WikipediA

Uniform Resource Identifier

A **Uniform Resource Identifier** (**URI**) is a string that provides a unique address (either on the Internet or a local Intranet) where a resource can be found. It looks and is similar to the more commonly known <u>Uniform Resource Locator</u> (URL) that is used to locate web sites and pages within those sites. A URL is a kind of URI. I.e., all URLs are <u>URIs</u> but not all <u>URIs</u> are <u>URLs</u>. URLs typically identify web pages that are meant to be viewed in a browser. URIs identify any type of resource, often meant to be used by a computer but not to be viewed in a browser. URIs often identify resources that are part of an ontology defined using the <u>Web Ontology Language</u> (OWL). For example, the class Agent or the property email in the <u>Friend of a Friend vocabulary</u> would each have an individual <u>URI</u>. URIs have mostly been replaced by <u>Internationalized Resource Identifiers</u> (IRI) which are identical to <u>URIs</u> except IRIs can handle character sets such as Kanji rather than being restricted to ASCII as <u>URIs</u> are. [1] [2]

Uniform Resource Identifier (URI)

Domain	World Wide Web
Abbreviation	URI

Contents

History

Conception Refinement

Design

URLs and URNs

Syntax

URI references

Resolution

Relation to XML namespaces

See also

Notes

References

Further reading

External links

History

Conception

URIs and URLs have a shared history. In 1994, <u>Tim Berners-Lee</u>'s proposals for <u>hypertext</u> implicitly introduced the idea of a URL as a short string representing a resource that is the target of a hyperlink. At the time, people referred to it as a "hypertext name" or "document name".

Over the next three and a half years, as the World Wide Web's core technologies of HTML, HTTP, and web browsers developed, a need to distinguish a string that provided an address for a resource from a string that merely named a resource emerged. Although not yet formally defined, the term *Uniform Resource Locator* came to represent the former, and the more contentious *Uniform Resource Name* came to represent the latter.

During the debate over defining URLs and URNs, it became evident that the concepts embodied by the two (2) terms were merely aspects of the fundamental, overarching, notion of resource *identification*. In June 1994, the <u>IETF</u> published <u>Berners-Lee</u>'s the first *Request for Comments* that acknowledged the existence of URLs and URNs. Most importantly, it defined a formal syntax for *Universal Resource Identifiers* (i.e. URL-like strings whose precise syntaxes and semantics depended on their schemes). In addition, the <u>RFC</u> (<u>RFC 1630</u>) attempted to summarize the syntaxes of URL schemes in use at the time. It acknowledged -- *but did not standardize* -- the existence of relative URLs and fragment identifiers. [5]

Refinement

In December 1994, IETF RFC 1738 formally defined relative and absolute URLs, refined the general URL syntax, defined how to resolve relative URLs to absolute form, and better enumerated the URL schemes then in use. [6] The agreed definition and syntax of URNs had to wait until the publication of IETF RFC 2141[7] in May 1997.

The publication of IETF RFC 2396^[8] in August 1998 saw the URI syntax become a separate specification^[9] and most of the parts of RFCs 1630 and 1738 relating to URIs and URLs in general were revised and expanded by the IETF. The new RFC changed the meaning of "U" in "URI" to "Uniform" from "Universal".

In December 1999, IETF RFC $2732^{[10]}$ provided a minor update to RFC 2396, allowing URIs to accommodate $\underline{\text{IPv6}}$ addresses. A number of shortcomings discovered in the two specifications led to a community effort, coordinated by RFC 2396 co-author Roy Fielding, that culminated in the publication of IETF RFC 3986 in January 2005. While obsoleting the prior standard, it did not render the details of existing URL schemes obsolete; RFC 1738 continues to govern such schemes except where otherwise superseded. IETF RFC 2616 for example, refines the http scheme. Simultaneously, the IETF published the content of RFC 3986 as the full standard STD 66, reflecting the establishment of the URI generic syntax as an official Internet protocol.

In 2001, the W3C's Technical Architecture Group (TAG) published a guide to <u>best practices</u> and canonical URIs for publishing multiple versions of a given resource. [13] For example, content might differ by language or by size to adjust for capacity or settings of the device used to access that content.

