

Hypertext Markup Language - 2.0

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Abstract

The Hypertext Markup Language (HTML) is a simple markup language used to create hypertext documents that are platform independent. HTML documents are SGML documents with generic semantics that are appropriate for representing information from a wide range of domains. HTML markup can represent hypertext news, mail, documentation, and hypermedia; menus of options; database query results; simple structured documents with in-lined graphics; and hypertext views of existing bodies of information.

HTML has been in use by the World Wide Web (WWW) global information initiative since 1990. This specification roughly corresponds to the capabilities of HTML in common use prior to June 1994. HTML is an application of ISO Standard 8879:1986 Information Processing Text and Office Systems; Standard Generalized Markup Language (SGML).

The "text/html" Internet Media Type (RFC 1590) and MIME Content Type (RFC 1521) is defined by this specification.

Table of Contents

1.	Introduction	2
1.1	Scope	3
1.2	Conformance	3
2.	Terms	6
3.	HTML as an Application of SGML	10
3.1	SGML Documents	10
3.2	HTML Lexical Syntax	12
3.3	HTML Public Text Identifiers	17
3.4	Example HTML Document	17
4.	HTML as an Internet Media Type	18

4.1	text/html media type	18
4.2	HTML Document Representation	19
5.	Document Structure	20
5.1	Document Element: HTML	21
5.2	Head: HEAD	21
5.3	Body: BODY	24
5.4	Headings: H1 ... H6	24
5.5	Block Structuring Elements	25
5.6	List Elements	28
5.7	Phrase Markup	30
5.8	Line Break: BR	34
5.9	Horizontal Rule: HR	34
5.10	Image: IMG	34
6.	Characters, Words, and Paragraphs	35
6.1	The HTML Document Character Set	36
7.	Hyperlinks	36
7.1	Accessing Resources	37
7.2	Activation of Hyperlinks	38
7.3	Simultaneous Presentation of Image Resources	38
7.4	Fragment Identifiers	38
7.5	Queries and Indexes	39
7.6	Image Maps	39
8.	Forms	40
8.1	Form Elements	40
8.2	Form Submission	45
9.	HTML Public Text	49
9.1	HTML DTD	49
9.2	Strict HTML DTD	61
9.3	Level 1 HTML DTD	62
9.4	Strict Level 1 HTML DTD	63
9.5	SGML Declaration for HTML	64
9.6	Sample SGML Open Entity Catalog for HTML	65
9.7	Character Entity Sets	66
10.	Security Considerations	69
11.	References	69
12.	Acknowledgments	71
12.1	Authors' Addresses	71
13.	The HTML Coded Character Set	72
14.	Proposed Entities	75

1. Introduction

The HyperText Markup Language (HTML) is a simple data format used to create hypertext documents that are portable from one platform to another. HTML documents are SGML documents with generic semantics that are appropriate for representing information from a wide range of domains.

As HTML is an application of SGML, this specification assumes a working knowledge of [SGML].

1.1. Scope

HTML has been in use by the World-Wide Web (WWW) global information initiative since 1990. Previously, informal documentation on HTML has been available from a number of sources on the Internet. This specification brings together, clarifies, and formalizes a set of features that roughly corresponds to the capabilities of HTML in common use prior to June 1994. A number of new features to HTML are being proposed and experimented in the Internet community.

This document thus defines a HTML 2.0 (to distinguish it from the previous informal specifications). Future (generally upwardly compatible) versions of HTML with new features will be released with higher version numbers.

HTML is an application of ISO Standard 8879:1986, "Information Processing Text and Office Systems; Standard Generalized Markup Language" (SGML). The HTML Document Type Definition (DTD) is a formal definition of the HTML syntax in terms of SGML.

This specification also defines HTML as an Internet Media Type[IMEDIA] and MIME Content Type[MIME] called 'text/html'. As such, it defines the semantics of the HTML syntax and how that syntax should be interpreted by user agents.

1.2. Conformance

This specification governs the syntax of HTML documents and aspects of the behavior of HTML user agents.

1.2.1. Documents

A document is a conforming HTML document if:

- * It is a conforming SGML document, and it conforms to the HTML DTD (see 9.1, "HTML DTD").

NOTE - There are a number of syntactic idioms that are not supported or are supported inconsistently in some historical user agent implementations. These idioms are identified in notes like this throughout this specification.

- * It conforms to the application conventions in this specification. For example, the value of the HREF attribute

of the <A> element must conform to the URI syntax.

* Its document character set includes [ISO-8859-1] and agrees with [ISO-10646]; that is, each code position listed in 13, "The HTML Coded Character Set" is included, and each code position in the document character set is mapped to the same character as [ISO-10646] designates for that code position.

NOTE - The document character set is somewhat independent of the character encoding scheme used to represent a document. For example, the 'ISO-2022-JP' character encoding scheme can be used for HTML documents, since its repertoire is a subset of the [ISO-10646] repertoire. The critical distinction is that numeric character references agree with [ISO-10646] regardless of how the document is encoded.

1.2.2. Feature Test Entities

The HTML DTD defines a standard HTML document type and several variations, by way of feature test entities. Feature test entities are declarations in the HTML DTD that control the inclusion or exclusion of portions of the DTD.

HTML.Recommended

Certain features of the language are necessary for compatibility with widespread usage, but they may compromise the structural integrity of a document. This feature test entity selects a more prescriptive document type definition that eliminates those features. It is set to 'IGNORE' by default.

For example, in order to preserve the structure of a document, an editing user agent may translate HTML documents to the recommended subset, or it may require that the documents be in the recommended subset for import.

HTML.Deprecated

Certain features of the language are necessary for compatibility with earlier versions of the specification, but they tend to be used and implemented inconsistently, and their use is deprecated. This feature test entity enables a document type definition that allows these features. It is set to 'INCLUDE' by default.

Documents generated by translation software or editing software should not contain deprecated idioms.

1.2.3. User Agents

An HTML user agent conforms to this specification if:

- * It parses the characters of an HTML document into data characters and markup according to [SGML].

NOTE - In the interest of robustness and extensibility, there are a number of widely deployed conventions for handling non-conforming documents. See 4.2.1, "Undeclared Markup Error Handling" for details.

- * It supports the 'ISO-8859-1' character encoding scheme and processes each character in the ISO Latin Alphabet No. 1 as specified in 6.1, "The HTML Document Character Set".

NOTE - To support non-western writing systems, HTML user agents are encouraged to support 'ISO-10646-UCS-2' or similar character encoding schemes and as much of the character repertoire of [ISO-10646] as is practical.

- * It behaves identically for documents whose parsed token sequences are identical.

For example, comments and the whitespace in tags disappear during tokenization, and hence they do not influence the behavior of conforming user agents.

- * It allows the user to traverse (or at least attempt to traverse, resources permitting) all hyperlinks from <A> elements in an HTML document.

An HTML user agent is a level 2 user agent if, additionally:

- * It allows the user to express all form field values specified in an HTML document and to (attempt to) submit the values as requests to information services.

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.