@Web Constraint Checking: Functional Specification

Leandro Lovisolo <leandro.lovisolo@supagro.inra.fr>,
INRA SupAgro and INRIA GraphiK,
Montpellier, France

December 2015

Abstract

In this document I summarize the changes proposed to the **@Web** platform in order to implement automatic constraint checking using SPARQL queries.

1 Core ontology changes

The following list provides a high level view of the changes proposed to the core ontology.

- A new OWL class `Constraint` is added to the core ontology, which represents constraints expressed as SPARQL queries.
- Instances of the `Constraint` class are associated to their respective relation classes via a new object property `hasForConstraint`.
- A new data property `hasForSPARQLQuery` is added, which connects instances of the `Constraint` class with a string literal holding the actual SPARQL query.
- Constraints are described in natural language with a textual guideline associated to `Constraint` instances via a SKOS scope note.
- A new `Error` concept is added, with the purpose of storing validation errors collected during the execution of the SPARQL constraints.
- Instances of the `Error` class are linked to the `Constraint` instance that produced the validation error via a new object property `hasForConstraint`.
- Instances of the `Error` class are linked to the `Row` instance associated to the unit operation relation instance that caused the error via a new object property `hasForRow`.

1.1 Biorefinery domain ontology

This particular domain requires the notion of experiment *categories*, which have associated constraints that require additional information. In the following sections, the concept of categories is explained, and then some changes to the domain ontology are proposed to support expressing the required constraints as SPARQL queries.

1.1.1 Categories

Categories are a way to group unit operation relation instances according to the kind of experiment they model. This grouping is currently done in the **@Web** platform via document topics (e.g., Bioref-PM, Bioref-PM-UFM, Bioref-PM-PC-EX-PS, etc.)

Categories are characterized by the two following points:

- Each unit operation relation instance must belong to exactly one category.
- Each category has clearly defined rules to decide whether an experiment (i.e. a set of unit operation relation instances) belongs to it or not.

Thus, given a unit operation relation instance and the category it belongs to, it is required to check said rules automatically. To this end, rules are encoded as constraints written as SPARQL queries.

1.1.2 Changes to the Biorefinery domain ontology

A new symblic concept `ProcessType` is created, with one subclass for each supported category. Each such subclass is listed below, with its proposed alternate label between parentheses:

- `Milling` (PM)
- `Milling_PhysicoChemical_Extrusion` (PM-PC-EX-PS)
- `Milling_PhysicoChemical` (PM-PC-PS)
- `Milling_PhysicoChemical_UltraFineMilling` (PM-PC-UFM)
- `Milling_PhysicoChemical_UltraFineMilling_PressSeparation` (PM-PC-UFM-PS)
- `Milling_UltraFineMilling` (PM-UFM)

A new argument is added to the n-ary relations that represent unit operations in this domain with the purpose of linking a relation instance with the category (i.e. `ProcessType` subclass) it belongs to.

The `Constraint` instances that verify the category inclusion rules are associated to the `Relation` class in order to make them available to all subclasses (i.e. all relations).

In a future version of **@Web** there should be a mechanism for ontology-level constraints that would allow a more correct way of expressing category constraints.

2 Uploading constraints in CSV format

A new screen is added to the **@Web** management UI with the purpose of uploading constraints as CSV files. Such files would have the following columns:

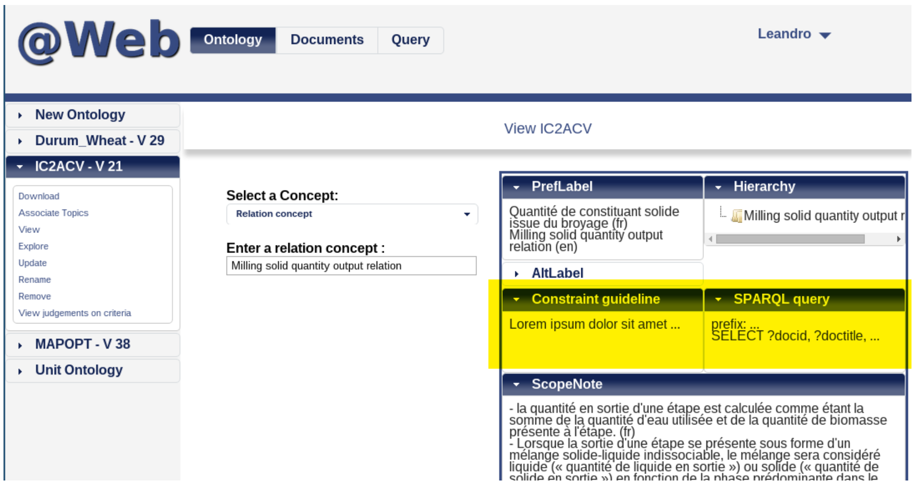
- `prefLabel=EN` (e.g. *Milling*)
- `altLabel=EN` (e.g. *PM*)
- `scopeNote=EN` (textual guideline)
- `Relation_Concept`
- `SPARQL_query`

3 User interface

The constraint visualization and verification user interface is described in the following sections.

