Style Control of Structured Documents using DSSSL

Kyong Ho Lee,   Jin Ho Lee,   Yoon Chul Choy

Dept. of Computer Science,   Yonsei University

134   Shinchon-dong, Soedaemun-ku, Seoul, Korea

E-mail : lkh@rainbow.yonsei.ac.kr

Abstract

SGML(Standard Generalized Markup Language) is the ISO standard for describing the logical structure of documents and is also adopted as the CALS standard for document description. Since then, there have been growing interests in SGML application in a variety of fields. However because SGML doesn't provide a standard method for describing various processing informations, ie, formatting and transformation, most applications have applied methods that are system dependent. Recently, ISO defined DSSSL(Document Style Semantics and Specification Language) as a standard mechanism to specify the formatting, transformation and retrieval of structured documents.

Therefore, in this paper, we present a DSSSL processing system for style control of structured documents such as SGML documents. The system processes DSSSL style sheet that describes layout of documents and browses the result of its application to a SGML document. We have conducted tests on a lot of SGML documents and DSSSL style sheets successfully. Now, we are developing the SGML document management system that supports creation, editing, storage and retrieval of SGML document based upon the DSSSL processor and the SGML parser which we have developed.

Keywords : multimedia, hypertext, sgml, dsssl

1. Introduction

As it is obvious to everyone, the Information Superhighway has become a reality with rapidly growing development in network processing of information in computer science.   As a natural consequence, the movement to share the information through, for example, the Internet can be seen and felt in the every works of life.   Therefore, the need to develop standard format to develop and maintain hypertext or any other computer transmitted document has been expressed by the international society and organization.   The reason can be traced to the lack of compatibility between word processing program belonging to other systems and to the limited means of presentation of HTML (HyperText Markup Language).   As a response to the current issues, ISO has developed SGML [1] to facilitate the standardization of and better means for electrical transmission of documents, and, at this point, information systems which could provide support for the above system has been urgently needed.

Generally, SGML is used as a standard for describing a logical structure of document.   However, it does not provide any means to describe various handling of documents such as document layout,

presentation, transformation, or retrieval. The research paper[2] published previously suggests a means to describe a layout information of structured document. Due to the fact that the suggestions are dependent on specified systems, it is inadequate to use it for transmission of documents. Therefore, in order to describe various process of information in terms of structured document, DSSSL[3] has been established as the standard.

In this paper, an introduction to DSSSL along with design and implementation of DSSSL Processor will be discussed. The proposed processing system checks the syntax of the DSSSL style sheet which deals with the format and layout of structured document, and displays the result of application of DSSSL to SGML document. Presently, we are in the process of developing the document management system which supports creation, editing, storage, and retrieval of SGML using the SGML Parser whose development has been finished last year and DSSSL Processor whose discussion will follow in this paper.

The content of this paper is as follows. In the second section following the introduction, DSSSL will be introduced, in the third section the description of the structure of DSSSL Processor, in the fourth section the description of the current state of the art, future trend and possible direction of the development of the discussed system, and in the last section the conclusion.

2. What is DSSSL ?

There has been numerous publications concerning the introduction[4][5][6][7][8][9] and application [10][11][12][13] of SGML. However, an inadequate number of papers has been published which introduces DSSSL and its implementation. Therefore, it seems fitting to begin this paper with a short introduction to DSSSL and its special features.
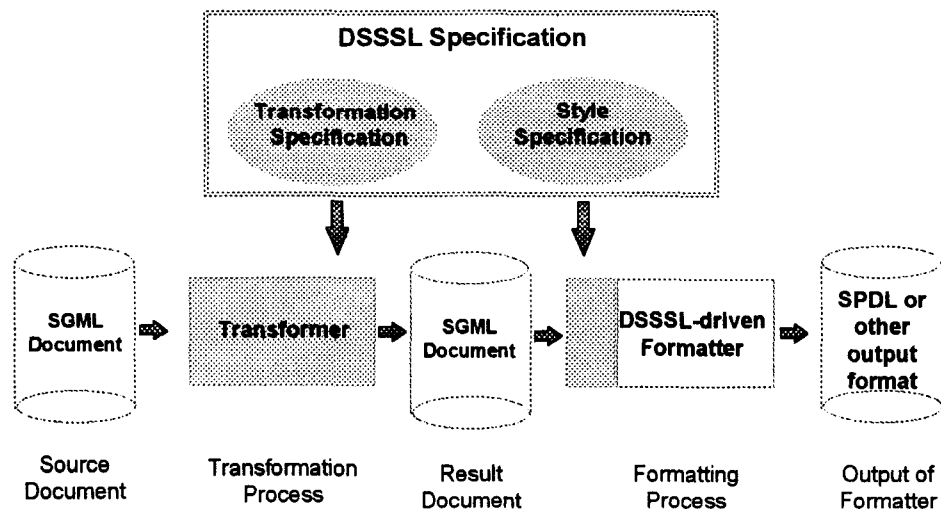

