Experimental data annotation guided by an ontology for decision support: @Web project

Patrice Buche
JRU IATE, Montpellier

Juliette Dibie
JRU MIA, Paris
# L’équipe projet @Web en 2014

## Concepteurs-développeurs

### Permanents:
- Patrice Buche IR HDR, INRA IATE
- Brigitte Charnomordic, INRA MISTEA
- Stéphane Dervaux, IE INRA MIA
- Sébastien Destercke, CR CNRS HEUDYASIC
- Juliette Dibie, PR AgroParisTech MIA
- Liliana Ibanescu, MC AgroParisTech MIA
- Luc Menut AI, INRA IATE
- Michard Rakotosone, IR, INRA Plateforme CEPIA
- Mathieu Roche, CR CIRAD TETIS

### Non permanents:
- Lilia Berrahou, Doct. UM2 IATE-LIRMM
- Mathilde Salthun-Lasalle, IE INRA IATE-MIA
- Rim Touhami, Doct. INRA IATE-MIA

## Utilisateurs

### Permanents:
- Cédric Baudrit, CR INRA I2M
- Abdellatif Barrakhat, CR INRA IATE
- Jean-Pierre Belaud, MC ENSIACET LGC
- Valérie Guillard, MC UM2 IATE
- Xavier Rouau, DR INRA IATE
- Caroline Sablayrolle, MC ENSIACET LCA

### Non permanents:
- Filippo Acerbi, Doct. UM2 IATE
- Fabien Bibi, Doct. UM2 IATE
- Guillaume Busset, Post-doc ENSIACET LCA
- Charlotte Loustau Cazalet, Doct. INRA IATE
- Mia Kurek, Post-doc UM2 IATE
- Anaïs Licari, Master INRA IATE
- Ali Motedayen, Doct., UM2 IATE
- Céline Mousties, Master UM2 IATE
- Caroline Wolf, Doct. UM2 IATE
Heterogeneous data reuse for decision support

(Destercke et al. 2011, Buche et al 2013, Destercke et al. 2013, Tamani et al. 2014, Guillard et al. 2015)
I want a packaging which preserves my product, made of renewable resources, but without GMO, if possible transparent and with a « material » cost < 3 € / kg …
Multi-criteria querying

- Fresh produce database
- Virtual MAP simulation
- Multi-criteria flexible querying
- Stakeholder preferences and needs
- Packaging database
- Ranked list of most relevant packagings

- Industrial constraints
- Consumer preferences
- Waste management policy
- etc
Life cycle analysis (LCA) permits to estimate environmental impacts using a complete inventory of matter flow, energy, and effluents generated by the production process.
Pre-treatment eco-design

Biomass

1. Physico-chemical treatments
   - Ultrafin milling

2. Physico-chemical treatments
   - Ultrafin milling

3. Physico-chemical treatments
   - Washing and filtration
   - Ultrafin milling

4. Physico-chemical treatments
   - Extrusion
   - Washing and filtration

5. Physico-chemical treatments

6. Physico-chemical treatments

Enzymatic hydrolysis

Xylose
Glucose
Pre-treatment comparison for rice straw

![Graph showing different pre-treatment methods for rice straw]

- CM then dry BM
- Hot water
- Oxidizing treatment
- Ionic liquid treatment
- Hot water treatment then wet disk milling
- CM then Wet DM
- Hot compressed water
- Nitric acid pretreatment
- CM then Blender
- Ball milling [0.1;0.8]mm
- Torrefaction
- Cutting milling 2mm
- Ball milling [0.177;0.833]mm
- Organosolv

En-factor vs. Glucose yield (%) graph

Projet IC2ACV – 24 septembre 2014
Heterogeneous data capitalisation guided by an ontology

Table 1: Permeabilities of MFC films and literature values for films of synthetic polymers and cellophane