3.1 Constraint visualization

Constraints are displayed next to the scope notes when exploring relations, as shown below.



Milling solid quantity output relation

Quantité de constituant solide issue du broyage (fr)
Milling solid quantity output relation (en)

Milling solid quantity output r

Lorem ipsum dolor sit amet ...

prefix: ...
SELECT ?docid, ?doctitle, ...

- la quantité en sortie d'une étape est calculée comme étant la somme de la quantité d'eau utilisée et de la quantité de biomasse présente à l'étape. (fr)
- Lorsque la sortie d'une étape se présente sous forme d'un mélange solide-liquide indissociable, le mélange sera considéré liquide (« quantité de liquide en sortie ») ou solide (« quantité de solide en sortie ») en fonction de la phase prédominante dans le

Figure 1: Constraint visualization.

Note that this visualization is read-only.

3.2 Constraint verification

The constraint verification process is done on a per-table basis, and is launched by right-clicking the target table and selecting *validate constraints*, as shown below.

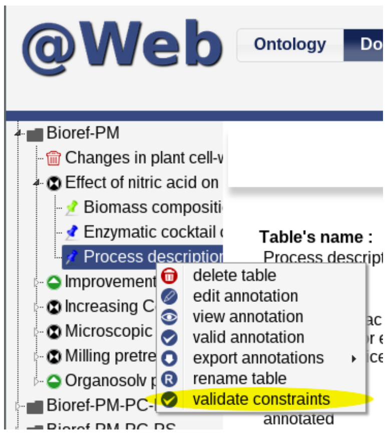


Figure 2: *validate constraints* menu item.

The user is then taken to a screen where they can select the constraints to verify. This list is compiled by finding all constraints linked to the categories associated with all relation instances in the table annotations.

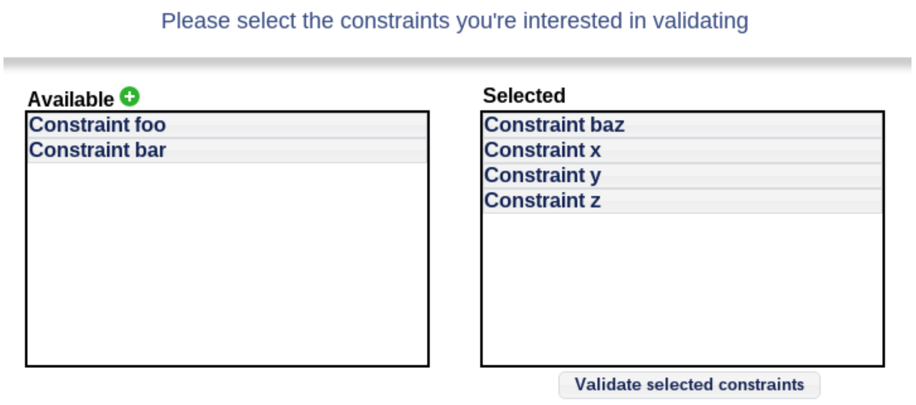


Figure 3: Constraint selection.

After clicking the *Validate selected constraints* button, a loading indicator is shown. When the queries finish running the user is taken to a results screen sketched below, where a summary of the errors is provided.

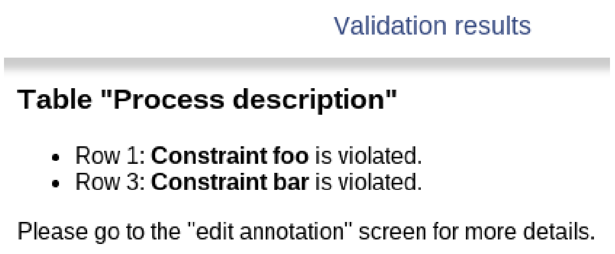

Validation results

Table "Process description"

- Row 1: **Constraint foo** is violated.
- Row 3: **Constraint bar** is violated.

Please go to the "edit annotation" screen for more details.

Figure 4: Summary of validation results.

The table icon is updated to reflect the presence or absence of errors.