Figure 1. DSSSL conceptual model

2.1 DSSSL overview

DSSSL(Document Style Semantics and Specification Language : ISO 10179) is an international

standard for providing a standardized method for association of processing information with the markup of SGML documents. It is intended to be used along with documents structured as a hierarchy of elements in terms of formatting and other processing information to be associated with these elements to produce a formatted document and its subsequent presentation.

The DSSSL has two distinct processes: formatting process and transformation process. Figure 1 below shows the DSSSL conceptual model.

2.2 Components of DSSSL

The DSSSL consists of three languages: the style language(specification), the transformation language(specification), and the query language. Especially, the DSSSL expression language is a common subset of the transformation language, the style language and query language. It is based on the Scheme programming language defined in the IEEE Scheme standard[14], R^4RS. The sections below provide a brief overview of the main parts of DSSSL.

2.2.1 Style language

It provides the standard method for describing the information concerning formatting and layout of SGML documents. The process that applies formatting to SGML documents is called the formatting process. Figure 2 shows the formatting process of DSSSL.
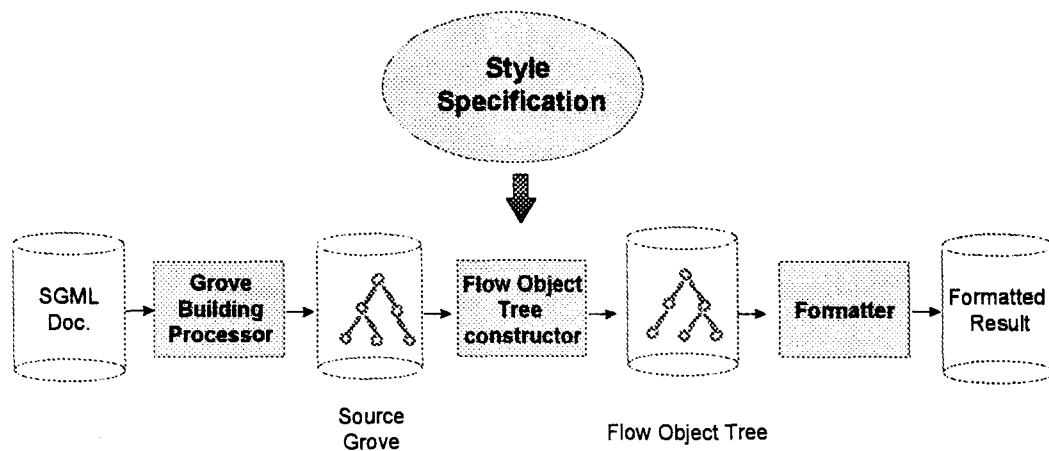

Figure 2. The formatting process

An SGML document is input to grove building process. The SGML document is parsed and is represented by a collection of nodes called a grove. DSSSL is based on the document model which describes a document as a grove. A grove is similar to an element tree, but may include other subtrees, for example, a subtree of attribute values. The grove is then further processed to create a flow object tree.

Flow objects provide a standard mechanism for describing the layout of a document. They represent layout constructs such as page sequences, paragraphs, hyperlinks and pictures. Flow objects have

characteristics such as a page sequence's margins, a paragraph's font size, a hyperlink's destination and a picture's height and width.

2.2.2 Transformation language

A transformation language transforms SGML document marked up according to one DTD(document type definition) into another. For example, it transforms TEI(Text Encoding Initiative) document into HTML document. A transformation language contains a list of associations. An association contains up to three parts: the query expressions, the transform expressions, and optional priority expressions.