<table>
<thead>
<tr>
<th>Sample</th>
<th>Grammage (g/m²)</th>
<th>Thickness (µm)</th>
<th>Air permeability (nm/Pa s)</th>
<th>Oxygen permeability in the material (ml m⁻² day⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFC film A</td>
<td>17 ±1</td>
<td>21 ±1</td>
<td>13 ± 2</td>
<td>17.0, 18.5</td>
</tr>
<tr>
<td>EVOH</td>
<td>–</td>
<td>25</td>
<td>–</td>
<td>3–5</td>
</tr>
<tr>
<td>Cellophane</td>
<td>–</td>
<td>21</td>
<td>–</td>
<td>3</td>
</tr>
</tbody>
</table>

**Symbolic concept**
- EthyleneVinylAlcohol
- Packaging

**Quantity**
- Thickness
- Oxygen Permeability

**O2PermeabilityRelation**
Data capitalisation guided by an ontology

- EcoBioCap DSS
- Web open data access
- RDB
- RDF DB
- Ontology
- Reliability assessment
- Table and text extraction
- Semantic annotation
- Rech document (WOS, …)
- PDF
- HTML

Web
Termino-ontological resource

Conceptual component (OWL)

- OTR_Concept
  - T_Concept
    - Relation
    - Argument

Terminological component (SKOS)

- Ethylene_Vinyl_Alcohol
  - skos:prefLabel ‘Ethylene vinyl alcohol’@en
  - skos:altLabel ‘EVOH’@en

Punning

- ‘Ethylène alcool vinylique’@fr
  - skos:altLabel ‘EVOH’@fr

(Up core Ontology)
- UM_Concept
- Unit_Concept
- Quantity
- Symbolic_Concept
- Packaging
- Temperature
- Thickness
- O2Permeability_Relation

(Down core Ontology)

(Touhami et al. 2011, Buche et al. 2013c)
A relation concept guarantees the data reusability.
A relation concept models a unit operation

- Output flow
  - Milling solid qty output
    - hasOutput

- Input flow
  - Biomass
    - hasInput
  - Biomass quantity
    - hasInput
  - Total pretreatment energy
    - hasInput

- Control parameters
  - Wet disk milling
    - hasInput
  - Treatment duration
    - hasInput
  - Milling rotation speed
    - hasInput

Core ontology facilitates software component reuse
Exemple : Supprimer la quantité partial_pressure :

- Supprimer le lien d’héirarchie entre le concept générique Quantity et le concept partial_pressure.

- Supprimer les liens avec les propriétés ayant partial_pressure comme domaine.

- Supprimer les liens avec les propriétés ayant partial_pressure comme co-domaine.

- Supprimer la terminologie associée à partial_pressure.
Proposition: Notion de kit de changement

Thèse Rim Touhami Sept 2014
Data capitalisation guided by an ontology

Web

EcoBioCap DSS

RDB

Web open data access

RDF DB

Ontology

Semantic annotation

Table and text extraction

Reliability assessment

Rech document (WOS, …)
2 categories of meta data:

- Nature et source reputation
  - Publication date
  - Times cited (WOS)
  - Source type

- Experimentation management
  - Repetition number
Reliability expert quotation

<table>
<thead>
<tr>
<th>repetition</th>
<th>unknown</th>
<th>not at all reliable</th>
<th>not at all or hardly reliable</th>
<th>hardly reliable</th>
<th>hardly or average</th>
<th>average reliable</th>
<th>average or reliable</th>
<th>reliable</th>
<th>reliable very</th>
<th>very reliable</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; No</td>
<td></td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>++</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>number of repetitions</th>
<th>unknown</th>
<th>not at all reliable</th>
<th>not at all or hardly reliable</th>
<th>hardly reliable</th>
<th>hardly or average</th>
<th>average reliable</th>
<th>average or reliable</th>
<th>reliable</th>
<th>reliable very</th>
<th>very reliable</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; no repetition</td>
<td></td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; once</td>
<td></td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; twice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>&gt; 3 times</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>&gt; more than 3 times</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>++</td>
<td>++</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>age and citation number</th>
<th>unknown</th>
<th>not at all reliable</th>
<th>not at all or hardly reliable</th>
<th>hardly reliable</th>
<th>hardly or average</th>
<th>average reliable</th>
<th>average or reliable</th>
<th>reliable</th>
<th>reliable very</th>
<th>very reliable</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; between 3 and 8 years old &amp; less than 10 citations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; between 3 and 8 years old &amp; between 10 and 20 citations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[-,+]</td>
<td>[+,-]</td>
</tr>
<tr>
<td>&gt; between 3 and 8 years old &amp; between 20 and 40 citations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[+,-]</td>
<td>[+,-]</td>
</tr>
</tbody>
</table>

