Krist V. Gernaey, Jakob K. Huusom and Rafiqul Gani (Eds.), 12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering.
31 May – 4 June 2015, Copenhagen, Denmark

Model Integration Using Ontology Input-Output Matching

Linsey Koo, Franjo Cecelja

PRISE, FEPS, University of Surrey, Guildford, Surrey. GU2 7XH, UK

Abstract

This paper introduces ontology controlled model integration framework using input-output matching in the domain of biorefining. The framework builds upon the existing framework and replaces the Common Object Request Broker Architecture (CORBA) object bus with more flexible semantic repository. Semantic Web Services Description Ontologies (OWL-S) are used to describe model inputs, outputs, preconditions, operating environment and its functionality. The OWL-S enables the automation of model integration through (i) discovery, (ii) selection, (iii) composition, and (iv) execution stages. This concept has been verified with a small scale model integration to demonstrate the flexibility of model integration through all four stages of the process.

Keywords: Model Integration, Ontology, Input-Output Matching

1. Introduction

Process System Engineering (PSE) and Computer Aided Process Engineering (CAPE) have traditionally been concerned with the development of systematic procedures and computer aided techniques for the design, control and operation of chemical process systems. Problems related to process optimisation, process integration, and process synthesis/design have been solved through the use of knowledge, and optimisation tools and methods, developed by the PSE/CAPE community. The traditional users of these tools and methods include the oil & gas industry, petrochemical industry, and the chemical industry (Alvarado-Morales et al., 2008).

Process modelling and simulation have become vital tools for process engineers, in order to plan, evaluate, assess, and develop different alternatives for the design of products and processes. In the process of developing a new model, the best suited modelling tools for different parts of the process are employed. However, to understand process design as a continuous work process from an integrated perspective requires the use of tools from diverse sources and disciplines simultaneously. The only existing model integration framework is the CAPE-OPEN, proposed by the European process industry which established a software based interface, Common Object Request Broker Architecture (CORBA). More precisely, CORBA provides a Common Interface Bus that allows individual models to reside in multiple computer systems and communicate with each other. Whilst the CAPE-OPEN standard is a widely recognised standard, a disadvantage has been identified in the lack of flexibility in input-output (I-O) architecture.

The exploitation and use of sustainable renewable resources have been recognised as an essential component in meeting future energy demand needs. Biofuels are alternative fuel sources to traditional petroleum based fuels that could be used to meet future demands. However, mainstream biofuel production faces a number of technical challenges due to the complexity of the characteristics of biomass feedstock, and associated processing technologies, which have very different characteristics when compared to traditional petroleum processes. Gani and Grossmann (2007) addressed the importance of developing methods and tools that represent comprehensive biorefining technologies, through the adaptation of current systems and development of new systems. They can be further integrated by a single, flexible, reliable, and efficient system that has an ability to integrate methods and tools from different sources, to provide meaningful and useful simulation and optimisation tools.

In response to the challenges stated, this paper introduces the use of ontology in biorefining as a method of model integration, which has strong synthesis capabilities and functions to invite degrees of freedom. In particular, the ontology controlled model integration builds upon the existing CAPE-OPEN framework and replaces the CORBA object bus with a more flexible semantic repository, as shown in Figure 1. Models are described by Semantic Web Services (SWS) using Ontology Web Service Description (OWL-S) as an enabler of web services through service discovery, selection, composition, and execution stages. The Web Ontology Language (OWL) defines three upper ontologies, including *ServiceProfile*, *ServiceGrounding*, and *ServiceModel*. *ServiceProfile* ontology represents what a service does, and determines whether the service/model meets the requirement. *ServiceGrounding* ontology supports the details of how to access a service/model. Finally, *ServiceModel* ontology defines how the service/model works through the description of the work flow and the potential interoperation paths.