This language allows the user to specify the creation of new structures, the replication of existing structures, and the reordering and regrouping of existing structures. Figure 3 shows the transformation process of DSSSL.
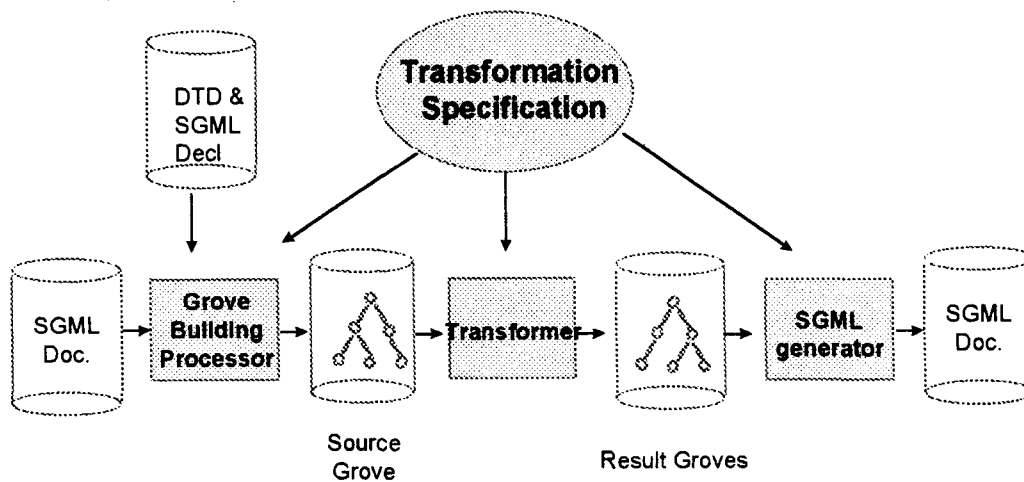

Figure 3. The transformation process

2.2.3 SDQL(standard document query language)

SDQL is part of both the transformation language and the style language. It is used for navigating through the hierarchical structure of the SGML document, and identifying the relevant pieces of the SGML markup and content on which processing is to be performed, that is, selecting and returning nodes of grove.

3. The structure of DSSSL Processor

In this section, we present the DSSSL processor which processes the layout information of the structured document and the formatter which enables an user to view structured document whose layout has been processed by DSSSL processor. They are C++ based systems that are available for Win32 platforms. The proposed overview of the system and its structure is as follows.

3.1 Overview

The proposed system of process by this research can be largely divided into three: The first is the input of DSSSL style sheet, and SGML grove as the result of the parsing of SGML document, the second is the application of style sheet to the SGML grove which results in the creation of Flow object tree, and the final third stage is the viewing of the Flow object tree on screen. The second stage can be sub-divided into expression language, style language, query language and transformation language interpreting modules. Figure 4 shows the overall structure of the DSSSL processor.
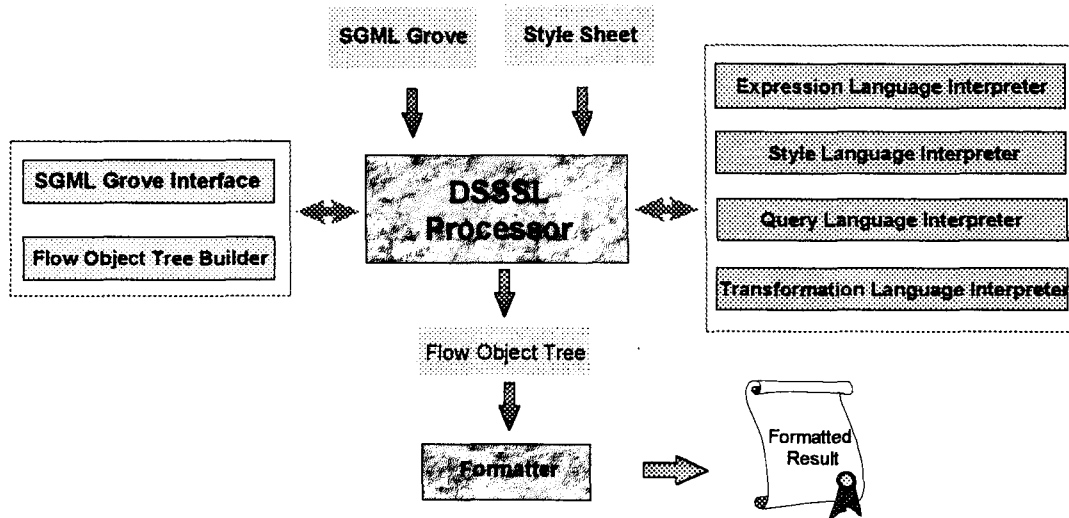
Figure 4. The structure of the system

3.2 Components of the system

The detailed description of the proposed structure of the system is as follows.