014 --- V 09-10-2014

@Web graphical user interface
<table>
<thead>
<tr>
<th>Citation Number</th>
<th>more than 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>more than 8 years old</td>
</tr>
</tbody>
</table>

**Criterion age and top citation**

<table>
<thead>
<tr>
<th>Age</th>
<th>more than 8 years old</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top Citation</td>
<td>top 0.10%</td>
</tr>
</tbody>
</table>

**Criterion source type**

| Source Type    | journal article       |

**Reliability evaluation's document information**

**Reliability results**

<table>
<thead>
<tr>
<th>Low expectation</th>
<th>4.94</th>
</tr>
</thead>
<tbody>
<tr>
<td>High expectation</td>
<td>5.0</td>
</tr>
</tbody>
</table>

**Known criteria values rate**

| 80 %            |

**Last evaluation date**

| 2014-09-29      |
## Classes

<table>
<thead>
<tr>
<th>Classes</th>
<th>Number of publications</th>
<th>Reliability Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very reliable</td>
<td>14</td>
<td>[ 4.97 ; 5 ]</td>
</tr>
<tr>
<td>Reliable</td>
<td>3</td>
<td>[ 3.2 ; 4.85 ]</td>
</tr>
<tr>
<td>In conflict</td>
<td>9</td>
<td>[ 1.01 ; 4.93 ]</td>
</tr>
<tr>
<td>Not at all reliable</td>
<td>4</td>
<td>[ 1.02 ; 1.33 ]</td>
</tr>
</tbody>
</table>

- **Very reliable**: peer-reviewed articles, with repetitions, high reputation
- **Reliable**: peer-reviewed articles, with repetitions, medium reputation
- **In conflict**: peer-reviewed articles but no repetition
- **Not at all reliable**: technical sheets without external review and no repetition
Data capitalisation guided by an ontology

EcoBioCap
DSS

Web open data access

RDB

RDF DB

Ontology

Semantic annotation

Table and text extraction

Reliability assessment

Rech document (WOS, ...)

Web

PDF

<html>
Verrou 2: Text-mining approach

Extracting experimental data

The oxygen permeability was measured according to the ASTM standard D3985-23 (3°C, 0% RH on the top side, 50% RH on the bottom side). The MFC films were mounted in a cell where 100% CO₂ was flushed on the top side and 100% N₂ on the bottom side. The amount of O₂ transferred through the films was assessed by a Mocon Coulux oxygen sensor in the N₂ gas flow. Two replicates were measured for each sample.

Results and Discussion

Parameterization

To perform the program, the parameters involved in equations must be estimated:

The permeability of the LDPE film was estimated independently by the cell permeability method. At 100% relative humidity and 20°C, CO₂ and O₂ permeability were respectively 1.87 and 4.14 amol x m⁻¹ x s⁻¹ x Pa⁻¹. These values did not change significantly when the relative humidity decreased data not shown and were in close agreement with the literature data for the same material (Page 198).

To design an oxygen-absorber equation, typical experimental data for AECOP®11140 compared with time are presented in Figure 1. The following absorption kinetic model was fitted to the experimental data and was a typical saturation exponential curve. The following mathematical model was then developed to express the number of oxygen molecules absorbed (Nₐₜ) compared with time:

Materials

Tomatoes (Solanum lycopersicum) were supplied from the Centre Technique Interprofessionnel des fruits et légumes (CTI) of Saint-Étienne-en-Provence (France) and were harvested by hand from the field. They were kept at 20°C under ambient air for 12 h before the experiments began.

Low density polyethylene film of 50 µm thickness was used (LDPE, films thickness: 50 µm,Micrometer). The oxygen absorbers, type AECOP®11140, were supplied by Sunco.

