Computers and Chemical Engineering 100 (2017) 219–231

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

A semantic framework for enabling model integration for biorefining

Linsey Koo[a,*], Nikolaos Trokanas[b], Franjo Cecelja[a]

[a] Centre for process & Information Systems Engineering, University of Surrey, Guildford, GU2 7XH, UK
[b] Centre for International Manufacturing, Institute for Manufacturing (IfM), University of Cambridge,17 Charles Babbage Road, Cambridge, CB3 0FS, UK

ARTICLE INFO

Article history:
Received 25 October 2016
Received in revised form 11 January 2017
Accepted 3 February 2017
Available online 11 February 2017

Keywords:
Ontology engineering
Model integration
Computer aided process engineering (CAPE)
Biorefining

ABSTRACT

This paper introduces a new paradigm for establishing a framework that enables interoperability between process models and datasets using ontology engineering. Semantics are used to model the knowledge in the domain of biorefining including both tacit and explicit knowledge, which supports registration and instantiation of the models and datasets. Semantic algorithms allow the formation of model integration through input/output matching based on semantic relevance between the models and datasets. In addition, partial matching is employed to facilitate flexibility to broaden the horizon to find opportunities in identifying an appropriate model and/or dataset. The proposed algorithm is implemented as a web service and demonstrated using a case study.

1. Introduction

In computer aided process engineering (CAPE) community, increased availability of mainstream commercial and free simulation software, as well as data from laboratory experiments or pilots to near commercial scale plants, has facilitated the development of a large number of custom-made models. As, historically, most of the models were developed to represent petrochemical processes, modelling and simulation for biorefining processes are still facing challenges due to lack of biochemical property data, complexity of feedstock characterisation, as well as a constant influx of new processes and technologies or adaptation to new environments. To develop an understanding of biochemical processes or to provide suitable design, development of a database system to support modelling and analysis of biochemical processes is vital. The development of these models, as practice has demonstrated, goes along the development of new models, integration and/or adaptation of existing models, or most commonly the combination of the two.

To increase reusability of existing models that are developed in disparate software tools and process simulators, CAPE-OPEN was initiated to conceptualise and develop a set of interface specifications as a method pertaining interoperability standard (Braunschweig et al., 2004; Morales-Rodríguez et al., 2008; Pons 2010). As such, CAPE-OPEN is a widely recognised standard which defines the interconnection representation of interfaces facilitated by a middleware service as a communication hub across heterogeneous software environments (Braunschweig et al., 2000; Bogusch et al., 2000). To take full advantage of reusability of existing models, the task of identifying the most sufficient model from the libraries is heavily dependent on the user's intuition and experience and remains as a manual process (Braunschweig et al., 2004). Yang et al. (2008) acknowledges that inadequate assessment for the suitability of models may lead to potential misuse of the models, which has the risk of insufficient or even wrong solution to the engineering problems. To better address the shortcoming associated with user intervention in CAPE-OPEN, ontology engineering is recognised as a viable solution to reduce the chance of these error occurring and to minimise the impact of any errors that do occur. Ontology has an ability to address the problem of automated support for the configuration of process models and data in a structured and proactive manner (Yang and Marquardt 2004; Yang et al., 2008) by accounting for complex relations, such as systematic knowledge of model as well as tacit knowledge extracted from user intuition. A large scale ontology, the OntoCAPE, has as a result been introduced to support various process engineering applications, mainly addressing two aspects: i) characterisation of models stored in the libraries and ii) description of the specific requirements of the models to be identified as potential candidates. To address the reconciliation of interoperability between process modelling components, COGents was developed to perform the registration and integration of the models stored in the libraries. This method was the first attempt to integrate process modelling components from heterogeneous sources using ontologies as a tool in the field of process engineering. As indicated by Yang et al. (2008), the integration of the models they

* Corresponding author.
E-mail address: m.koo@surrey.ac.uk (L. Koo).

http://dx.doi.org/10.1016/j.compchemeng.2017.02.004

used was based on the full-scale matching. Partial matching which extends the search scope was first introduced by the eSymbiosis project to enable and hence to support processing technologies participation in Industrial Symbiosis (IS) and concomitant integration (Raafat et al., 2012; Raafat et al., 2013; Cecelja et al., 2015). The framework employed semantic technologies to automate widely used manual procedure of synergy identification of IS by finding the semantic relevance of participant's profile based on practical experience in the form of tacit knowledge and explicit knowledge acquired from users. The measure of semantic relevance requires obtaining appropriate description of processing technology to further use in the discovery process. To support these processes, the process of IS was semantically formulated in an IS domain ontology (Trokanas et al., 2012). Recently, a number of ontologies have been developed in the domain of biorefining, which focuses on the knowledge representation of biomass and bioprocessing technologies (Trokanas and Bussemaker, 2015) and process systems design and optimisation of biorefining processes (Siougkrou and Kokossis 2016; Magioglou et al., 2015). These ontologies, however, although in the domain of biorefining do not address the process of model and data integration, and, to the best of our knowledge, they are not yet available in public domain for reuse.