In August 2002, IETF RFC 3305^[14] pointed out that the term "URL" had, despite widespread public use, faded into near obsolescence, and serves only as a reminder that some URIs act as addresses by having schemes implying network accessibility, regardless of any such actual use. As URI-based standards such as Resource Description Framework make evident, resource identification need not suggest the retrieval of resource representations over the Internet, nor need they imply network-based resources at all.

The Semantic Web uses the HTTP URI scheme to identify both documents and concepts in the real world, a distinction which has caused confusion as to how to distinguish the two. The TAG published an e-mail in 2005 on how to solve the problem, which became known as the httpRange-14 resolution. The W3C subsequently published an Interest Group Note titled Cool URIs for the Semantic Web, which explained the use of content negotiation and the HTTP 303 response code for redirections in more detail.

Design

URLs and URNs

A <u>Uniform Resource Name</u> (URN) is a URI that identifies a resource by name in a particular namespace. A URN may be used to talk about a resource without implying its location or how to access it. For example, in the <u>International Standard Book Number</u> (ISBN) system, *ISBN 0-486-27557-4* identifies a specific edition of Shakespeare's play *Romeo and Juliet*. The URN for that edition would be *urn:isbn:o-486-27557-4*. However, it gives no information as to where to find a copy of that book.

A <u>Uniform Resource Locator</u> (URL) is a URI that specifies the means of acting upon or obtaining the representation of a resource, i.e. specifying both its primary access mechanism and network location. For example, the URL http://example.org/wiki/Main_Page refers to a resource identified as /wiki/Main_Page, whose representation, in the form of HTML and related code, is obtainable via the Hypertext Transfer Protocol (http:) from a network host whose domain name is example.org.

A URN may be compared to a person's name, while a URL may be compared to their street address. In other words, a URN identifies an item and a URL provides a method for finding it.

Technical publications, especially standards produced by the $\underline{\text{IETF}}$ and by the $\underline{\text{W3C}}$, normally reflect a view outlined in a $\underline{\text{W3C}}$ Recommendation of 2001, which acknowledges the precedence of the term URI rather than endorsing any formal subdivision into URL and URN. {{cquote|URL is a useful but informal concept: a URL is a type of URI that identifies a resource via a representation of its primary access mechanism (e.g., its network "location"), rather than by some other attributes it may have. $\underline{^{[17]}}$

As such, a URL is simply a URI that happens to point to a resource over a network. [a][18] However, in non-technical contexts and in software for the World Wide Web, the term "URL" remains widely used. Additionally, the term "web address" (which has no formal definition) often occurs in non-technical publications as a synonym for a URI that uses the *http* or *https* schemes. Such assumptions can lead to confusion, for example, in the case of XML namespaces that have a visual similarity to resolvable URIs.

Specifications produced by the WHATWG prefer URL over URI, and so newer HTML5 APIs use URL over URI. [19]

Standardize on the term URL. URI and IRI [Internationalized Resource Identifier] are just confusing. In practice a single algorithm is used for both so keeping them distinct is not helping anyone. URL also easily wins the search result popularity contest. [20]

While most URI schemes were originally designed to be used with a particular <u>protocol</u>, and often have the same name, they are semantically different from protocols. For example, the scheme *http* is generally used for interacting with web resources using HTTP, but the scheme *file* has no protocol.

Syntax

Each URI begins with a scheme name that refers to a specification for assigning identifiers within that scheme. As such, the URI syntax is a federated and extensible naming system wherein each scheme's specification may further restrict the syntax and semantics of identifiers using that scheme. The URI generic syntax is a superset of the syntax of all URI schemes. It was first defined in RFC 2396 (https://tools.ietf.org/html/rfc2396), published in August 1998, and finalized in RFC 3986 (https://tools.ietf.org/htm l/rfc3986), published in January 2005.

