

Smush

The concept of smush is a way of combining the graphs of rdf data when they share certain links. When a property is determined to be identifying for a resource it can be renamed so that any graphs that have equivalent values in that property are condensed down to one resource.

Arguments

Short Option	Long Option	Parameter Value Map	Description	Required
i	inputJena-config	CONFIG_FILE	inputJena JENA configuration filename	true
I	inputOverride	override the JENA_PARAM of inputJena jena model config using VALUE	false	
o	outputJena-config	CONFIG_FILE	outputConfig JENA configuration filename	true
O	outputOverride	override the JENA_PARAM of outputJena jena model config using VALUE	false	
P	inputJena-predicates	PREDICATE	PREDICATE(s) on which, to match. Multiples are done in series not simultaneously.	true
n	namespace	NAMESPACE	only match rdf nodes in inputJena whose URI begin with NAMESPACE	false
r	replace		replace input model with changed / output model	false

Usage

Smush operates on a chosen Jena [model](#), the one shown here is using an h2 database backend.

If multiple predicates are supplied each one is operated on in turn and **will not produce an intersection of the sets**.

```
//from the env file
Smush="java $OPTS -Dprocess-task=Smush org.vivoweb.harvester.qualify.Smush"

//from the script file
SCOREINPUT="-i $H2MODEL -ImodelName=$MODELNAME -Idburl=$MODELDBURL -IcheckEmpty=$CHECKEMPTY"
CONNUM="http://vivo.ufl.edu/ontology/vivo-ufl/psContractNumber"
UFID="http://vivo.ufl.edu/ontology/vivo-ufl/ufid"
BASEURI="http://vivoweb.org/harvest/ufl/dsr/"

//smushes in-place(-r) on the contract number THEN on the UFID
$Smush $SCOREINPUT -P $CONNUM -P $UFID -n ${BASEURI} -r
```