

Inhaltsverzeichnis

1.0	Der URL	2
1.1	Uniform Resource Locator (URL) ☺☺☺	2
1.2	URL-Schema bei »http«.....	3
1.3	URL-Schema bei »mailto«.....	4
1.4	URL-Schema bei »news«	5
2.0	Auszug aus dem RFC 1738:	5

Diese Dokumentation wurde verfasst von
Bernhard Kreinz
Systemspezialist und Webmaster

Der URL

1.0 Der URL

Uniform Resource Locators (URL)

1.1 Uniform Resource Locator (URL) ☺☺☺

Das World Wide Web-Projekt hat neben den beiden Features HTTP und der Hypertext-Funktionalität sowie dem DNS noch ein viertes Standbein, daß die revolutionäre Verbreitung erst ermöglichte: Dank der URL-Schemata ist es möglich, jede Ressource im Internet (auch außerhalb des WWW) eindeutig zu adressieren und anzusprechen.

Das grundlegende Schema

Ein gültiger URL hat grundsätzlich folgende Syntax:

<Protokoll>://<Userid>@<Host>:<Port>/<Pfad>

»Protokoll« steht für das anzuwendende Internet-Protokoll, also z.B. »http« für Hypertext Transfer Protocol, »ftp« für File Transfer Protocol oder »mailto« für den eMail-Dienst. Gefolgt wird diese Protokollkennzeichnung von einem Doppelpunkt und, falls der URL eine nicht lokale Aktion auslösen soll, mit zwei Querstrichen (»//«).

»UserID« enthält User-ID und/oder Paßwort-Informationen, die zum Benutzen der Ressource unbedingt erforderlich sind. Wird zum Beispiel eine paßwortgeschützte HTTP-Ressource angesprochen, können alternativ zur Dialogfenstereingabe diese Angaben auch hier eingesetzt werden, was z.B. so einen Bookmark ermöglicht. Also z.B. :

<http://ErnstEiswuerfel:GeheimesPasswort@www.barnes.ch/Telematik/index.html>

Standardmässig meldest du dich an jedem System (Webserver) im Internet an. Und zwar mit dem „anonymous“ – Benutzer, den ein Webserver für den public Access braucht an.

»Host« enthält den Rechner, der angesprochen werden soll, also z.B. »barnes.ch«.

»Port« enthält die Nummer des TCP/IP-Ports, an den die Anfrage geschickt wird. Z.B. werden HTTP-Anfragen voreingestellt an Port 80 geschickt.

Der URL

»Pfad« enthält nun die Informationen, die »vor Ort« beim Server verarbeitet werden. Dazu gehören Pfadangaben, aber auch Übergabeparameter, Sprungziele und Abfrageinformationen.

1.2 URL-Schema bei »http«

»http« kennzeichnet alle URL, die mit dem HTTP-Protokoll arbeiten. Voreingestellt ist bei »http« die Portnummer 80, bei Bedarf kann jedoch explizit auch ein anderer Port angesprochen werden (z.B. zum Ansprechen eines Proxy-Servers).

`http://www.netplanet.org`

Der Browser schickt eine GET-Anfrage direkt an den Host »www.netplanet.org«. Da explizit keine Datei angefordert wird, schickt der Server entweder einen Index des Verzeichnisses oder eine Default-Datei, d.h. die Datei, die vom Administrator des Servers dazu auserwählt wurde, immer dann verschickt zu werden, wenn nur ein Verzeichnis angesprochen wird.

`http://user-ID:Passwort@www.netplanet.org`

Informationen, die vor den Klammeraffen (»@«) gesetzt werden, der seinerseits vor der Hostadresse sitzt, werden als Identifikationsparameter, z.B. für serverseitig paßwortgeschützte Bereiche, benutzt. Moderne Browser blenden bei Anwahl einer solchen Seite ein Dialogfeld ein, in das die Zugangskennungen eingegeben werden müssen. Intern übersetzt der Browser diese Informationen in diese URL-Form. Durch direkte Eingabe der Zugangskennungen in einen URL können Sie z.B. einen Bookmark setzen, der direkten Zugriff auf die paßwortgeschützte Seite ermöglicht. Hinweis: Werden die Zugangskennungen direkt in die URL eingesetzt, werden sie nicht verborgen angezeigt!