Published experimental data in scientific documents

Enriching an Ontological and Terminological Resource (OTR)

Quantitative data

{(LDPE, Packaging: Low Density Polyethylene),
(50 µm, Thickness: value: 50, unit concept: Micrometer),
(1078 amol x m⁻¹ x s⁻¹ x Pa⁻¹, O₂ Permeability: value: 1078, unit concept: Attomole per meter per second per pascal))

Localisation of relevant information (packaging characteristics)
Reduction of the search space in the text and enrichment of the termino-ontological resource by adding units of measure

Motivation

Automatic extraction from the text of quantitative data

• Water Vapor Permeability (WVP) of $1.81 \times 10^{-9}$ g m$^{-1}$ s$^{-1}$ Pa$^{-1}$

• O[2] permeability increasing from 7.12 to $7.68 \times 10^{-15}$ g·(Pa s m)$^{-1}$

Scientific locks

- Locate relevant information in the text
- Identify and extract units of measure taking into account their specific syntactical rules
Data capitalisation guided by an ontology
Concept description and associated guidelines

@Web graphical user interface
Data table annotation (pack example)

@Web graphical user interface
Data table annotation (bioraf example)

Manual Annotation of Comparison of sugar yields, crystallinity, and energy consumption between three pretreatments in various...

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sugar yields (%)</th>
<th>Crf (%)</th>
<th>Energy consumption (MJ/kg rice straw)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu</td>
<td>Xyl 23.4±2.3</td>
<td>Ara 18.7±2.7</td>
<td>Total 22.5±1.3</td>
</tr>
<tr>
<td>CM</td>
<td>Ara 20.7±1.4</td>
<td>Ara 51.9</td>
<td></td>
</tr>
<tr>
<td>DBM 5 min</td>
<td>Xyl 52.2±3.5</td>
<td>Ara 16.5±0.4</td>
<td>Total 41.8±2.3</td>
</tr>
<tr>
<td>DBM 15 min</td>
<td>Xyl 66.0±0.5</td>
<td>Ara 28.0±0.3</td>
<td>Total 54.5±0.4</td>
</tr>
<tr>
<td>DBM 30 min</td>
<td>Xyl 75.9±0.5</td>
<td>Ara 38.3±0.2</td>
<td>Total 64.4±0.4</td>
</tr>
<tr>
<td>DBM 60 min</td>
<td>Xyl 80.4±2.0</td>
<td>Ara 54.3±1.3</td>
<td>Total 78.2±1.7</td>
</tr>
<tr>
<td>HCWIT 160</td>
<td>Xyl 70.3±3.3</td>
<td>Ara 88.6±4.4</td>
<td>Total 74.1±3.6</td>
</tr>
</tbody>
</table>

