A termino-ontological resource to compare ligno-cellulosic biomass and agro-waste valorisation routes

Amélie Breysse[1], Abdelatif Barakat[1], Hélène Angellier-Coussy[1], Claire Mayer-Laigle[1], Patrice Buche*[1,2]

1. *JRU IATE, INRA, Montpelier University, Montpellier, France;*
2. *LIRMM and INRIA GraphiK, Montpellier, France*

Corresponding author: patrice.buche@inra.fr

ABSTRACT

Lignocellulosic biomass and agro-waste valorization routes are two of the promising methods towards a more sustainable bio-economy. Scientific literature in this domain is increasing fast and could be a valuable source of data. As these abundant scientific data are mostly in textual format and heterogeneously structured, using them to compute biomass treatment efficiency is not straightforward. The implementation of a ***Decision Support System*** (DSS) based on an original pipeline coupling knowledge management (KM) based on semantic web technologies, soft computing techniques and environmental factor computation has been already done for lignocellulosic biomass valorization routes into glucose [Lousteau-Cazalet et al. 2016]. The DSS allows using data found in the literature to assess environmental sustainability of biorefinery systems. The pipeline permits to: (1) structure and integrate relevant experimental data, (2) assess data source reliability [Destercke et al. 2013], (3) compute and visualize green indicators taking into account data imprecision and source reliability. This pipeline has been made possible thanks to innovative researches in the coupling of ontologies, uncertainty management and propagation. In this first version, data acquisition is done by experts and facilitated by a ***Termino-Ontological Resource*** (TOR) called *Biorefinery* and available at http://www6.inra.fr/cati-icat-atweb/Ontologies/Biorefinery. Data source reliability assessment is based on domain knowledge and done by experts. The operational prototype has been used by field experts on realistic use cases (glucose extraction for biofuel production [Licari et al. 2016]). The obtained results have validated the usefulness of the system.

Keywords: Web of data, data standard in agriculture; knowledge management, reliability and uncertainty assessment, uncertainty management, system performance and environmental impact assessment, France

1. INTRODUCTION

The *Biorefinery* TOR is a formal specification of the concepts and relationships shared by a community of experts, in order to enable knowledge exchange and reuse [Gruber, 1993]. *Biorefinery* TOR has been recently extended to manage data acquisition about new agro-waste valorization routes to extract additional constituents namely particles dedicated to food packaging material reinforcement. This work has been done with a new group of experts which has validated and extended *Biorefinery* TOR.

