

Discussion

[Benjamin De Boe](#) · Nov 9, 2020

Accent sensitivity and other internationalization issues

Hi,

InterSystems IRIS has long supported the obvious translation functions required to for converting to upper or lowercase to enforce case-insensitive string comparison (e.g. in ObjectScript with [\\$zconvert](#)) and sorting (e.g. with [SQL collation functions](#), not to be confused with [NLS collation](#)). Customers in international contexts have at times used custom workarounds to also treat accent insensitivity or even more advanced normalization duct tape. We 're looking to address such use cases at the system and SQL level to increase convenience for this international audience, which is well represented on the Developer Community.

No rocket science here, but we hope to make this comprehensive yet at the same time keep it practical to build in and easy to use for customers (many of which will only need the basics):

- Extend `$zconvert()` with an "A" flag (next to "U" and "L") for stripping accents based on the recommended [ICU transformation](#). This will be implemented through an internal translation table which can be customized if need by (as has always been the case in IRIS)
- Add a new `%SQLGENERIC(transformation-spec [, truncate-length])` collation function for use in DDL (defining columns or indices) or SQL queries, using parameters to control the primary "transformations" our existing collation functions offer (uppercasing, trimming whitespace, ...), plus the new accent stripping feature.

To illustrate how this would look, here's how you could express the existing collation functions using `%SQLGENERIC`:

Current function	Using <code>%SQLGENERIC()</code>
<code>%EXACT</code>	<code>%SQLGENERIC</code>
<code>%SQLSTRING(n)</code>	<code>%SQLGENERIC('prepend-space, strip-trailing-whitespace', n)</code>
<code>%SQLUPPER(n)</code>	<code>%SQLGENERIC('prepend-space, strip-trailing-whitespace, mode=U', n)</code>
<code>%TRUNCATE(n)</code>	<code>%SQLGENERIC(", n)</code>
<code>%MINUS</code>	<code>%SQLGENERIC('mode=-')</code>
<code>%PLUS</code>	<code>%SQLGENERIC('mode=+')</code>

If you have faced such requirements and/or implemented any of those creative workarounds in the past, please use the poll below or chime in how you've approached this in the past and how you think the above addresses this.

PLEASE NOTE: this is a proposal shared here for soliciting your valued feedback. We are not committing to this specific approach nor a particular target release at this point. We are awaiting your feedback, after all :-)

Thanks,
benjamin

[#ObjectScript](#) [#SQL](#) [#InterSystems](#) [IRIS](#)

Source URL: <https://community.intersystems.com/post/accent-sensitivity-and-other-internationalization-issues>