3.2.1 SGML grove interface

SGML grove interface should produce a tree structure grove as the result of parsing of SGML document due to the fact that DSSSL is based on a document model called grove. Therefore, the interface provides creation of grove and an access to it. Also, in order to check the syntax of the SGML document, the parser already developed by our research team was used.

Figure 5. Block diagram of Expression Language interpretation module

3.2.2 Expression language interpreting module

As it has been stated already, expression language is a common subset of style language and transformation language. Considering the importance of the language, the development of module to interpret the language is essential. In this research we have undertook this much needed study to develop t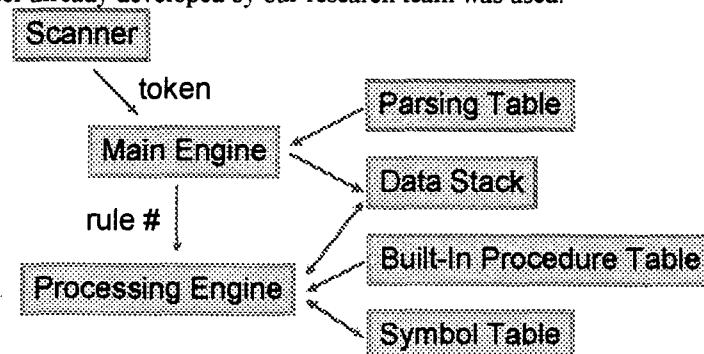he module which supports full set of language. We have used LALR(1) Parsing Table in the implementation of Expression Language Interpreting Module, and its structure described in Figure 5.

3.2.3 Style Language Interpreting Module

Style Language is the one that describes format and layout of SGML document. It, especially, supports the detailed presentation of the document. Therefore, the style language interpreting module developed by our research team is able to process basic information concerning the layout to more complicated one such as color, graphic, table, and mathematics equations. The following Figure 6 displays an example of SGML and DSSSL style sheet, and Figure 7 shows the result of SGML document after applying layout information.
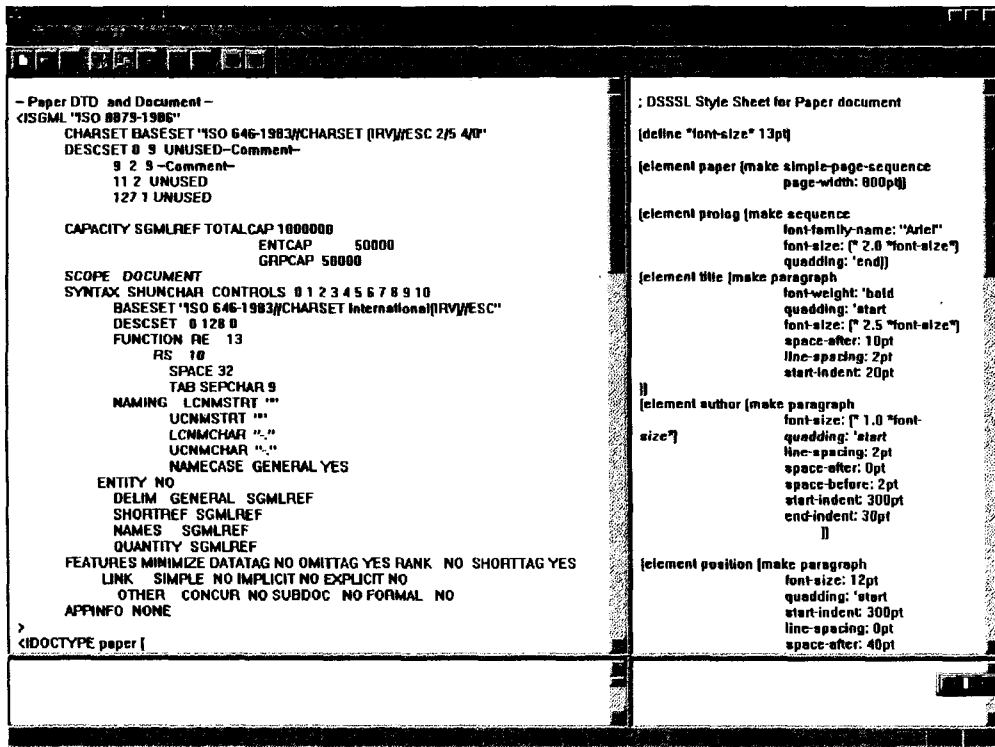