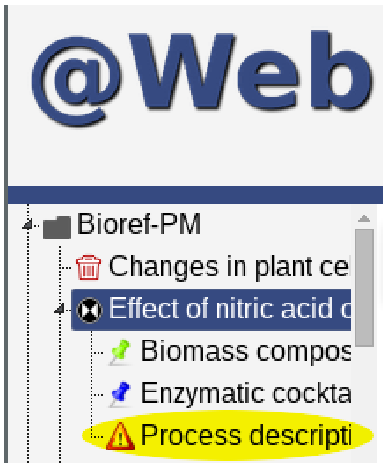
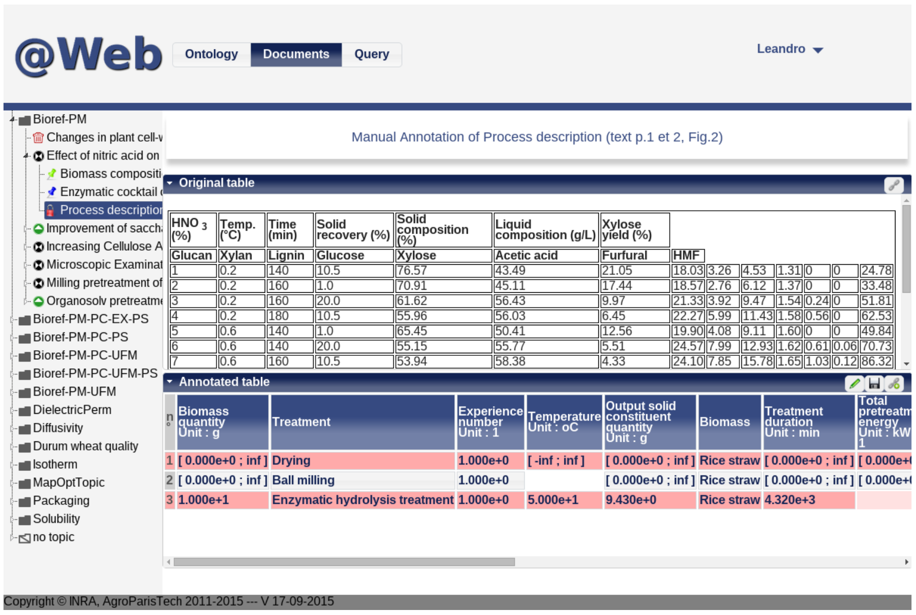


Figure 5: Table icon changes to reflect the presence of errors.

In the case of errors, when the user enters the table edition screen they will see the rows in a table containing errors highlighted in red.



Manual Annotation of Process description (text p.1 et 2, Fig.2)

Original table

HNO 3 (%)	Temp. (°C)	Time (min)	Solid recovery (%)	Solid composition (%)	Liquid composition (g/L)	Xylose yield (%)							
Glucan	Xylan	Lignin	Glucose	Xylose	Acetic acid	Furfural	HMF						
1	0.2	140	10.5	76.57	43.49	21.05	18.03	3.26	4.53	1.31	0	0	24.78
2	0.2	160	1.0	70.91	45.11	17.44	18.57	2.76	6.12	1.37	0	0	33.48
3	0.2	160	20.0	61.62	56.43	9.97	21.33	3.92	9.47	1.54	0.24	0	51.81
4	0.2	180	10.5	55.96	56.03	6.45	22.27	5.99	11.43	1.58	0.56	0	62.53
5	0.6	140	1.0	65.45	50.41	12.56	19.90	4.08	9.11	1.60	0	0	49.84
6	0.6	140	20.0	55.15	55.77	5.51	24.57	7.99	12.93	1.62	0.61	0.06	70.73
7	0.6	160	10.5	53.94	58.38	4.33	24.10	7.85	15.78	1.65	1.03	0.12	86.32