Original table

Annotated table

<table>
<thead>
<tr>
<th>n°</th>
<th>Output solid constituent size Unit : mm</th>
<th>Treatment</th>
<th>Experience number Unit : 1</th>
<th>Process step number Unit : 1</th>
<th>Biomass</th>
<th>Biomass quantity Unit : g</th>
<th>Total pretreatment energy Unit : MJ/kg</th>
<th>Water quantity Unit : ml</th>
<th>Milling rotation speed Unit : min-1</th>
<th>Treatment duration Unit : min</th>
<th>Output solid constituent quantity Unit : g</th>
<th>Temperature Unit : °C</th>
<th>Output liquid constituent quantity Unit : ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.000e+0</td>
<td>Cutting milling</td>
<td>1.000e+0</td>
<td>1.000e+0</td>
<td>Rice straw 3.000e+1</td>
<td>[ -inf ; inf ]</td>
<td>0.000e+0</td>
<td>[ -inf ; inf ]</td>
<td>[ -inf ; inf ]</td>
<td>[ -inf ; inf ]</td>
<td>6.000e+1</td>
<td>3.000e+1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3.000e+0</td>
<td>Drying</td>
<td>1.000e+0</td>
<td>1.000e+0</td>
<td>Rice straw 3.000e+1</td>
<td>[ -inf ; inf ]</td>
<td>[ -inf ; inf ]</td>
<td>3.000e+1</td>
<td>3.000e+1</td>
<td>1.600e+2</td>
<td>3.000e+1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4.000e+0</td>
<td>Hot water treatment</td>
<td>1.000e+0</td>
<td>1.000e+0</td>
<td>Rice straw 3.000e+1</td>
<td>[ -inf ; inf ]</td>
<td>5.700e+0</td>
<td>3.000e+2</td>
<td>3.000e+1</td>
<td>4.320e+3</td>
<td>4.320e+3</td>
<td>4.670e+2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5.000e+0</td>
<td>Enzymatic hydrolysis treatment</td>
<td>1.000e+0</td>
<td>1.000e+0</td>
<td>Rice straw [4.000e-2 ; 6.000e-2 ]</td>
<td>[ -inf ; inf ]</td>
<td>0.000e+0</td>
<td>[ -inf ; inf ]</td>
<td>[ -inf ; inf ]</td>
<td>[ -inf ; inf ]</td>
<td>4.500e+1</td>
<td>4.500e+1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.000e+0</td>
<td>Cutting milling</td>
<td>1.000e+0</td>
<td>1.000e+0</td>
<td>Rice straw 3.000e+1</td>
<td>[ -inf ; inf ]</td>
<td>0.000e+0</td>
<td>[ -inf ; inf ]</td>
<td>[ -inf ; inf ]</td>
<td>[ -inf ; inf ]</td>
<td>4.500e+1</td>
<td>4.500e+1</td>
<td></td>
</tr>
</tbody>
</table>

2011-2014 --- V 06-03-2014
Data capitalisation guided by an ontology

EcoBioCap DSS

RDB

Web open data access

Web

<html>

PDF

Rech document (WOS, …)

Reliability assessment

Table and text extraction

Semantic annotation

Ontology

RDF DB
Annotated data table access using permalink

CO2 permeability (export)

<table>
<thead>
<tr>
<th>n°</th>
<th>CO2 Permeability Unit : kg.m.m-2.s-1.pa-1</th>
<th>Partial pressure difference Unit : atm</th>
<th>Packaging</th>
<th>Relative_Humidity Unit : %</th>
<th>Temperature Unit : °C</th>
<th>Thickness Unit : μm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[2.720e+17; 2.820e+17]</td>
<td>-inf; inf</td>
<td>poly(98% L-lactide)/Polylactic Acid</td>
<td>0.000e+0</td>
<td>2.500e+1</td>
<td>[0.000e+0 ; inf]</td>
</tr>
<tr>
<td>2</td>
<td>[1.930e+17; 2.050e+17]</td>
<td>-inf; inf</td>
<td>poly(98% L-lactide)/Polylactic Acid</td>
<td>0.000e+0</td>
<td>2.500e+1</td>
<td>[0.000e+0 ; inf]</td>
</tr>
<tr>
<td>3</td>
<td>[3.070e+17; 3.170e+17]</td>
<td>-inf; inf</td>
<td>poly(98% L-lactide)/Polylactic Acid</td>
<td>0.000e+0</td>
<td>3.000e+1</td>
<td>[0.000e+0 ; inf]</td>
</tr>
<tr>
<td>4</td>
<td>[2.230e+17; 2.350e+17]</td>
<td>-inf; inf</td>
<td>poly(98% L-lactide)/Polylactic Acid</td>
<td>0.000e+0</td>
<td>3.000e+1</td>
<td>[0.000e+0 ; inf]</td>
</tr>
<tr>
<td>5</td>
<td>[3.360e+17; 3.480e+17]</td>
<td>-inf; inf</td>
<td>poly(98% L-lactide)/Polylactic Acid</td>
<td>0.000e+0</td>
<td>3.500e+1</td>
<td>[0.000e+0 ; inf]</td>
</tr>
<tr>
<td>6</td>
<td>[2.460e+17; 2.580e+17]</td>
<td>-inf; inf</td>
<td>poly(98% L-lactide)/Polylactic Acid</td>
<td>0.000e+0</td>
<td>3.500e+1</td>
<td>[0.000e+0 ; inf]</td>
</tr>
<tr>
<td>7</td>
<td>[3.720e+17; 3.840e+17]</td>
<td>-inf; inf</td>
<td>poly(98% L-lactide)/Polylactic Acid</td>
<td>0.000e+0</td>
<td>4.000e+1</td>
<td>[0.000e+0 ; inf]</td>
</tr>
<tr>
<td>8</td>
<td>[2.830e+17; 2.910e+17]</td>
<td>-inf; inf</td>
<td>poly(98% L-lactide)/Polylactic Acid</td>
<td>0.000e+0</td>
<td>4.000e+1</td>
<td>[0.000e+0 ; inf]</td>
</tr>
<tr>
<td>9</td>
<td>[4.030e+17; 4.330e+17]</td>
<td>-inf; inf</td>
<td>poly(98% L-lactide)/Polylactic Acid</td>
<td>0.000e+0</td>
<td>4.500e+1</td>
<td>[0.000e+0 ; inf]</td>
</tr>
<tr>
<td>10</td>
<td>[3.290e+17; 3.410e+17]</td>
<td>-inf; inf</td>
<td>poly(98% L-lactide)/Polylactic Acid</td>
<td>0.000e+0</td>
<td>4.500e+1</td>
<td>[0.000e+0 ; inf]</td>
</tr>
</tbody>
</table>
SPARQL querying of the RDF base