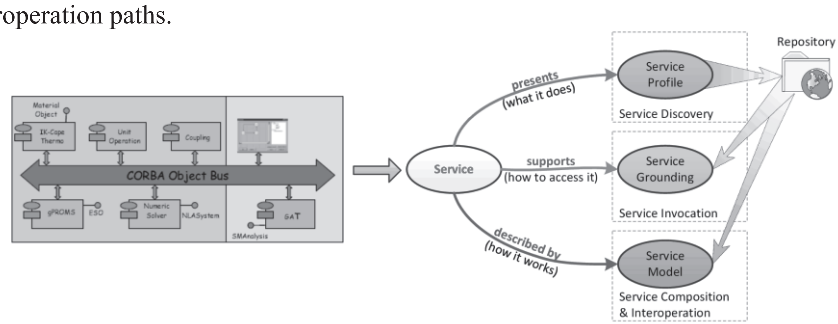

Figure 1. Introduction of repository for model integration

2. Ontological Approach to Model Integration

2.1. Basic Concept of Domain Ontology based on OWL-S Framework

Ontology is a set of interlinked common concepts and relationships between concepts, which defines the knowledge, both tacit and explicit, in a specific domain (Gruber, 1995). The synergistic relationship and interplay between tacit and explicit knowledge developed in the ontology domain, enable the generation of new knowledge. The domain ontology of biorefining, with a particular view to coordinate model integration, provides a common set of vocabulary to describe the model and data, and characterises functionality in relation to the processes of biorefining. Each model is described by the SWS to allow the models to be semantically annotated via the use of concepts from domain ontology, to form a comprehensive knowledge base which includes; model input(s), output(s), precondition(s), the environment in which it operates, as well as the functionality it performs (Raafat et. al., 2013a; Trokanas et. al., 2014). The semantically described model is then registered as an instance of the domain ontology and implemented in OWL-S framework. The respective instances are published in the purposely built public repository for I-O matching with other available models, as shown in Figure 2.

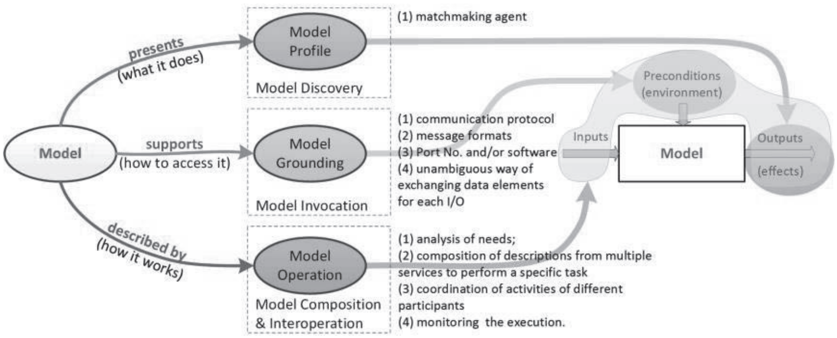

Figure 2. Semantically described model

The particular role of OWL-S ontology is to facilitate the automation of model integration through the model discovery, selection, composition, and execution stages. The discovery stage allows formation of an integrated model to organise repositories and supports the automated process of locating a model from a public repository(ies). The best option that satisfies the requestor's functionality is selected in the model selection stage. Then, the model composition stage formulates the chain of integrated models in response to the requestor's functionality. The instances of the model are created, executed and managed during the execution stage. The model selection and model composition stages are performed by semantic I-O matching (Trokanas et. al., 2014), which allows for partial matching and permits a high degree of flexibility in model integration. It is also important to mention that the proposed framework allows for integration of data residing in various forms of databases, and is concomitantly semantically described in the same way as the model with the omission of the input. The performance of the concept has been tested by large scale technology integration to satisfy requirements of industrial symbiosis, where technologies were represented by respective models (Raafat et. al., 2013b).

In a biorefining domain, a simplified model is used, which includes the conversion process itself as a black box model, desired input, targeted output streams in the form of characterisation of resources, and the environment supporting the unit operation, i.e. operating condition or software platform. To establish a common set of vocabulary and to enable the discovery of models and data that characterises functionality, a classification of biorefining is conceptualised based on two distinct concepts, which include conversion technology and resource. The term 'conversion technology' in biorefining refers to a wide array of state-of-art processes that are capable of producing value-added products, such as biofuels and biochemicals from biomass. Conversion technology has been characterised by the biorefinery platform (i.e. sugar platform, thermochemical platform), the process stage (i.e. pretreatment stage, conversion stage, separation stage), the resultant products in terms of material (i.e. biofuel, biochemical), and energy (i.e. heat, electricity). The term 'resource' in biorefining refers to both energy and material, which can be described as an input(s), output(s) and precondition(s) of the conversion process. Resources have been further characterised, by the type of material (i.e. electricity, heat, inorganic compound, organic compound) and the characteristics of property (i.e. physical property, chemical property) so that resources are processed based on the classification of the conversion process.

The relation in ontology describes the semantics of the domain between the classes and individuals. For example, the relationship *hasComposite* creates a link between type of material and characteristics of property to further describe the mereology relation. In order to encompass additional semantic information concerning the concept, the properties that link between concepts and its data value that assist characterisation are used to define a precise characteristics, i.e. properties *hasSize* and *hasMoistureContent*. In addition, the semantic restrictions on properties provide the following: (i) the I-O matching by calculating similarity, (ii) the user navigation process through cardinality restrictions, and (iii) the inferences.

