

## Supported Characters in Semantic Web Middleware Manager (SWKM)

Date: 22 / 2 / 2007  
Author: Dimitris Andreou

Semantic Web Knowledge Middleware (SWKM) supports Unicode RQL/RUL query strings. Not every character is allowable though:

- A URI can have the following characters only:  
# - . / : ? \_ ~  
0..9  
a..z  
A..Z
- An RDF literal can be composed of the following characters:  
0x0009 (tab character)  
{0x0010..0xFFFF} (space character and beyond)

Care should be taken when the query in the client side is not stored in a Unicode-capable format (a java.lang.String with the appropriate characters should be fine). Furthermore, the query strings needs to be sent over the network, so it is up to the implementation of the client stub to preserve the queries' characters. JAX-WS has been tested for implementing the client side with no character problems whatsoever.

SWKM operations that serve XML will take care to appropriately encode characters that cannot be contained in a valid XML document. These characters are (followed by their respective encodings):

< → &lt;  
> → &gt;  
“ → &quot;  
' → &apos;  
& → &amp;