UDC: 007:389.6
Review article

A REFERENCE MODEL FOR DATA INTERCHANGE STANDARDS

Mirko Varga

University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia
E-mail: mvarga@foi.hr

This paper discusses data interchange standards. It introduces a reference model in which the relevant issues are studied on several levels and in a layered manner: 1) the atom level, 2) the elementary level, 3) the structural level and 4) the application level. Each of these levels is specific and is subject to a variety of standards. The heterogeneous nature of the media (text, graphics, sound, video, animation, etc.) involves a multitude of corresponding formats that bring out the need to establish and apply appropriate standards in order to facilitate data interchange. WWW demands a standard that will guarantee format stability and data compatibility.

Keywords: reference model, data standards.

1. INTRODUCTION

WWW is based on the following assumption - all documents are represented in a standard format characterised by a structure defined by HTML (Hypertext Markup Language). A standard scheme is used for document identification, along with a standard search form and HTTP protocol (hypertext transfer protocol).

In 1989 the file transfer protocol (FTP) was developed in order to use it as a standard for access to remote data. The FTP repertoire was not developed enough and proved to be too slow for an optimal utilization of the Web. In view of these shortcomings, Hypertext Transfer Protocol (HTTP) was designed and meant that quicker operations with hypertext links were made possible. HTTP allows the client to specify the language and selection of data formats. This feature, called the mediator format, is a key element of the independence between the HTTP and the HTML specifications.

HTML, URL (Uniform Resource Locator) and HTTP standards are simplifications of several more comprehensive standards that serve as a basis for SGML (Standard Generalised Markup Language), DSSSL (Document Style Semantic and Specification Language), Hytime and CCL (Common Command Language). Each of these standards has a great impact on the interchange of electronic information.

SGML represents a standard designed by the International Organization for Information Structuring Standards (ISO-standard 8879). This standard identifies and codes document elements (title, heading, paragraphs, etc.). The user can define the rules and relations of data structuring. In a SGML document, data is saved in a neutral form (American Standard Code for Information Interchange) which makes it possible

for it to be used on various computer platforms and in a variety of operating systems. The familiar programs (*Acrobat, Framemaker, WordPerfect, etc.*) in the Windows and Macintosh versions include the SGML format and their corresponding auxiliary tools.

This development, however promising, indicates that greater attention should be paid to the standards for data interchange. For example, WWW requires a standard that will guarantee format stability and data compatibility. Rapid evolution from the origins of the HTML to its version 2.0 and later to the HTML 3.0, which has a transitional and extended adaptability variation, has caused quite a lot of instability. The question now is whether standardization of data transfer is possible, bearing in mind the rapidly changing technology.

A significant response to this is represented by the World Wide Web Consortium's (W3C's) development of the eXtensible Markup Language (XML) for logical structuring and Cascading Style Sheets (CSS) for presentation as a balance between the author and the reader. XML is basically a dialect of SGML, but is specially intended for use on the Web.

As a publishing method, XML/CSS is comparable to using text processing software with styles (LaTeX) , style templates or macros.

The World Wide Web Consortium wants the Web to be useful for everyone. To solve this problem, W3C has initiated a Resource Description Framework (RDF) specification. RDF will provide an interoperable and uniform means of exchanging the metadata between programs and across the Web. There are a huge number of Web publishing methods, for example MS Word, PostScript, PDF, AutoCad, etc.

2. A MULTITUDE OF FORMATS

The fundamental concept on data interchange standardization says that N various formats require $N \bullet (N-1)$ converters for the information to be transferred from one format to another.

The most important formats include [1]:

Area	*Number of formats*
Data bases	5
Graphics	37
Text processing	6
Page-description	3
Table calculations	10
Sound	12
Windows domain	11

Table 1 *Number of formats in relation to the area of application*

Common standards need to resolve the question of formats in data interchange as well as the operations on them. Text documents are usually saved in the following formats:

ASCII, LaTeX, HTML, SGML and PostScript. The sound and video areas demand a lot more.

Furthermore, along with the formats, the operations on data must also be standardised. For example, the SGML syntax is specific and strict while its operational semantics are generally insufficiently defined. Significant breakthroughs are to be expected on the semantic level, as is the case with the DBMS operations [3].

Common standards must resolve the question of formats in data interchange as well as operations on them. The reference model helps to specify the standards for data entities.

