimiters take effect during the content layout process by deleting them and the inserted character sequence.

6.5 Graphic characters

The set of graphic characters used in the content of a basic component, and their coded representation, are specified by presentation attributes and code extension control functions (see clause 12 and 13.1.17).

Any set or sets of graphic characters may be used in the content of basic components, subject to the restrictions associated with the particular content architecture in use and subject to proper designation and invocation in accordance with ISO 2022.

Any non-spacing characters included in a graphic character set are not to be used in isolation but only in combination with spacing characters.

6.6 Space characters

The character SPACE (SP) is considered both as a logical control function and as a graphic character. As a graphic character, it has a graphical representation consisting of the absence of a graphic symbol. As a control function, it indicates a potential line break point (see 14.2.1.3.2).

NOTE – NBSP (No Break SPace) and any fixed-width space characters, such as "digit space", "em space" and "en space" are regarded as graphic characters i.e. are not regarded as line break points.

6.7 Coding of content information

The coded representation of the content information within a content portion is in accordance with the rules specified in ISO 2022. If ODIF is used as the interchange format, the coding shall be performed for an 8-bit environment.

NOTE 1 – This is equivalent to presuming a code extension announcer value of ESC 2/0 4/7 for the C1 set.

If ODL is used in the interchange format, the coding should where possible be performed for an 8-bit environment. (The use of ODL is applicable to ISO/IEC 8613 only.)

NOTE 2 – If an ODL application requires coding for a 7-bit environment, this should be specified by the document application profile, preferably also stating the appropriate code extension announcer(s).

Coded representations of control functions are defined in ISO 6429 and are summarized in Annex C.

6.8 Internal structure

6.8.1 Formatted content

Formatted content is content for which all the necessary information relating to the layout and imaging of that content has been specified. Content in this form is intended to be imaged as specified and is not intended to be revised by an editing process or to be reformatted.

The content of a basic component conforming to a formatted character content architecture consists of one or more lines of characters. Each pair of successive lines is separated by a hard line terminator. The end of the content of a basic component implicitly terminates the last line.

6.8.2 Processable content

Processable content is content which has not been laid out. Content in this form is suitable for revision by an editing process.

NOTE – The editing process is implementation dependent and is not described in ITU-T Rec. T.410-Series | ISO/IEC 8613.

In order to image content in this form, it is necessary to apply a content layout process (see clause 14) to the content which converts the processable content into formatted content (see 6.8.1) or into formatted processable content (see 6.8.3).

To assist the processing (i.e. editing or layout processes) of processable content, a number of logical presentation attributes and control functions have been defined (see clauses 9 and 13). In addition, the character SPACE is regarded both as a graphic character and as a control function that indicates where a line break may occur when the content is laid out.

The content of a basic component conforming to a processable character content architecture consists of one or more sequences of characters. Each pair of successive character sequences is separated by a hard line terminator control function.

At the end of the content of a basic logical component to which another basic logical component is concatenated (see ITU-T Rec. T.412 | ISO/IEC 8613-2), the last character sequence continues into the content of the next basic logical component. In all other cases, the end of the content of the basic logical component implicitly terminates the last character sequence.

The division into character sequences represents the internal structure of the processable content of a basic logical component. Each character sequence is anonymous, in that no name or identifier is associated with it, and no relationship exists among character sequences except that of sequence.

6.8.3 Formatted processable content

Formatted processable content is content that is structured such that it contains both the formatted content and the processable content as subsets. It is identical in structure to the processable content, except that it may contain additional control functions and graphic characters that have been added as a result of the content layout process. It is identical in structure to the formatted content, except that it may contain logical control functions and delimiters as well as deleted logical content.

Thus, formatted processable content can be converted to processable content by deleting (or ignoring) all layout control functions, all occurrences of the delimiters and all control functions and characters inserted within those delimiters.

Alternatively, formatted processable content can be converted to formatted content by deleting (or ignoring) all logical control functions and the delimiters as well as deleted content but retaining the control functions and characters inserted within the delimiters.

NOTE – The conversion of formatted processable content to processable content is a reversible process (providing the same layout constraints are applicable to the content layout process) but converting formatted processable content to formatted content is irreversible.