Query Summary

Query scope

Ontology
IC2ACV

Topics

Relation
Biomass glucose composition relation

Value domains wanted for attributes

Mandatory
(1) Biomass : [ Grasses and energetic plants : 1 ]

Desirable
(1) Glucose rate : [0 ; 100 ; 100 ; 100] - unit: Percent

Parameters
(default parameters)

Define numeric value domain

Run query

@Web graphical user interface
<table>
<thead>
<tr>
<th>rank</th>
<th>reliability score</th>
<th>Biomass [Grasses and energetic plants]</th>
<th>Glucose rate [0.000e+00; 1.000e+02; 0.000e+02; 1.000e+02].%</th>
<th>Biomass state</th>
<th>Experience number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.2317</td>
<td>Rice straw</td>
<td>[5.333e+01; 5.600e+01].%</td>
<td>Untreated biomass</td>
<td>[2.000e+00].1</td>
</tr>
<tr>
<td>2</td>
<td>2.2523</td>
<td>Bagasse</td>
<td>[4.666e+01].%</td>
<td>Untreated biomass</td>
<td>[3.000e+00].1</td>
</tr>
<tr>
<td>3</td>
<td>1.2488</td>
<td>Rice straw</td>
<td>[4.633e+01].%</td>
<td>Untreated biomass</td>
<td>[1.000e+00].1</td>
</tr>
<tr>
<td>3</td>
<td>0.2488</td>
<td>Rice straw</td>
<td>[4.633e+01].%</td>
<td>Untreated biomass</td>
<td>[0.000e+00].1</td>
</tr>
<tr>
<td>4</td>
<td>5.2506</td>
<td>Rice straw</td>
<td>[4.522e+01].%</td>
<td>Untreated biomass</td>
<td>[5.000e+00].1</td>
</tr>
<tr>
<td>5</td>
<td>5.2546</td>
<td>Rice straw</td>
<td>[4.522e+01].%</td>
<td>Untreated biomass</td>
<td>[5.000e+00].1</td>
</tr>
</tbody>
</table>

@Web graphical user interface
Conclusion

- A generic and reusable ontological model to capitalize experimental data
- Ontology consistency management
- Data reliability assessment
- Manual annotation guided by the ontology using text-mining assistant
- Flexible querying of annotated data combining 3 kinds of reasoning (specialization, fuzzy pattern matching, reliability satisfaction)
Future works

- CSV data file import
- RDF DB consistency management in case of ontology evolution
- Ontology mapping for Linked Open Data
- Assistant development to facilitate the manual annotation work
  - Text mining approach
  - Guidelines formalization using rules (OBDA approach)
- Extending core ontology to represent semantic links between n-ary relation concepts (temporal links, …)
Mapping to DOLCE upper ontology

