Experiments based on "Guidelines for the homogenization of VTL 1.1 syntax"

1 Introduction

This document collects some experiments where the new "Guidelines for the homogenization of VTL 1.1 syntax" are applied to existing VTL 1.1 operators.

2 Set operators

Example 1 *In this example, we calculate the union of two datasets expressed as a list of datasets.*

$ds_1 := \texttt{union}([ds_2, ds_3])$

Example 2 *In this example, we calculate the union of three datasets expressed as a list of datasets. Moreover, one of the involved datasets is in turn obtained from another VTL expression.*

$ds_1 := \texttt{union}([ds_2, ds_3, (ds_4 + ds_5)/2])$

Example 3 *In this example we report the template invocation for* `union`*.*

$ds := \texttt{union}([ds_1, \ldots, ds_n])$

Observe that in the prototypical invocation in Example 3, the *deduplication functions*, which are present in the current specification, are not shown as they may be removed from the next release for the sake of simplicity. Moreover, there are some concerns regarding their use in multi-measure cases.

3 Data validation operators

Example 4 *In this example we check that a dataset corresponds to the sum of two datasets and returns only the measures of invalid cases.*

$ds_1 := \texttt{check}(ds_4 = ds_2 + ds_3, \texttt{return\_only} := \text{"not\_valid"}, \texttt{output\_components} := \text{"measures"})$

Example 5 *In this example we try to check that the absolute difference between two datasets is not above a certain threshold.*

$$\begin{aligned} ds_1 := \texttt{check}(&ds_2 = ds_3,\\ &\texttt{imbalance} := \texttt{abs}(abs(ds_2 - ds_3)),\\ &\texttt{threshold} := 0.2,\\ &\texttt{output\_components} := \text{"condition"},\\ &\texttt{errorlevel} := 5) \end{aligned}$$

Example 6 *The following example is semantically equivalent to Example 5 and shows some features of the new syntactic conventions.*

$ds_1 :=$ `check`($\mathit{dataset} := ds_2 = ds_3$,
$\mathit{imbalance} :=$ `abs`$(abs(ds_2 - ds_3))$,
$\mathit{threshold} := 0.2$,
$\mathit{errorlevel} := 5$,
$\mathit{output\_components} :=$ *"condition"*)

Example 7 *In this example we report a template invocation for* `check`*.*

$ds_1 :=$ `check`$(dataset, \{threshold, return\_only, output\_components, imbalance, errorcode, errorlevel\})$