Following on previous developments and use of ontology to address challenges of identifying most suitable model or data to achieve the best solution for a particular engineering problem, ontology engineering is employed to describe them in a comprehensive manner to distinguish between them (Koo and Cecelja 2015; Koo et al., 2016). It has been demonstrated that the differences of models and data can be addressed by explicit descriptions using defined terms to further improve consistency as well as understanding of the heterogeneity and concomitant consequences. The semantically enriched and reconciled process models and data are then applicable to facilitate semantic interoperability between them. The semantic interoperability is achieved by employing different matchmaking algorithms to benefit from partial matching to measure a meaningful similarity between models that are not identical. We argue that this approach allows to improve the decision making process and broaden the horizon to find opportunities in identifying appropriate models and/or datasets whilst increasing awareness of existing models.

This paper proposes a new paradigm for model and data integration with focus on biorefining and which is built around the ontology to i) model tacit knowledge in the domain of biorefining including the advances in biorefining process, biomaterial and technologies classifications, and ii) model explicit knowledge which includes a complete set of model input, output and auxiliary parameter properties, as well as known and otherwise identified potential model and data integration solutions. Tacit knowledge is built in the ontology structure (Cecelja et al., 2015), i.e. subsumption and object properties with respective and domain dictated restrictions. Explicit knowledge is captured during the instantiation process from data collected on model/data entities presented as ontology instances and characterised by input, output and auxiliary parameter properties. The proposed ontology enables instance matching with the view of model integration, expanding knowledge base, generating new knowledge in the process of model integration for biorefining, and knowledge sharing. Designed ontology is open to further development in response to advances in the domain of biorefining. The proposed matching algorithm is tuned to match models and data based on i) tacit knowledge formulation to observe process synthesis logic by employing semantic distance measurements between the two or more instances of the ontology, and ii) explicit knowledge formulation by employing similarity calculation between input/output parameters of candidate models/data identified suitable for integration. In addition, matching process allows for recursive matching towards complex model/data integration solutions, matching for integration of models developed in heterogeneous software environments to generate a meaningful solution for particular engineering tasks, as well as for partial matching to broaden the search domain and to find comparable replacement model rather than focusing only on an exact match. This paper explicitly formulates theoretical concept of knowledge model and design of ontology and matching algorithm, as well as auxiliary conditions used in the process of model/data integration. The usefulness and operation of the proposed formalism is demonstrated by a case study to guide the user to make an informed decision by taking into consideration of users' intuition and their experience in modelling.

2. Theoretical concepts of model and data integration

2.1. Model and data representation

A process model represents a part of the actual system in which physical and chemical processes are taking place and describes the behaviour of a process system within well-defined boundaries together with inputs and outputs and under certain environmental conditions as a requirement (Hangos and Cameron 2001). The process models used to address process modelling, simulation and optimisation problems are arguably classified into two distinct types i) sequential modular models, and ii) equation based models. Sequential modular models represent individual units as a pre-configured block model where modelling equations are grouped to represent a particular process equipment. The sequence of calculation is initiated from one unit to the next in the process flowsheet through the process streams that connect the units using thermodynamics and physical property calculations. Equation based models are considered as custom modelling packages which have a set of equations from the various units in the process into a single large set to be solved.

Each model is semantically described by its type, i.e. its functionality in terms of the process and/or unit it represents. In addition, each model is (semantically) described by requirements and other characteristics that form a comprehensive knowledge model (Koo et al. (2016)) which includes model input(s), output(s), precondition(s), and the environment in which each process model operates (Trokanas et al., 2014; Trokanas et al., 2015). The inputs and outputs are not limited to physical properties and can be extended to additional data or other properties. The number of output variables can be purposely adjusted or extended to include additional data or parameters to consider the dynamic nature of models. Contrary to the models, data is semantically annotated with regards to output(s), functionality, and precondition(s) required to process data (Koo et al., 2016).

2.2. Concept of model integration

The integration of model and data is a process of assembling heterogeneous tools and methods to generate new knowledge that is meaningful and useful for particular engineering tasks. The CAPE-OPEN interface specification (Belaud and Pons 2002) is developed as a standard requirement for the unit operation components (such as process unit operation, thermodynamics, and numerical solvers packages) to be compliant with any simulator without modification, compiling, or linking. The standard mainly provides the details for the interface specifications of sequential modular simulators and the granularity of the interface design was restricted to the unit operation level (CAPE-OPEN Project team 2000; van Baten and Pons 2014).

The structure of unit is configured by a coupling through the different inlet and outlet ports where a unit can be connected to