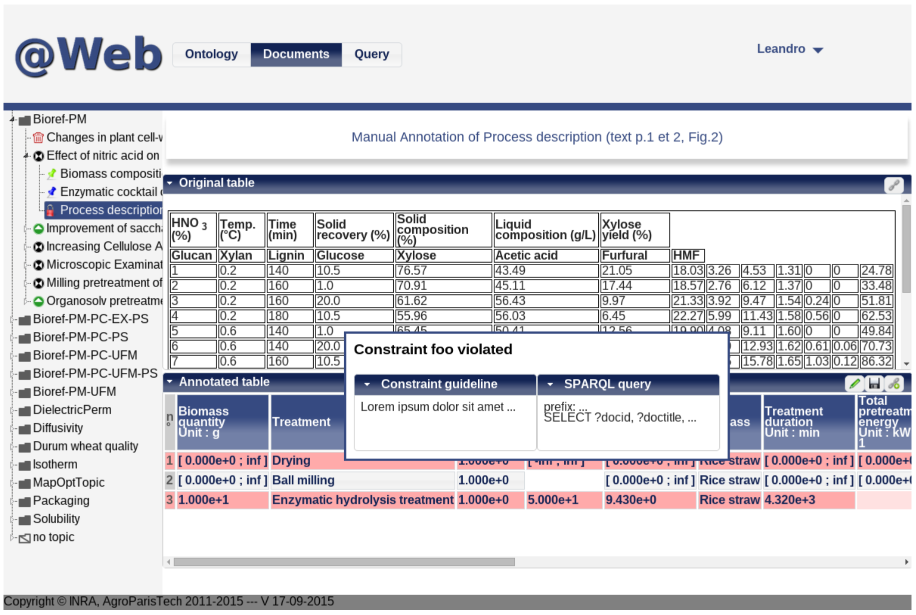
Annotated table

n°	Biomass quantity Unit : g	Treatment	Experience number Unit : 1	Temperature Unit : oC	Output solid constituent quantity Unit : g	Biomass	Treatment duration Unit : min	Total pretreatm energy Unit : kW 1
1	[0.000e+0 ; inf]	Drying	1.000e+0	[-inf ; inf]	[0.000e+0 ; inf]	Rice straw	[0.000e+0 ; inf]	[0.000e+
2	[0.000e+0 ; inf]	Ball milling	1.000e+0		[0.000e+0 ; inf]	Rice straw	[0.000e+0 ; inf]	[0.000e+
3	1.000e+1	Enzymatic hydrolysis treatment	1.000e+0	5.000e+1	9.430e+0	Rice straw	4.320e+3	

Copyright © INRA, AgroParisTech 2011-2015 --- V 17-09-2015

Figure 6: Annotated table with validation errors.

The user can get additional information (such as the name of the constraint violated) by hovering the mouse pointer over the affected row.



Manual Annotation of Process description (text p.1 et 2, Fig.2)

Original table

HNO 3 (%)	Temp. (°C)	Time (min)	Solid recovery (%)	Solid composition (%)	Liquid composition (g/L)	Xylose yield (%)							
Glucan	Xylan	Lignin	Glucose	Xylose	Acetic acid	Furfural	HMF						
1	0.2	140	10.5	76.57	43.49	21.05	18.03	3.26	4.53	1.31	0	0	24.78
2	0.2	160	1.0	70.91	45.11	17.44	18.57	2.76	6.12	1.37	0	0	33.48
3	0.2	160	20.0	61.62	56.43	9.97	21.33	3.92	9.47	1.54	0.24	0	51.81
4	0.2	180	10.5	55.96	56.03	6.45	22.27	5.99	11.43	1.58	0.56	0	62.53
5	0.6	140	1.0	65.45	50.41	12.56	19.90	4.08	9.11	1.60	0	0	49.84
6	0.6	140	20.0						12.93	1.62	0.61	0.06	70.73
7	0.6	160	10.5						15.78	1.65	1.03	0.12	86.32

Constraint foo violated

Constraint guideline

Lorem ipsum dolor sit amet ...

