

Output from serene module with data that is not working.

This example is trained using five data files from the same dataset (column names are not consistent)

Predictions only work when one of the training files is used.

ML lib predictions => look sensible, 7 classes (including unknown) and 23 columns

Loads the alignment graph.

Ontology cache – no. of classes:5
no. of all properties: 17
data properties: 7
object properties: 5

no. of properties explicitly defined as owl: DataTypeProperty: 12

no. of properties explicitly defined as owl: ObjectProperty: 10

Semantic modeller got 23 dataset predictions

Semantic modeller will use 8 ds predictions

Converted ds predictions correct

ModelLearningGraph – loading the alignment graph

GraphBuilder adds edges

KarmaSuggestModel – mappings appear correct (flightNo, dob, passportNo, name, gender, event date)

Confidence against correct type high > 90 – 100%

All other confidences < 10%

AutoAssignment happening and columns NotAssigned correctly

Alignment graph: 35 nodes, 11 links

8 steiner nodes are obtained

Model learner initialised

Column	No. of matches for column	No.of candidate steiner sets before update	No. of candidates steiner sets after update
flightNo	14	0	14
passport	20	14	50
entry date	26	50	50
middlename	32	50	50
surname	38	50	50
sex	43	50	50
birthdate	48	50	50
firstname	48	50	50

No. of possible mappings: 877032898560

Graph nodes: 75

Graph links: 51

No. of steiner sets: 50

No. of candidate steiner sets after update: 50

ModelLearner: results are ready

0 (zero) suggestions by Karma

One thing I noticed was if I run the prediction again using the same data as before, in the ModelLearner phase, the columns are processed in a different sequence but the output is exactly the same as the table above e.g. middlename (14,0,14), firstname (20,14,50) etc. Not sure if that means anything.