`http://www.netplanet.org:8080`

Zusätzlich kann auch direkt der Port gewählt werden, an den die Anfrage beim Host geschickt werden soll. Ist im URL kein Port angegeben, wird standardmäßig die Anfrage an den Port 80 geschickt, der für HTTP reserviert ist.

`http://www.netplanet.org/index.html`

Bei diesem Aufruf wird explizit die Datei »index.html« vom Host angefordert.

`http://www.netplanet.org/index.html#sprungziel`

Ist in einem URL eine Raute vorhanden, so wird die Information hinter der Raute bis zu einem nächsten URL-spezifischen Zeichen (oder dem Ende des URL) vom Browser als Sprungziel innerhalb der spezifizierten Datei definiert. Die komplette Datei wird angefordert und lokal nach dem angegebenen Sprungziel abgesucht. Ist das Sprungziel nicht vorhanden, wird die Datei so angezeigt, als wäre sie ohne Sprungziel angefordert worden.

Der URL

<http://www.netplanet.org/index.html?ak=Abfrage>

Ein Fragezeichen in einem URL kennzeichnet bis zu einem nächsten URL-relevanten Zeichen (oder dem Ende der URL) den sogenannten »Abfrageteil«. In diesem Bereich lassen sich Parameter oder Informationen mit dem URL verschicken und entfernte Aktionen, z.B. CGI-Programme, füttern. Zum Versenden muß jedoch der Text der Variablen angepaßt werden, d.h. Sonderzeichen und Leerschritte müssen maskiert werden.

<http://www.netplanet.org/editorial/>

In diesem Fall wird ein Zugriff auf das Verzeichnis »editorial« auf diesem Host vorgenommen. Da es sich um ein Verzeichnis handelt, schickt der Server automatisch die Default-Datei, ansonsten einen Index vom Verzeichnis. WICHTIG: Wird direkt auf ein Verzeichnis zugegriffen, sollte immer ein abschließender Schrägstrich folgen, da der Server bei fehlendem Schrägstrich immer davon ausgeht, daß es sich um eine Datei handelt. Findet er keine Datei, schickt er eine HTTP-Fehlermeldung (Klasse 3xx - Redirection), die jedoch, dank fehlertoleranter Programmierung moderner Browser, nicht angezeigt wird, sondern den Browser dazu veranlaßt, den gleichen URL nochmals mit abschließendem Schrägstrich abzusenden.

1.3 URL-Schema bei »mailto«

Beginnt ein URL mit »mailto«, so ruft ein Klick auf diesen Link in der Regel ein Mailprogramm auf und übernimmt die Parameter des URL. Der URL wird also nicht verschickt, sondern löst nur eine lokale Aktion aus.

<mailto:besim@netplanet.org>

Der »klassische« mailto-URL, der ein Mailfenster öffnet und lediglich die Adresse des Empfängers vorgibt.

<mailto:besim@netplanet.org?to=hans@mustermann.de>

Mit »to=« können zusätzliche Empfänger angegeben werden.

<mailto:besim@netplanet.org?cc=hans@mustermann.de>

»cc=« ermöglicht die Angaben von eMail-Adressen, die als CC-Empfänger (»Carbon Copy«) angegeben werden können.

<mailto:besim@netplanet.org?bcc=hans@mustermann.de>

Mit diesem, ebenfalls fiktiven, URL wird hans@mustermann.de als BCC-Empfänger (»Blind Carbon Copy«) voreingestellt.

<mailto:besim@netplanet.org?newsgroups=de.test>

Dieser URL bereitet die Mail gleichzeitig auf den Versand zur eMail-Adresse und zur news-Gruppe »de.test« vor.