3. THE REFERENCE MODEL

The reference model has to state the problems and identify the areas of standardization. The most important areas are:

- Layering
- Standard range
- Separating entity representation from the operations
- Identification of meta-standards

3.1. Layering

In modular standard elaboration, the effect of change in a standard may be isolated. Layering, i.e. the elaboration of a standard on several levels, is widely used in relating standards, e.g. in the case of the Windows system. The reference model uses four layers:

- the atom layer
- the element layer
- the structure layer
- the application layer

The atom layer represents the basic data form known as UNICODE, ASCII and RGB triplets. On this level we concentrate on the nature of the record, i.e. the medium, e.g. text, image, audio, language, video record, animation, etc. The most widely used is the ASCII code (American Standard Code for Information Interchange) which usually saves data in 8 bits so that a byte can represent 256 values (2^8) which is more than enough for the representation of ASCII and non-ASCII symbols.

The element layer involves formats for saving data entities on the level of the atom. The element data level supposes the data bit to be higher than the atom layer, making it possible to interpret the element structure.

The structure layer provides two additional types of information. On this level we can distinguish between various types of atom structures structured together, as

well as specify structural and relational dependencies among the elements. For example, in the SGML language we can include text and graphics and specify their dependence. The standards on this level include XDR (External Data Representations) and SGML.

The application layer adds specific professional information from a given branch to structural standards which can frequently be found under the abbreviations CALS (Computer-Aided Acquisition and Logistics Support) and TEI (Text Encoding Initiative).

3.2. Standard Range

The reference model also needs to define the range, i.e. the scope for each standard. On the atom level, the range of definition guarantees that all suitable data forms will be processed. On higher levels of complex structure and applications, the standard range is likely to be an "open document architecture".

3.3. Separating Representation From Operations

The data interchange standards call attention to entity representations. More recent standards, such as SGML, distinguish between these two activities. The role of the reference model is to identify representation standards as well as the standards that define possible operations on entities. This involves standardization and categorization of operations into the following groups [5]:

1. Create
2. Display
3. Save, and
4. Search

Let's illustrate this with an example of a text document.

Media type	*Text*
	Representation
	ASCII
	ISO
	Marking the text
	Structuring the text
	Hypertext
	Operations
	Symbol operations
	String operations
	Editing
	Formatting
	Searching
	Sorting
	Compression
	Encryption
	Specific language operations

Figure 1 *Separating operations from representation in a text*

Text representation is viewed separately from the operations on the text. Only major operations on text documents are stated. For a more detailed overview of text processing, see [5].

3.4. Identification of Meta-Standards

Some data interchange standards are so-called *meta-standards* and they represent a skeleton from which derived standards can develop. The reference model needs to find ways to relate the derived standards and the meta-standards on which they are based.

4. CONCLUSION

The reference model discussed here provides a mode for classification and organization of data interchange standards. The reference model requires additional clarification and extension.

The model may be extended, i.e. visualised as a cube. A three-dimensional approach includes three model parameters:

1. different levels
2. standard range, and
3. representation, providing suitable operations on objects.

The Web requires Renaissance. HTML as a simple, scalable, document format that can be used for information exchange on virtually any platform is momentarily in version 4 and further versions are envisioned.

Media independent publishing is actually a much greater problem than data exchange. The key to understanding the revolutionary potential of XML is that it is just one piece of a larger picture. XML by itself can provide standardized interchange formats for databases and spreadsheets. Standards are the body and soul of the Web.

REFERENCES

[1] G. Born. *The File Formats Handbook.* Int'l Thomson Computer Press, London, 1995.

[2] C. Cargill. *Information Technology Standardization: Theory, Process and Organizations*, Digital Press, Bedford, Mass, 1989.

[3] D. Raymond, F. Tompa , D. Wood. From data representation to data model: metasemantic issues in the evolution of SGML". *Standards and Interfaces*, Jan. 1996, pp. 25-36.

[4] H. Sleurink. *The Multimedia Dictionary*, Academic Press, London, 1995.

[5] M. Spring, T. Carlo-Bearman, Information standards: models for future development, *Book Research Quarterly*, Fall 1988, pp. 38-47.

Received: 16 November 1998
Accepted: 2 February 1999

Mirko Varga

REFERENTNI MODEL ZA STANDARDE RAZMJENE PODATAKA

Sažetak

U radu se razmatra problematika standarda za razmjenu podataka. Uvodi se referentni model koji navedenu problematiku proučava slojevito na sljedećim razinama: 1) atomska, 2) elementarna, 3) strukturna te 4) aplikacijska. Svaka od navedenih razina je specifična i podliježe različitim standardima. Raznovrsnost medija (tekst, grafika, slike, zvuk, video, animacija itd.) uključuje mnoštvo pripadajućih formata čime se potencira utvrđivanje i primjena odgovarajućih standarda u cilju osiguravanja kompatibilnosti te lakše razmjene podataka. WWW zahtijeva standard koji jamči stabilnost i kompatibilnost podataka.

Ključne riječi: referentni model, standardi podataka.