SPARQL query

```
prefix: ...
SELECT ?docid, ?doctitle, ...
```

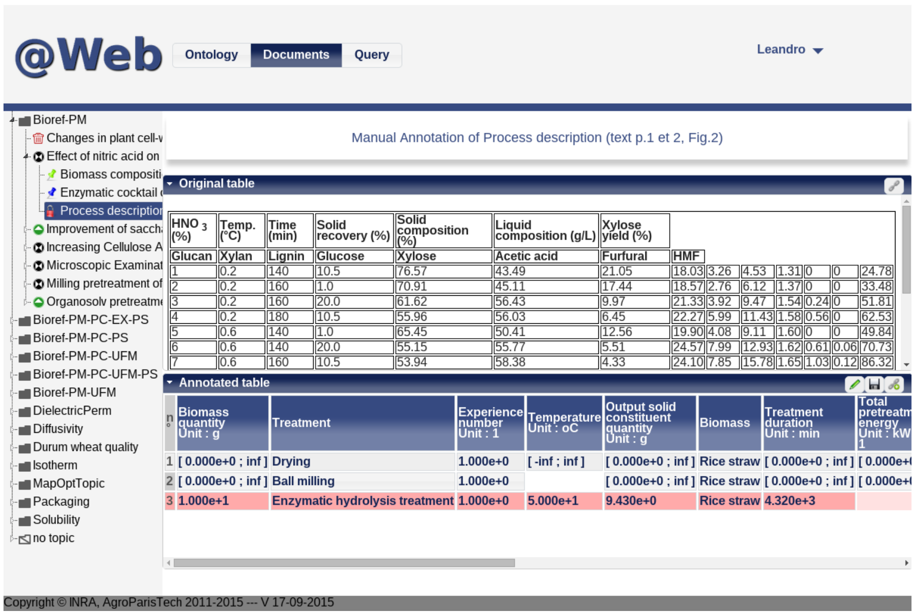
Annotated table

n°	Biomass quantity Unit : g	Treatment				ass	Treatment duration Unit : min	Total pretreatm energy Unit : kW 1
1	[0.000e+0 ; inf]	Drying	1.000e+0	[-inf ; inf]	[0.000e+0 ; inf]	Rice straw	[0.000e+0 ; inf]	[0.000e+
2	[0.000e+0 ; inf]	Ball milling	1.000e+0		[0.000e+0 ; inf]	Rice straw	[0.000e+0 ; inf]	[0.000e+
3	1.000e+1	Enzymatic hydrolysis treatment	1.000e+0	5.000e+1	9.430e+0	Rice straw	4.320e+3	

Copyright © INRA, AgroParisTech 2011-2015 --- V 17-09-2015

Figure 7: Validation error details.

After performing modifications on an affected row, the row color goes back to gray.



Manual Annotation of Process description (text p.1 et 2, Fig.2)

HNO 3 (%)	Temp. (°C)	Time (min)	Solid recovery (%)	Solid composition (%)	Liquid composition (g/L)	Xylose yield (%)							
Glucan	Xylan	Lignin	Glucose	Xylose	Acetic acid	Furfural	HMF						
1	0.2	140	10.5	76.57	43.49	21.05	18.03	3.26	4.53	1.31	0	0	24.78
2	0.2	160	1.0	70.91	45.11	17.44	18.57	2.76	6.12	1.37	0	0	33.48
3	0.2	160	20.0	61.62	56.43	9.97	21.33	3.92	9.47	1.54	0.24	0	51.81
4	0.2	180	10.5	55.96	56.03	6.45	22.27	5.99	11.43	1.58	0.56	0	62.53
5	0.6	140	1.0	65.45	50.41	12.56	19.90	4.08	9.11	1.60	0	0	49.84
6	0.6	140	20.0	55.15	55.77	5.51	24.57	7.99	12.93	1.62	0.61	0.06	70.73
7	0.6	160	10.5	53.94	58.38	4.33	24.10	7.85	15.78	1.65	1.03	0.12	86.32

n°	Biomass quantity Unit : g	Treatment	Experience number Unit : 1	Temperature Unit : oC	Output solid constituent quantity Unit : g	Biomass	Treatment duration Unit : min	Total pretreatm energy Unit : kW 1
1	[0.000e+0 ; inf]	Drying	1.000e+0	[-inf ; inf]	[0.000e+0 ; inf]	Rice straw	[0.000e+0 ; inf]	[0.000e+(
2	[0.000e+0 ; inf]	Ball milling	1.000e+0		[0.000e+0 ; inf]	Rice straw	[0.000e+0 ; inf]	[0.000e+(
3	1.000e+1	Enzymatic hydrolysis treatment	1.000e+0	5.000e+1	9.430e+0	Rice straw	4.320e+3	

Copyright © INRA, AgroParisTech 2011-2015 --- V 17-09-2015

Figure 8: Row with validation errors no longer in red after modifying its contents.

The table icon also changes color to reflect this new state, i.e. the table used to have errors and it has since been modified, but its validation constraints have yet to be rerun.

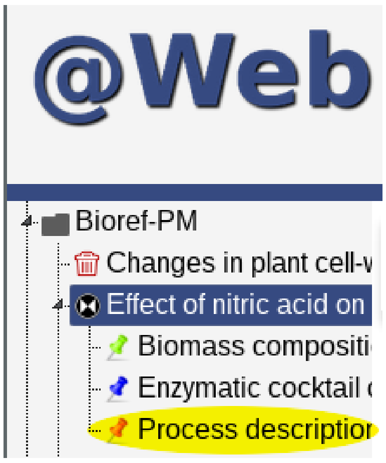


Figure 9: Table icon after modifying its annotations in the presence of validation errors.

If no errors are found after running the validation constraints, the table icon should be reset to the default icon.

4 Other implementation details

`Error` instances should be stored in the annotations ontology.

An `Error` instance should be eliminated after a modification to its related row is performed.

A new table should be created in the relational database to keep a list of all annotated tables that should be shown in the UI with an icon indicating the presence of errors. That is, annotated tables that have or had validation errors and have been modified since those errors were found, but their validation constraints haven't been rerun.