2. BIOREFINERY TERMINO ONTOLOGICAL RESOURCE

The communication will be dedicated to the extended version of *Biorefinery* TOR, which is based on a generic TOR model to represent scientific experiments in order to annotate data tables in a given domain (see Buche et al., 2013 for more details). We made the choice to represent an experiment by using n-ary relations between several experimental parameters and a given result. This structures information in a simple way. As recommended by W3C [Noy et al. 2006], we used the design pattern which represents a n-ary relation thanks to a concept associated with its arguments via properties. Let us illustrate this notion by using the example of n-ary relation *Biomass_Glucose_Composition_Relation*. This relation is characterized by 4 arguments: 1) the glucose rate, which is the experimental result, 2) the biomass, which is the studied object, and associated experimental parameters being 3) the biomass state (untreated or treated) and 4) the experiment number reported in the document. This relation is used to create annotated tables whose data are extracted from scientific documents (by example peer-reviewed papers). The *Biorefinery* TOR is composed of a conceptual component and a terminological component. The conceptual component of *Biorefinery* TOR is composed of a **core ontology** to represent n-ary relations between experimental data and a **domain ontology** to represent specific concepts of a given application domain**.** In the Up core ontology, generic concepts **Relation** and **Argument** represent respectively n-ary relations and arguments. The representation of n-ary relations between experimental data requires a particular focus on the management of quantities and their associated units of measure. In the Down core ontology, generic concepts **Dimension**, **UM_Concept**, **Unit_Concept** and **Quantity** allow the management of quantities and their associated units of measure. The sub-concepts of the generic concept Symbolic_Concept represent the non-numerical arguments of n-ary relations between experimental data. The domain ontology contains specific concepts of a given application domain, the *Biorefinery* domain. They appear as sub concepts of the generic concepts of the core ontology. In the *Biorefinery* TOR, relations represent either experiments which characterize biomass (by example the *Biomass_Glucose_Composition_Relation*) or experiments involving unit operations performed on biomass. For instance, the milling operation represented by the n-ary relation *Milling_Solid_Quantity_Output_Relation* is characterized by 7 arguments and represents the milling solid quantity output, which is the milling experimental result for a given biomass associated with a set of controlled parameters being the biomass input quantity, the total treatment energy used for the milling, the treatment duration, the milling rotation speed and the type of milling. In the *Biorefinery* TOR conceptual component part, all concepts are represented as OWL classes, hierarchically organized by the subsumption relation **subClassOf** and pairwise disjoints. The terminological component of the *Biorefinery* TOR contains the domain-related set of terms used to annotate data tables. Sub concepts of the generic concepts Relation, Symbolic_Concept and Quantity, as well as instances of the generic concept Unit_Concept, are all denoted by at least one term of the terminological component. Each of these sub concepts or instances are, in a given language, denoted by a preferred label and optionally by a set of alternative labels, which correspond to synonyms or abbreviations. Labels are associated with a concept or an instance thanks to SKOS[1] labelling properties, recommended by W3C to represent controlled vocabularies associated with.

Extended version of the *Biorefinery* TOR is currently composed of 52 Relation concepts, 69 Quantity concepts, 213 Symbolic concepts and 56 Unit instances of concepts. 10015 instances of

[1] Simple Knowledge Organization System

relation concepts have been created extracted from more than 36 publications representing more than 405 experimentations.

ACKNOWLEDGEMENT

The research leading to these results is part of **NoAW project** that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688338.

REFERENCES

Patrice Buche, Juliette Dibie-Barthélemy, Liliana Ibanescu, Lydie Soler (2013): Fuzzy Web Data Tables Integration Guided by an Ontological and Terminological Resource. IEEE Trans. Knowl. Data Eng. 25(4): 805-819

Sébastien Destercke, Patrice Buche, Brigitte Charnomordic (2013) Evaluating Data Reliability: An Evidential Answer with Application to a Web-Enabled Data Warehouse. IEEE Trans. Knowl. Data Eng. 25(1): 92-105

Charlotte Lousteau-Cazalet, Abdellatif Barakat, Jean-Pierre Belaud, Patrice Buche, Guillaume Busset, Brigitte Charnomordic, Stephane Dervaux, Sebastien Destercke, Juliette Dibie, Caroline Sablayrolles, Claire Vialle (2016) Decision support system for eco-efficient biorefinery process comparison using a semantic approach. Computers and Electronics in Agriculture 127:351-367.

Charlotte Lousteau-Cazalet, Abdellatif Barakat, Jean-Pierre Belaud, Patrice Buche, Guillaume Busset, Brigitte Charnomordic, Stephane Dervaux, Sebastien Destercke, Juliette Dibie, Caroline Sablayrolles, Claire Vialle (2016) A decision support system using multi-source scientific data, an ontological approach and soft computing - Application to eco-efficient biorefinery. FUZZ-IEEE 2016.

Anaïs Licari, Florian Monliau, Abderrahim Solhy, Patrice Buche, Abdellatif Barakat (2016) Comparison of various milling modes combined to the enzymatic hydrolysis of lignocellulosic biomass for bioenergy production: Glucose yield and energy efficiency. Energy 102:335-342.

Gruber, T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition*, *5*(2), 199–220.

Noy, N., Rector, A., Hayes, P., Welty, C., 2006 Defining n-ary relations on the semantic web. W3C working group note http://www.w3.org/TR/swbp-n-aryRelations.