- **Chemical reagent**
- **Biomass**
- **Quantity**
- **Treatment**
- **Treatment expe**
- **Biomass characterization**
Which varieties of grapes having a “high” rate of tannin extraction from marc (vinification co-product) and “good” wine color parameters?
Modeling Guidelines associated with processes as rules

- Topic Bioref-PM. This Topic must contain experiences with only one milling followed by the enzymatic hydrolysis. It does not include a physico-chemical step but it can include a washing and separation step. (en)
- Topic Bioref-PM-PC-EX-PS. This Topic must contain experiences composed of a pre-milling step, then a physico-chemical treatment and an extrusion treatment and finally a press and separation step (washing and filtration) followed by the enzymatic hydrolysis step. (en)
- Topic Bioref-PM-PC-PS. This Topic must contain experiences composed of at least one pre-milling step, then a physico-chemical treatment and a washing and filtration step (washing and separation) and finally the enzymatic hydrolysis step. (en)
- Topic Bioref-PM-PC-UFM. This Topic must contain experiences composed of a pre-milling step, then a physico-chemical treatment followed by an ultrafine milling step (ball milling...) and finally the enzymatic hydrolysis step. This topic doesn’t require a step of press and separation because it’s a process with a low intake of effluent. The second milling step must give an “Output solid constituent size” smaller than 1 mm. (en)
- Topic Bioref-PM-PC-UFM-PS. This Topic must contain experiences composed of a pre-milling step, then a physico-chemical treatment followed by an ultrafine milling step (wet milling...) and a press and separation step (washing and filtration) and finally the enzymatic hydrolysis step. This topic requires a press and separation step because there are a lot of effluents in the physico-chemical step or because the milling is made with effluent. The second milling step must give an “Output solid constituent size” smaller than 1 mm. (en)
La gestion de l'évolution

Implémentation

Rim Touhami
La gestion de l'évolution

Implémentation

Rim Touhami

Changes:

Relation changes  Argument changes  Parameters

Evolution strategies

How to deal orphan concepts?

Orphan concepts are:
- deleted
- reconnected to their parents
- reconnected to the root concept

How to deal orphan properties?

Orphan properties are:
- deleted
- reconnected to their parents
- left alone

How to deal restrictions with undefined class in their definition?

Restrictions are:
- deleted
- updated: replace the class representing the restricted range with its subclasses
- updated: replace the class representing the restricted range with one or more subclasses selected by the user

How to deal restrictions with undefined property in their definition?

Restrictions are:
- deleted
- updated: replace the property used in the restricted range with one property selected by the user

How to deal instances whose concept is deleted?

Instances are:
- deleted
- reconnected to the parents

How to deal instances whose property is deleted?

Instances are:
- deleted
- defined for the parent properties

How to deal properties with undefined class in their domain/range?

Undefined class is:
- deleted from the list of domains or co-domains
- replaced by its subclasses

How to deal labels whose concept is deleted?

Labels are:
- deleted
- reconnected to the subclasses of the deleted concept as alternative labels
- reconnected as alternative labels to one or more subclasses selected by the user

OK  Reset
## Querying the RDF base

### Query Summary

<table>
<thead>
<tr>
<th>Query scope</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ontology</strong></td>
<td>MAP0PT</td>
</tr>
<tr>
<td><strong>Topics</strong></td>
<td>&quot;PackPermeability&quot;</td>
</tr>
<tr>
<td><strong>Relation</strong></td>
<td>O2 Permeability_relation</td>
</tr>
</tbody>
</table>