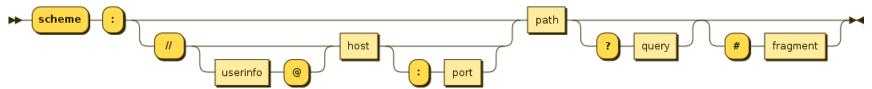
The *URI generic syntax* consists of a hierarchical sequence of five *components*: [22]

```
URI = scheme:[//authority]path[?query][#fragment]
```

where the authority component divides into three *subcomponents*:

```
authority = [userinfo@]host[:port]
```

This is represented in a syntax diagram as:



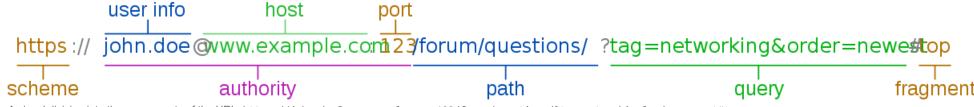
The URI comprises:

- A non-empty **scheme** component followed by a colon (:), consisting of a sequence of characters beginning with a letter and followed by any combination of letters, digits, plus (+), period (.), or hyphen (-). Although schemes are case-insensitive, the canonical form is lowercase and documents that specify schemes must do so with lowercase letters. Examples of popular schemes include https.ftp, mailto.file, data, and irc. URI schemes should be registered with the Internet Assigned
 Numbers Authority (IANA), although non-registered schemes are used in practice. [b]
- An optional **authority** component preceded by two slashes (//), comprising:
 - An optional **userinfo** subcomponent that may consist of a <u>user name</u> and an optional <u>password</u> preceded by a colon (:), followed by an at symbol (@). Use of the format username: password in the userinfo subcomponent is deprecated for security reasons. Applications should not render as clear text any data after the first colon (:) found within a userinfo subcomponent unless the data after the colon is the empty string (indicating no password).
 - A **host** subcomponent, consisting of either a registered name (including but not limited to a <u>hostname</u>), or an <u>IP address</u>. <u>IPv4</u> addresses must be in <u>dot-decimal</u> notation, and <u>IPv6</u> addresses must be enclosed in brackets ([]).
 - An optional **port** subcomponent preceded by a colon (:).
- A path component, consisting of a sequence of path segments separated by a slash (/). A path is always defined for a URI, though the defined path may be empty (zero length). A segment may also be empty, resulting in two consecutive slashes (//) in the path component. A path component may resemble or map exactly to a file system path, but does not always imply a relation to one. If an authority component is present, then the path component must either be empty or begin with a slash (/). If an authority component is absent, then the path cannot begin with an empty segment, that is with two slashes (//), as the following characters would be interpreted as an authority component. [26] The final segment of the path may be referred to as a 'slug'.
- An optional **query** component preceded by a question mark (?), containing a <u>query string</u> of non-hierarchical data. Its syntax is not well defined, but by convention is most often a sequence of attribute–value pairs separated by a delimiter.
- An optional **fragment** component preceded by a <u>hash</u> (#). The fragment contains a <u>fragment identifier</u> providing direction to a secondary resource, such as a section heading in an article identified by the remainder of the URI. When the primary resource is an <u>HTML</u> document, the fragment is often an <u>id attribute</u> of a specific element, and web browsers will scroll this element into view.

Query delimiter	Example
Ampersand (&)	key1=value1&key2=value2
Semicolon (;) ^[d]	key1=value1;key2=value2

Strings of data octets within a URI are represented as characters. Permitted characters within a URI are the ASCII characters for the lowercase and uppercase letters of the modern English alphabet, the Arabic numerals, hyphen, period, underscore, and tilde. Octets represented by any other character must be percent-encoded.

The following figure displays example URIs and their component parts.



A visual division into the components of the URL: https://john.doe@www.example.com:123/forum/questions/?tag=networking&order=newest#top



URI references

A *URI reference* is either a URI, or a *relative reference* when it does not begin with a scheme component followed by a colon (:).[32] A path segment that contains a colon character (e.g., foo:bar) cannot be used as the first path segment of a relative reference if its path component does not begin with a slash (/), as it would be mistaken for a scheme component. Such a path segment must be preceded by a dot path segment (e.g., ./foo:bar).[33]

Web document markup languages frequently use URI references to point to other resources, such as external documents or specific portions of the same logical document: [34]

- in HTML, the value of the src attribute of the img element provides a URI reference, as does the value of the href attribute of the a or link element;
- in XML, the system identifier appearing after the SYSTEM keyword in a DTD is a fragmentless URI reference;
- in XSLT, the value of the href attribute of the xsl:import element/instruction is a URI reference; likewise the first argument to the document() function.

```
https://example.com/path/resource.txt#fragment
//example.com/path/resource.txt
/path/resource.txt
path/resource.txt
../resource.txt
../resource.txt
#fragment
```

Resolution

Resolving a URI reference against a base URI results in a target URI. This implies that the base URI exists and is an absolute URI (a URI with no fragment component). The base URI can be obtained, in order of precedence, from: [35]

- the reference URI itself if it is a URI;
- the content of the representation;
- the entity encapsulating the representation;
- the URI used for the actual retrieval of the representation;

• the context of the application.