Figure 6. SGML document and DSSSL style sheet

3.2.4 Flow Object Tree Builder

DSSSL defines the basic mechanism of Flow object for the representation of SGML document; thus, we provide a builder which creates the flow object tree. The flow object tree is a combination of layout information and SGML grove.

3.2.5 Formatter

We have developed a Formatter which displays the result of application of DSSSL style sheet to SGML document directly on screen. In another words, it enables one to browse the flow object tree. Figure 7 shows an example of the Formatter's capability to view Flow Object Tree.
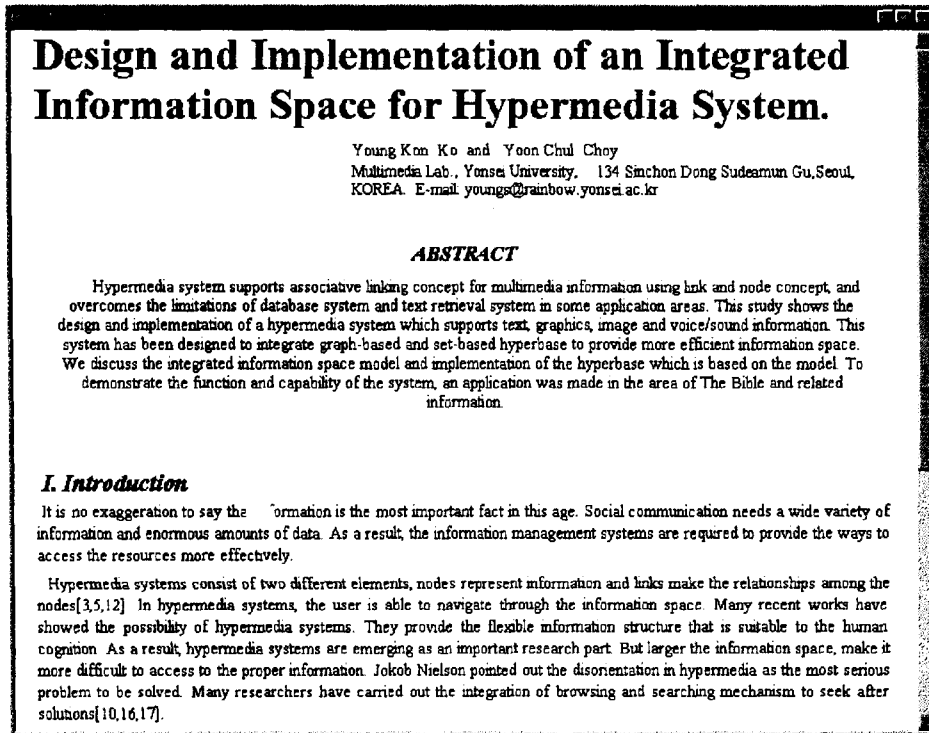

Design and Implementation of an Integrated Information Space for Hypermedia System.

Young Kon Ko and Yoon Chul Choy
Multimedia Lab., Yonsei University. 134 Sinchon Dong Sudaemun Gu.Seoul
KOREA E-mail: youngs@rainbow.yonsei.ac.kr

ABSTRACT

Hypermedia system supports associative linking concept for multimedia information using link and node concept, and overcomes the limitations of database system and text retrieval system in some application areas. This study shows the design and implementation of a hypermedia system which supports text, graphics, image and voice/sound information. This system has been designed to integrate graph-based and set-based hyperbase to provide more efficient information space. We discuss the integrated information space model and implementation of the hyperbase which is based on the model. To demonstrate the function and capability of the system, an application was made in the area of The Bible and related information.

I. Introduction

It is no exaggeration to say the "ormation is the most important fact in this age. Social communication needs a wide variety of information and enormous amounts of data. As a result, the information management systems are required to provide the ways to access the resources more effectively.