### Value domains wanted for attributes

**Mandatory**

1. O2 Permeability: [5e-13; 1.27e-11; 1.5e-11; 1e-10] - unit: Mole per Meter per second per pascal

### Paramotors

- Run query
<table>
<thead>
<tr>
<th>rank</th>
<th>reliability score</th>
<th>O2 Permeability</th>
<th>Temperature</th>
<th>Thickness</th>
<th>Relative Humidity</th>
<th>Packaging</th>
<th>Partial pressure difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>![Green icon]</td>
<td>[2.950e+01; 3.150e+01] cm^3 mm m^-2 day^-1 atm^-1</td>
<td>[2.400e+01; 2.500e+01] °C</td>
<td>[2.000e+02; 2.200e+02] μm</td>
<td>[\inf, +\inf] %</td>
<td>Polylactic acid</td>
<td>[1.000e+00] atm</td>
</tr>
<tr>
<td>1</td>
<td>![Green icon]</td>
<td>[2.290e+01; 2.490e+01] cm^3 mm m^-2 day^-1 atm^-1</td>
<td>[2.400e+01; 2.600e+01] °C</td>
<td>[2.000e+02; 2.200e+02] μm</td>
<td>[\inf, +\inf] %</td>
<td>(1wt%)Ag/Polyactic Acid</td>
<td>[1.000e+00] atm</td>
</tr>
<tr>
<td>1</td>
<td>![Green icon]</td>
<td>[1.680e+01; 1.880e+01] cm^3 mm m^-2 day^-1 atm^-1</td>
<td>[2.400e+01; 2.600e+01] °C</td>
<td>[2.000e+02; 2.200e+02] μm</td>
<td>[\inf, +\inf] %</td>
<td>(5wt%)Cellulose nanocrystals/(1wt%)Ag/Polyactic Acid</td>
<td>[1.000e+00] atm</td>
</tr>
<tr>
<td>1</td>
<td>![Green icon]</td>
<td>[7.000e+01; 8.000e+01] cm^3 mm m^-2 day^-1 atm^-1</td>
<td>[2.300e+01] °C</td>
<td>[8.500e+01] μm</td>
<td>[0.000e+00, +\inf] %</td>
<td>Chitosan/paper</td>
<td>[1.000e+00] atm</td>
</tr>
<tr>
<td>1</td>
<td>![Green icon]</td>
<td>[1.150e+01; 1.350e+01] cm^3 mm m^-2 day^-1 atm^-1</td>
<td>[2.400e+01; 2.600e+01] °C</td>
<td>[2.000e+02; 2.200e+02] μm</td>
<td>[\inf, +\inf] %</td>
<td>(5wt%)Modified cellulose nanocrystals/(1wt%)Ag/Polyactic Acid</td>
<td>[1.000e+00] atm</td>
</tr>
<tr>
<td>1</td>
<td>![Green icon]</td>
<td>[1.320e+01; 1.520e+01] cm^3 mm m^-2 day^-1 atm^-1</td>
<td>[2.400e+01; 2.600e+01] °C</td>
<td>[2.000e+02; 2.200e+02] μm</td>
<td>[\inf, +\inf] %</td>
<td>(1wt%)Modified cellulose nanocrystals/(1wt%)Ag/Polyactic Acid</td>
<td>[1.000e+00] atm</td>
</tr>
</tbody>
</table>
Learning corpus

Model

Set of descriptors

New similarity measure $SM_{DB}$

Reduction of the search space for unit of measure

Iterative enrichment of the ontology

Raw corpus

Learning on Packaging corpus

Biorefinery Corpus

Semi-automatic method

Identification of new units of measure
### Data

<table>
<thead>
<tr>
<th>«Corpus Biorefinery»</th>
<th>«Corpus Packaging»</th>
</tr>
</thead>
<tbody>
<tr>
<td>243 scientific documents</td>
<td>115 scientific documents</td>
</tr>
<tr>
<td>Unit terms extracted from the ontology: 36 terms</td>
<td>Unit terms extracted from the ontology: 211 terms</td>
</tr>
</tbody>
</table>

### Locate relevant information by learning

<table>
<thead>
<tr>
<th>Reduction of «Corpus Biorefinery»</th>
<th>Reduction of «Corpus Packaging»</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>86%</td>
</tr>
</tbody>
</table>

### Enrichment of the termino-ontological resource

<table>
<thead>
<tr>
<th>Enrichment «OTR Bioraffinerie»</th>
<th>Enrichment «OTR Packaging»</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ than 100% of new units</td>
<td>18% of new units</td>
</tr>
</tbody>
</table>