The formatted view of a basic component conforming to a formatted processable character content architecture consists of one or more lines of characters. Each pair of successive lines is separated by either a hard or soft line terminator. The end of the content of a basic layout component implicitly terminates the last line.

The processable view of a basic component conforming to a formatted processable character content architecture consists of one or more sequences of characters. Each pair of successive character sequences is separated by a hard line terminator.

At the end of the content of a basic logical component to which another basic logical component is concatenated (see ITU-T Rec. T.412 | ISO/IEC 8613-2), the last character sequence continues into the content of the next basic logical component. In all other cases, the end of the content of the basic logical component implicitly terminates the last character sequence.

Soft line terminators are used as separators between lines within a character sequence. The division into character sequences represents the internal structure of the content of a basic logical component. Each character sequence is anonymous, in that no name or identifier is associated with it, and no relationship exists among character sequences except that of sequence.

7 Character positioning

This clause specifies how characters are to be positioned within a basic layout object. The intention is to aid understanding of the presentation attributes and control functions that relate to character positioning.

This clause provides for the positioning of any font that is defined in accordance with ISO/IEC 9541-1. This clause also caters for the positioning of characters pertaining to different fonts within the same basic layout object.

7.1 Basic concepts

7.1.1 Character fonts

In the context of this part of ITU-T Rec. T.410-Series | ISO/IEC 8613, the term *graphic character* is used in its abstract sense; that is, this term refers to a member of a set of graphic symbols used for the representation of information. The term *character image* is then used to refer to the rendition of a graphic character on a presentation medium.

A *font* is a set of character images, normally with a common design and size. A set of font attributes is associated with the font as a whole and a set of character attributes is associated with each individual character. These attributes are defined in ISO/IEC 9541-1.

The main purpose of the font attributes is for the recipient to identify the font used by the originator and, in case the specified font is not available, the font and character attributes serve as guidance for the recipient to find an appropriate substitute font among those available.

Further information concerning the designation and invocation of different fonts within a basic object is given in clause 8.

7.1.2 Directions

In the context of this part of ITU-T Rec. T.410-Series | ISO/IEC 8613, all directions are expressed as counter clockwise angles of rotation (in degrees) relative to a specified reference direction (an example is given in Figure 1).

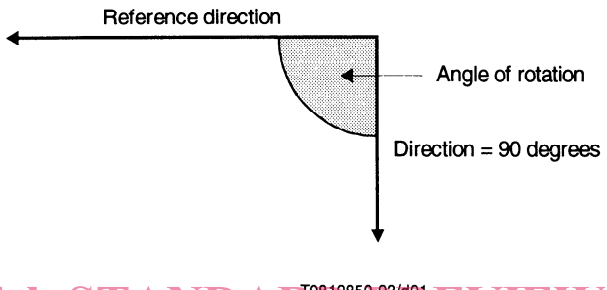

Figure 1 – Example of direction

The *character path* is the direction of progression of successive character images within a line box (defined in 7.1.7) and is expressed as a direction relative to the horizontal direction of the layout object (see Figure 4).

The *line progression* is the direction of progression of successive line boxes within the basic layout object and is expressed as a direction relative to the character path (see Figure 11).

The *character orientation* is the direction of the character baseline (defined in 7.1.3) relative to the character path.

Only one value for the character path, line progression and character orientation may be specified for a basic component.

7.1.3 Character image model

The *position point* is a reference point associated with a character image (see Figure 2). It is used for the positioning of the character image within a line box. The *escapement point* is a reference point associated with a character image (see Figure 2). It is used for the positioning of the next character image.

The *character baseline* is an imaginary line across a character image, for the purpose of defining the character orientation. The character baseline is a horizontal line when the character image is in its intended viewing orientation (see Figure 3).

A position point and escapement point shall be defined for each character orientation which is intended to be used (see Figure 3); i.e. "writing modes" corresponding to the required character orientations shall be defined in the font description, or fall-backs shall be defined in document application profiles.

A *kern* is that part of the character image that extends beyond its position and escapement points (see Figure 2).