Hypermedia systems consist of two different elements, nodes represent information and links make the relationships among the nodes[3,5,12] In hypermedia systems, the user is able to navigate through the information space. Many recent works have showed the possibility of hypermedia systems. They provide the flexible information structure that is suitable to the human cognition. As a result, hypermedia systems are emerging as an important research part. But larger the information space, make it more difficult to access to the proper information. Jokob Nielson pointed out the disorientation in hypermedia as the most serious problem to be solved. Many researchers have carried out the integration of browsing and searching mechanism to seek after solutions[10,16,17].

Figure 7. The result of presentation

4. The State of the Art

DSSSL Processor developed and explained in this paper is, by no means, completed. DSSSL is able to process various features such as presentation, transformation, and retrieval of document. We have been more inclined to process a style language that describes the presentation of document. At this point, DSSSL Processor is able to process the full set of expression language and also the subset of style language and query language. In the case of the formatter, it displays the flow object tree, a combination of layout information and SGML grove.

We have plans to further develop the formatter to display of the flow object tree not only on screen but also as SPDL(Standard Page Description Langauge)[15] and Transformation Language Interpreter. In addition, we are in the process of developing the SGML Document Management System based on the SGML Parser and DSSSL Processor.

5. Conclusion

This paper dealt with the DSSSL Processor which can process the layout of SGML documents. To

re-iterate the main functions of the DSSSL Processor are as follows. First, it processes the full set of expression language which is a common subset of DSSSL style language and Transformation Language. Second, it supports SGML Grove which is the DSSSL Document Model. Third, it supports SDQL in order to query the structured document. Fourth, it produces Flow Object Tree which results from applying DSSSL Style Sheet to SGML document. lastly, it supports the Formatter, SGML Browser, which displays the Flow Object Tree.

Although the development of our DSSSL Processor is not fully functional in terms of all its possibilities, it has been proven that it is not restricted to specified SGML Document or DSSSL Style Sheet after a numerous number of tests we have conducted on several different types of document.

Reference

[1] ISO 8879, Information processing -- Text and office systems -- Standard Generalized Markup Language (SGML), ISO, 1986

[2] Yaron Wolfsthal, "Style Control in the Quill Document Editing System," Software-Practice and Experience, Vol. 21, No. 6, pp. 625- 638, June, 1991

[3] ISO/IEC 10179, Information technology - Text and office system - Document Style Semantics and Specification Language(DSSSL), ISO, 1996

[4] Martin Bryan, SGML - an author's guide to the standard generalized markup language, Addison-Wesley, 1988

[5] Eric Van Herwijnen, Practical SGML Second Edition, Kluwer Academic Publishers, 1995

[6] Heather Brown, "Standards for structured documents," The Computer Journal, Vol. 32, No. 6, pp. 505-514, 1989

[7] H. Wright, "SGML Free Information," BYTE, Vol. 17, No. 6, pp. 279-286, 1992

[8] David Stern, "SGML Document: A better system for communicating knowledge," Special Libraries, Vol. 86, Iss. 2, pp. 117-124, Spring, 1995

[9] Wood, E. M., "So what is SGML?, " Journal of Systems Management, Vol 46, Iss. 5, pp.24-29, September/October, 1995

[10] Weiss, E. H., "Of document databases, SGML, and rhetorical neutrality," IEEE Transactions on Professional Communications, Vol. 36, Iss. 2, pp. 58-61, June, 1993

[11] Sacks-Davis R., Arnold-Moore T., Zobel J.,"Database systems for structured documents," IEICE Transactions on Information and System, Vol. E78-D, Iss. 11, pp.1335-1342, November, 1995

[12] V.Quint, C.Roisin, I.Vatton, "A structured authoring environment for the World-Wide Web," Computer Networks and ISDN Systems, Vol. 27, Iss. 6, pp. 831-840, April, 1995

[13] Paoli J., "Cooperative work on the network : edit the WWW!," Computer Networks and ISDN Systems, Vol. 27, Iss. 6, pp. 841-847, April, 1995

[14] IEEE Standard 1178, IEEE Standard for the Scheme Programming Language, IEEE, 1991.

[15] ISO/IEC 10180, Information technology - Text Communication - Standard Page Description Language (SPDL), ISO, 1995