



SCHENKER

URL Encoding

Version 2.0

Revision history

Latest changes first!

Date	Version	Revision
2008-08-21	2.0	Document updated.
1999-11-01	1.0	Document created.

Table of contents

REVISION HISTORY	2
1. BACKGROUND	3
2. GENERAL	3
2.1 URLEncode	3
2.2 Character set	3
2.3 MIME-Type	3
2.4 Allowed characters	3
2.5 Convertering of hexadecimal	3
2.6 Containers	3
3. FURTHER READING	4

1. Background

In DB Schenkers WEB Development Rules (WDR) communication with containers are described. When the above are used between two applications communicating with HTTP, the serialising needs to be converted from 8-bits to a 7-bits protocol. This protocol is called URLEncode. Today this protocol is used primarily for communication between a web browser and a web server.

All platforms have (or can have) different character sets as a foundation in their systems. In order for two systems to be able to exchange information they have to agree on which character set to apply.

In WDR the mutual character set is ISO 8859-1, also called Latin-1. This character set is supported on most platforms, and is standard in the Unix and Windows world. The large character set Unicode is using Latin-1 as a base for it's two first blocks. A few applications are adjusted to UTF-8.

2. General

URLEncode is a 7-bits code for communication with HTTP. Apart from the allowed characters all others are replaced with the hexadecimal value for the character with a '%' (percent) as prefix. The exception is ' ' (blank space) which is replaced by the '+' (plus) character.

IMPORTANT!!! It is only the data in a serialised message which should be URLEncoded. The delimiters in the message should not be converted, i.e. '=' (equals) and '&' (Ampersand).

2.1 URLEncode

Based on the standard RFC 1738.

2.2 Character set

Communication based on ISO 8859-1 (Latin-1).

2.3 MIME-Type

Application/x-www-form-urlencoded

2.4 Allowed characters

Allowed characters are 'A' to 'Z', 'a' to 'z', '0' to '9'. These characters doesn't need to be converted. Blank spaces should be converted to a '+' (plus). All other characters must be converted.

2.5 Convertering of hexadecimal

All characters that should be converted must be converted to the characters hexadecimal value (2-digits) with one '%' as prefix. The hexadecimal characters 'A'-'F' could be either lowercase or uppercase.

Example:

Exclamation mark ("!") has in ISO 8859-1 the decimal value '31' and in hexadecimal the value '21'. The converted value is '%21'.

2.6 Containers

Containers consist of an array there every row in the array consist of two parts, firstly the name of the key and secondly it's value. Serialised it takes the format Key=Value&.

IMPORTANT!!! It is only the name of the key and it's value that should be converted, not the delimiters '=' and '&'.

Example:

Document: URL Encoding
Version: 2.0
Document name: URL_Encoding_en.pdf
Date: 2008-08-21



Schenker AB, IT Division

Sida 4 av 4

If you are to send the field “e_mail” with the value “info@schenker.com” and the field “code” with the value “k%9S2!” to a service, you would send the string “e%5Fmail=info%40schenker%2Ecom&code=k%259S2%21&”.

3. Further reading

Character entity references in HTML 4.0, 24.2 ISO 8859-1

<http://www.w3.org/TR/REC-html40/sgml/entities.html>

Specification of URL, RFC 1738

<http://www.faqs.org/rfcs/rfc1738.html>