Within a representation with a well defined base URI of

```
http://a/b/c/d;p?q
```

a relative reference is resolved to its target URI as follows: [36]

```
"g:h"
"g"
         -> "http://a/b/c/g"
         -> "http://a/b/c/g"
"g/"
         -> "http://a/b/c/g/"
"/g"
         -> "http://a/g"
         -> "http://g"
"?y"
         -> "http://a/b/c/d;p?y'
"g?y"
         -> "http://a/b/c/g?y"
         -> "http://a/b/c/d;p?q#s"
         -> "http://a/b/c/g#s"
        -> "http://a/b/c/g?y#s"
";x"
         -> "http://a/b/c/;x"
"g;x"
        -> "http://a/b/c/g;x"
"g;x?y#s" -> "http://a/b/c/g;x?y#s"
         -> "http://a/b/c/d;p?q"
         -> "http://a/b/c/"
         -> "http://a/b/c/"
         -> "http://a/b/"
         -> "http://a/b/"
       -> "http://a/b/g"
        -> "http://a/"
"../../" -> "http://a/"
"../../g" -> "http://a/g"
```

Relation to XML namespaces

In \underline{XML} , a namespace is an abstract domain to which a collection of element and attribute names can be assigned. The namespace name is a character string which must adhere to the generic URI syntax. [37] However, the name is generally not considered to be a URI, because the URI specification bases the decision not only on lexical components, but also on their intended use. A namespace name does not necessarily imply any of the semantics of URI schemes; for example, a namespace name beginning with http: may have no connotation to the use of the HTTP.

Originally, the namespace name could match the syntax of any non-empty URI reference, but the use of relative URI references was deprecated by the W₃C. [39] A separate W₃C specification for namespaces in XML 1.1 permits internationalized resource identifier (IRI) references to serve as the basis for namespace names in addition to URI references. [40]

See also

- CURIE
- Dereferenceable Uniform Resource Identifier
- Extensible Resource Identifier
- Internationalized Resource Identifier
- Internet Resource Locator
- Persistent uniform resource locator

- Uniform Naming Convention
- Resource Directory Description Language
- Universally unique identifier
- List of URI schemes

Notes

- a. A report published in 2002 by a joint W3C/IETF working group aimed to normalize the divergent views held within the IETF and W3C over the relationship between the various 'UR*' terms and standards. While not published as a full standard by either organization, it has become the basis for the above common understanding and has informed many standards since then.
- b. The procedures for registering new URI schemes were originally defined in 1999 by RFC 2717 (https://tools.ietf.org/html/rfc2717), and are now defined by RFC 7595 (htt ps://tools.ietf.org/html/rfc7595), published in June 2015.[23]
- c. For URIs relating to resources on the World Wide Web, some web browsers allow .0 portions of dot-decimal notation to be dropped or raw integer IP addresses to be used [25]
- d. Historic RFC 1866 (https://tools.ietf.org/html/rfc1866) (obsoleted by RFC 2854 (https://tools.ietf.org/html/rfc2854)) encourages CGI authors to support ';' in addition to '&' [27]