2.2. Semantic Matching

The model integration and creation of interoperability between methods and tools from different sources is referred to as the I-O matching. This allows for the automation process, and built in intelligence for automated discovery of potential and relevant models that will support model integration (Raafat et. al., 2013b).

The process of integration is performed by a purposely built matchmaker, which matches inputs and outputs of all models and data available in the public repository. There are two techniques to enable the matching, being forward and backward matching processes. The forward matching process is initialised to satisfy the requester's input parameters through an expansion, by matching inputs of the rest of the services with the outputs of the last identified service. The backward matching process is generated by identifying a suitable model, which satisfies the requested outputs. The backward matching process, compared to forward matching, avoids the processes of expansion, and can therefore be considered more efficient. After matching is completed, the respective model and data integration is reversed and perform a forward integration process. The whole process is visualised for better control and to allow intervention by the user as well as for result recording and analysis.

The semantic partial matching is considered in order to facilitate the flexibility of model integration, when the registered model is partially satisfying the input criteria. The partial matching is established through a direct matching process using semantic similarities between I-O type and property. The semantic measure for the matching by type of I-O is calculated by the distance between the respective concepts, along the class-subclass relationships in the domain ontology. The property matching is calculated from the object properties defined in respective SWS ontology. The method of measuring the similarity of the properties, which was established by industrial symbiosis case (T. Raafat et. al., 2013b), was adapted to calculate the property similarity between the property sets. Each set of property is represented as a vector:

$$p_j = \{v(d_1)\cdot w_1, v(d_2)\cdot w_2, \cdots, v(d_n)\cdot w_n\} \tag{1}$$

where p_j is set of vector properties in SWS, d_i is individual property representing concepts, $v(d_i)$ is value of the property, and w_i is weight of the property. The cosine similarity $h_k^{V,C}$ is calculated as a cosine of angle between the model vectors p_r and p_i.

$$h_k^{V,C} = \frac{p_r \cdot p_i}{\|p_r\|\|p_i\|} = \frac{\sum_{i=1}^{n} p_{r,i} \times p_{i,i}}{\sqrt{\sum_{i=1}^{n}(p_{r,i})^2} \times \sqrt{\sum_{i=1}^{n}(p_{i,i})^2}} \tag{2}$$

where n is a number of properties used for similarity calculation. To compensate in the case of the missing property value, Euclidean similarity $h_k^{V,E}$ is introduced as:

$$h_k^{V,E} = 1 - \frac{\sqrt{\sum_{i=1}^{n} w_i (p_{r,i} - p_{i,i})^2}}{max\sqrt{\sum_{i=1}^{n}(p_{r,i} - p_{i,i})^2}} \tag{3}$$

The vector similarity h_k^V is a mean average between combine cosine similarity $h_k^{V,C}$ and Euclidean similarity $h_k^{V,E}$:

$$h_k^V = \frac{h_k^{V,C} + h_k^{V,E}}{2} \tag{4}$$

The distance and property similarity, h_k^C and h_k^V respectively, are aggregate together as a fuzzy weighted average, where α and β are weighing parameters.

$$h_k = \frac{\alpha h_k^C + \beta h_k^V}{\alpha + \beta} \tag{5}$$

The above matching is comprised of three distinct stages: i) elimination, ii) semantic matching and ranking, iii) performance ranking. As a result, an optimum solution and multi-criteria analysis of model integration is established (F. Cecelja et. al., 2014).

3. Demonstration of Input-Output (I-O) Matching

A small scale model integration representing biochemical conversion technology is used to demonstrate the performance of the proposed approach. The model of co-fermentation (CF) uses the bacterium *Zymomonas Mobilis* to produce 315M litre of ethanol per year as an output from corn stover (A. Aden et. al., 2002). The operating condition of this model is temperature at 41°C with a resident time of 1.5 days and initial fermentation solids level of 20%, with an assumption that the hemicellulose sugars (arabinose, mannose, and galactose) have the same reaction and conversion as xylose, as a result, 56% of the fermentation broth is fermentable sugar.

Two models are selected as potential matches by the backward matching process. The set of respective requirements are listed in Table 1. Both of the models are pretreatment methods that break down the feedstock into fermentable sugars, dilute acid and enzymatic hydrolysis (DAEH) method and ammonia fiber expansion (AFEX) method. The DAEH process converts barley straw with composition of 38.1% glucan, 26.9% xylan, and 2.6% arabinan to monomer sugars, with yields of approximately 96% and 57% for glucose and xylose, respectively (M. Yang et. al., 2013). The AFEX process converts switchgrass with composition of 34.2% glucan, 22.1% xylan, and 3.1% arabinan to monomer sugars, with yields of approximately 93% and 70% for glucose and xylose, respectively (H. Alizadeh et. al., 2005).