[mailto:besim@netplanet.org?body=Dies ist ein Text](mailto:besim@netplanet.org?body=Dies%20ist%20ein%20Text)

Der URL

Mit »body=« kann der Body-Text der eMail vorgeschrieben werden, der nach Anklicken schon im Mailfenster erscheint (selbstverständlich aber gelöscht oder noch editiert werden kann).

1.4 URL-Schema bei »news«

URL beginnend mit »news« lösen ebenfalls nur lokale Aktionen aus, in dem sie den news-Client öffnen und ggf. die news-Gruppe und evt. den news-Server vorbestimmen.

news:

Dieser URL allein öffnet nur den news-Client und baut evtl. eine Verbindung zu angegebenen news-Servern auf, um die vorhandene Zahl der Beiträge bei evtl. abonnierten news-Gruppen zu aktualisieren.

news:de.test

Bei Eingabe dieses URL wird der news-Client aufgerufen und direkt in die angegebene news-Gruppe gesprungen. Meist wird die Gruppe damit auch gleich abonniert. Ist lokal kein news-Server angegeben oder die angegebene news-Gruppe auf dem lokalen news-Server nicht vorhanden, wird eine Fehlermeldung ausgegeben.

news:news.server.de/de.test

In diesem URL ist neben der news-Gruppe auch der news-Server angegeben, zu dem konnektiert werden soll. Die news-Gruppe wird dann von diesem news-Server abonniert.

2.0 Auszug aus dem RFC 1738:

RFC 1738 Uniform Resource Locators (URL) December 1994

...

3. Specific Schemes

The mapping for some existing standard and experimental protocols is outlined in the BNF syntax definition. Notes on particular protocols follow. The schemes covered are:

ftp	File Transfer protocol
http	Hypertext Transfer Protocol
gopher	The Gopher protocol
mailto	Electronic mail address
news	USENET news
nntp	USENET news using NNTP access
telnet	Reference to interactive sessions

Der URL

wais	Wide Area Information Servers
file	Host-specific file names
prospero	Prospero Directory Service

Other schemes may be specified by future specifications. Section 4 of this document describes how new schemes may be registered, and lists some scheme names that are under development.

...

3.1. Common Internet Scheme Syntax

While the syntax for the rest of the URL may vary depending on the particular scheme selected, URL schemes that involve the direct use of an IP-based protocol to a specified host on the Internet use a common syntax for the scheme-specific data:

```
://:@:/
```

Some or all of the parts ":", "@", ":", and "/" may be excluded. The scheme specific data start with a double slash "/" to indicate that it complies with the common Internet scheme syntax. The different components obey the following rules:

user

An optional user name. Some schemes (e.g., ftp) allow the specification of a user name.

password

An optional password. If present, it follows the user name separated from it by a colon.

The user name (and password), if present, are followed by a commercial at-sign "@". Within the user and password field, any ":", "@", or "/" must be encoded.

Note that an empty user name or password is different than no user name or password; there is no way to specify a password without specifying a user name. E.g., has an empty user name and no password, has no user name, while has a user name of "foo" and an empty password.

host

The fully qualified domain name of a network host, or its IP address as a set of four decimal digit groups separated by ".". Fully qualified domain names take the form as described in Section 3.5 of RFC 1034 [13] and Section 2.1 of RFC 1123 [5]: a sequence of domain labels separated by ".", each domain label starting and ending with an alphanumerical character and possibly also containing "-" characters. The rightmost domain label will never start with a digit, though, which

Der URL

syntactically distinguishes all domain names from the IP addresses.

port

The port number to connect to. Most schemes designate protocols that have a default port number. Another port number may optionally be supplied, in decimal, separated from the host by a colon. If the port is omitted, the colon is as well.

url-path

The rest of the locator consists of data specific to the scheme, and is known as the "url-path". It supplies the details of how the specified resource can be accessed. Note that the "/" between the host (or port) and the url-path is NOT part of the url-path.

The url-path syntax depends on the scheme being used, as does the manner in which it is interpreted.