References

- DuCharme, Bob (2013-07-03). Learning SPARQL (https://rk1bukz.cf/read.php?id =s2RgGxd572QC). O'Reilly Media, Inc. pp. 20–23. ISBN 9781449371470.
- W3C. "Universal Resource Identifiers in WWW" (https://www.w3.org/Addressing/ URL/uri-spec.html). W3.Org. Internet Engineering Task Force (IETF). Retrieved 2020-12-05.
- 3. Palmer, Sean. "The Early History of HTML" (http://infomesh.net/html/history/early/). infomesh.net. Retrieved 2020-12-06.
- 4. "W3 Naming Schemes" (https://www.w3.org/History/19921103-hypertext/hypertext/www/Addressing/Addressing.html). www.w3.org. 1992. Retrieved 2020-12-06.
- 5. Berners-Lee, Tim (June 1994). "Universal Resource Identifiers in WWW" (https://t ools.ietf.org/html/rfc1630). Network Working Group. Retrieved 2020-12-06.
- 6. Berners-Lee, Tim (December 1994). "Request for Comments: 1738: Uniform Resource Locators (URL)" (https://tools.ietf.org/html/rfc1738). tools.ietf.org/html. Retrieved 2020-12-06.
- Moats, R. (May 1997). "Request for Comments: 2141: URN Syntax" (https://tools.ietf.org/html/rfc2141). tools.ietf.org. Retrieved 2020-12-06.
- Berners-Lee, Tim (August 1998). "RFC 2396: Uniform Resource Identifiers (URI): Generic Syntax" (https://tools.ietf.org/html/rfc2396). tools.ietf.org. Retrieved 2020-12-06.
- 9. RFC 2396 (1998).
- 10. Hinden, R. (December 1999). "RFC 2732:Format for Literal IPv6 Addresses in URL's" (https://tools.ietf.org/html/rfc2732). tools.ietf.org. Retrieved 2020-12-06.
- 11. Berners-Lee, Tim (January 2005). "RFC 3986: Uniform Resource Identifier (URI): Generic Syntax" (https://tools.ietf.org/html/rfc3986). tools.ietf.org. Retrieved 2020-12-06.

- 12. Fielding, R. (June 1999). "RFC 2616: Hypertext Transfer Protocol -- HTTP/1.1" (https://tools.ietf.org/html/rfc2616). *tools.ietf.org*. Retrieved 2020-12-06.
- 13. Raman, T.V. (2006-11-01). "On Linking Alternative Representations To Enable Discovery And Publishing" (https://www.w3.org/2001/tag/doc/alternatives-discovery.html). www.w3.org. Retrieved 2020-12-06.
- Mealling, M. (August 2002). "RFC 3305: Uniform Resource Identifiers (URIs), URLs, and Uniform Resource Names" (https://tools.ietf.org/html/rfc3305). tools.ietf.org. Retrieved 2020-12-06.
- 15. Fielding, Roy (2005-06-18). "[httpRange-14] Resolved" (https://lists.w3.org/Archives/Public/www-tag/2005Jun/0039.html). *lists.w3.org*. Retrieved 2020-12-06.
- 16. Sauermann, Leo (December 2008). "Cool URIs for the Semantic Web" (https://www.w3.org/TR/cooluris/). www.w3.org. Retrieved 2020-12-06.
- URI Planning Interest Group, W3C/IETF (September 2001). "URIs, URLs, and URNs: Clarifications and Recommendations 1.0" (https://www.w3.org/TR/uri-clarif ication/). www.w3.org. W3C/IETF. Retrieved 2020-12-08.
- 18. Joint W3C/IETF URI Planning Interest Group (2002).
- 19. "URL Standard: 6.3. URL APIs elsewhere" (https://url.spec.whatwg.org/#url-apis-elsewhere).
- "URL Standard: Goals" (https://url.spec.whatwg.org/#goals).
- 21. RFC 3986 (2005).
- 22. RFC 3986, section 3 (2005).
- 23. IETF (2015).
- 24. RFC 3986 (2005), §3.2.2.
- 25. Lawrence (2014).
- 26. RFC 2396 (1998), §3.3.

- 27. RFC 1866 (1995), §8.2.1. 28. RFC 3986 (2005), §2.
- 29. RFC 3986 (2005), §2.2.
- 30. RFC 3986 (2005), §3.3.
- 31. RFC 3986 (2005), §3.4.
- 32. RFC 3986 (2005), §4.1.
- 33. RFC 3986 (2005), §4.2.

- 34. RFC 3986 (2005), §4.4.
- 35. RFC 3986 (2005), §5.1.
- 36. RFC 3986 (2005), §5.4.
- 37. Morrison (2006).
- 38. Harold (2004).
- 39. W3C (2009).
- 40. W3C (2006).