Table 1. Process requirements

Process	Sugar monomer flowrate	% quantity matching	Temperature
CF (Requestor)	56 t/h	100%	41°C
DAEH	51 t/h	89%	50°C
AFEX	46 t/h	81%	37°C

In the second stage of matching, the semantic similarity by type of I-O is measured based on type of feedstock materials through class-subclass relationship. The vector similarity of the properties is then calculated related to quantity/flow rate of sugar monomers in the fermentable broth and operating temperature. As a default, the weight of the individual property and the fuzzy weight for the aggregated similarity, mentioned in Eq. 1 and Eq. 5, respectively, are equal to 1, unless user defines otherwise.

Table 2. Results from second stage matching

Resource Type	Semantic similarity	Property similarity	Aggregated results
DAEH	60%	63%	62%
AFEX	20%	50%	35%

The matching results are shown in Table 2, the DAEH and AFEX process with 62% and 35% matching similarities are found, respectively. As a result, the DAEH will be a preferable option for the requestor.

The challenges encountered while implementing the integration strategy for the models included identifying the key parameters that should be matched, and to obtain an understanding of the impact of partial matching associated with those parameters.

4. Conclusion

The wide range of models in the domain of biorefining can be described using SWS at any appropriate level of detail. A new approach was introduced to support model integration through I-O matching, which is implemented in OWL-S framework. The conversion technologies and resources in biorefining are semantically described and the relationships of SWS ontologies are established, which enables I-O matching. The backward matching process is proposed to identify a suitable model based on the input criteria. In addition, the semantic partial matching is performed to facilitate the flexibility of model integration. The proposed approach is implemented in a small scale model integration to demonstrate the flexibility, by showing model repository interaction between simulation interfaces and the data repository in the domain of biorefining.

References

A. Aden, M. Ruth, K. Ibsen, J. Jechura, K. Neeves, J. Sheehan, B. Wallace, 2002, Lignocellulosic biomass to ethanol process design and economics utilizing co-current dilute acid prehydrolysis and enzymatic hydrolysis for corn stover, NREL/TP-510-32438.

H. Alizadeh, F. Teymouri, T.I. Gilbert, B.E. Dale, 2005, Pretreatment of switchgrass by Ammonia Fiber Explosion (AFEX), Applied Biochemistry and Biotechnology, vol. 124 (1-3), pp. 1133-1141.

M. Alvarado-Morales, N. Al-Haque, K. Gernaey, J. Woodley, R. Gani, 2008, CAPE methods and tools for systematic analysis of new chemical product design and development, Computer Aided Chemical Engineering, on CD. vol. 25, pp. 997-1002.

F. Cecelja, T. Raafat, N. Trokanas, S. Innes, M. Smith, A. Yang, Y. Zorgios, A. Korkofygas, A. Kokossis, 2014, e-Symbiosis: technology-enabled support for industrial symbiosis targeting SMEs and innovation, Journal of Cleaner Production, In press.

R. Gani, I. E. Grossman, 2007, Process Systems Engineering and CAPE – What Next?, 17th ESCAPE Proceeding Book, Bucharest, Romania, vol. 24, pp. 1-5.

T. Gruber, 1995, Toward principles for the design of ontologies used for knowledge sharing, International Journal of Human-Computer Studies, vol. 43 (5-6), pp. 907-928.

R. Morales-Rodriguez, R. Gani, S. Dechelotte, A. Vacher, O. Baudouin, 2008, Use of CAPE-OPEN standards in the interoperability between modelling tools (MOT) and process simulators (Simulis Thermodynamics and ProSimPlus), Chemical Engineering Research and Design, vol. 86, pp. 823-833.

T. Raafat, F. Cecelja, N. Trokanas, B. Xrisha, 2013a, Semantic approach for pre-assessment of environmental indicator in industrial symbiosis, Computers & Chemical Engineering, vol. 59, pp. 33-46.

T. Raafat, N. Trokanas, F. Cecelja, X. Bimi, 2013b, An ontological approach towards enabling processing technologies participation in industrial symbiosis, Computers & Chemical Engineering, vol. 59, pp. 33-46.

N. Trokanas, F. Cecelja, T. Raafat, 2014, Semantic input/output matching for water processing in industrial symbiosis, Computer & Chemical Engineering, vol. 66, pp. 259-268.

M. Yang, S. Kuittinen, J. Zhang, M. Keinanen, A. Pappinen, 2013, Effect of dilute acid pretreatment on the conversion of barley straw with grains to fermentable sugars, Bioresource Technology, vol 146, pp. 444-450.