Further reading

- Harold, Elliotte Rusty (2004). XML 1.1 Bible (Third ed.). Wiley Publishing. p. 291.
 ISBN 978-0-7645-4986-1.
- Hansen, Tony; Hardie, Ted (June 2015). Thaler, Dave (ed.). "Guidelines and Registration Procedures for URI Schemes" (https://tools.ietf.org/html/rfc7595). leff Request for Comments (RFC) Pages Test. Internet Engineering Task Force. ISSN 2070-1721 (https://www.worldcat.org/issn/2070-1721).
- Morrison, Michael Wayne (2006). "Hour 5: Putting Namespaces to Use". Sams Teach Yourself XML. Sams Publishing. p. 91.
- URI Planning Interest Group, W3C/IETF (2001-09-21). "URIs, URLs, and URNs: Clarifications and Recommendations 1.0" (http://www.w3.org/TR/uri-clarification/). Retrieved 2009-07-27.
- "On Linking Alternative Representations To Enable Discovery And Publishing" (htt p://www.w3.org/2001/tag/doc/alternatives-discovery.html). World Wide Web Consortium. 2006 [2001]. Retrieved 2012-04-03.
- Bray, Tim; Hollander, Dave; Layman, Andrew; Tobin, Richard, eds. (2006-08-16).
 "Namespaces in XML 1.1 (Second Edition)" (http://www.w3.org/TR/REC-xml-names/#iri-use).
 World Wide Web Consortium.
 2.2 Use of URIs as Namespace Names.
 Retrieved 2015-08-31.
- Ayers, Danny; Völkel, Max (2008-12-03). Sauermann, Leo; Cyganiak, Richard (eds.). "Cool URIs for the Semantic Web" (http://www.w3.org/TR/cooluris/). World Wide Web Consortium. Retrieved 2012-04-03.

- <u>Bray, Tim;</u> Hollander, Dave; Layman, Andrew; Tobin, Richard; Thompson, Henry S., eds. (2009-12-08). "Namespaces in XML 1.0 (Third Edition)" (http://www.w3.org/TR/REC-xml-names/#iri-use). World Wide Web Consortium. 2.2 Use of URIs as Namespace Names. Retrieved 2015-08-31.
- Berners-Lee, Tim; Connolly, Daniel "Dan" (November 1995). "Hypertext Markup Language 2.0" (https://tools.ietf.org/html/rfc1866#section-8.2.1). Internet Engineering Task Force. Retrieved 2015-09-13.
- Berners-Lee, Tim; Fielding, Roy T.; Masinter, Larry (August 1998). *Uniform Resource Identifiers (URI): Generic Syntax* (http://tools.ietf.org/html/rfc2396). Internet Engineering Task Force. doi:10.17487/RFC2396 (https://doi.org/10.1748 7%2FRFC2396). RFC 2396 (https://tools.ietf.org/html/rfc2396). Retrieved 2015-08-31.
- Berners-Lee, Tim; Fielding, Roy T.; Masinter, Larry (January 2005). <u>Uniform Resource Identifiers (URI): Generic Syntax</u> (http://tools.ietf.org/html/rfc3986). Internet Engineering Task Force. doi:10.17487/RFC3986 (https://doi.org/10.1748 7%2FRFC3986). <u>RFC</u> 3986 (https://tools.ietf.org/html/rfc3986). Retrieved 2015-08-31.
- Berners-Lee, Tim; Fielding, Roy T.; Masinter, Larry (January 2005). Uniform Resource Identifiers (URI): Generic Syntax, section 3, Syntax Components (https://tools.ietf.org/html/rfc3986#section-3). Internet Engineering Task Force. doi:10.17487/RFC3986 (https://doi.org/10.17487%2FRFC3986). RFC 3986 (https://tools.ietf.org/html/rfc3986). Retrieved 2015-08-31.
- Lawrence, Eric (2014-03-06). "Browser Arcana: IP Literals in URLs" (http://blogs.msdn.com/b/ieinternals/archive/2014/03/06/browser-arcana-ipv4-ipv6-literal-urls-dotted-va-dotless.aspx). IEInternals. Microsoft. Retrieved 2016-04-25.

External links

- <u>URI Schemes (https://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml)</u> <u>IANA</u>-maintained registry of URI Schemes
- URI schemes on the W3C wiki (https://www.w3.org/wiki/UriSchemes)
- Architecture of the World Wide Web, Volume One, §2: Identification (https://www.w3.org/TR/webarch/#identification) by W3C
- W3C URI Clarification (https://www.w3.org/TR/uri-clarification/)

This page was last edited on 8 December 2020, at 00:51 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.