

The Need for an n -ary Existential Quantifier in Description Logics

Manfred Theißen
Lehrstuhl für Prozesstechnik
RWTH Aachen University, Germany
`theissen@lpt.rwth-aachen.de`

Lars von Wedel
AixCAPE e.V., Aachen, Germany
`vonwedel@aixcape.org`

Abstract

Motivated by the development of an information system based on Description Logics (DL), which provides efficient support for searching mathematical models in chemical process engineering, an n -ary existential quantifier for Description Logics is proposed. We show that the quantifier can be paraphrased using concept conjunctions and disjunctions and qualified number restrictions. Due to its high cost in current DL systems, the paraphrasis does not allow for practical applications. We state the need for an efficient implementation of the proposed n -ary existential quantifier.

1 Introduction

The use of mathematical models in chemical process engineering is becoming more and more popular because models permit the study of design alternatives by means of simulation on a cheap computer instead of actually having to build a real plant to experiment with. A major problem which impedes more widespread use of modelling and simulation is the cost and effort required to develop suitable models. However, chemical engineering is based on recurring pieces of process equipments to be reused in various combinations and thus there is a potential for reusing models already developed in different contexts.

As model development is performed in various business units using different simulation tools, a specific information system to support model reuse has been prototypically developed [1]. An important feature of such a system is its ability to organize all created models in a way which is meaningful to the users of the

system. Further, the system must provide an efficient search functionality which can be used to look up models that are candidates for being reused. In this contribution, two promising methods to find appropriate models are examined: the formulation of queries, in which the desired properties of the model are specified by the user, and the exploration of a class hierarchy, in which models are arranged according to certain criteria.

A rather simple realization of *queries* can be a keyword search in textual model documentations. More sophisticated solutions could make use of a database in which all relevant model information is stored. A model search via *exploration* is impeded by at least three critical issues: the definition of suitable model classes, their arrangement within a taxonomy, and the assignment of implemented models to classes. As a model library grows, the need to split classes into subclasses may arise, resulting in the necessity to reassign existing models to the new classes. This assignment should be done automatically. Things are complicated by the fact that multiple inheritance of classes must be provided as several discriminators are appropriate to define subclasses.

Both search method can make use of a DL-based knowledge base. In a conventional approach, one would distinguish between a TBox, providing the concepts and roles relevant to the domain, and an ABox, in which existing models are described by means of the vocabulary of the TBox. In Section 2, we will expand on this approach. An alternative, that overcomes the difficulties of the first strategy, is discussed in Section 3. The idea is to represent mathematical models as concepts, in a way such that subsumptions between the concepts can be interpreted as specialization relationships between the corresponding models. However, this approach requires an n -ary existential quantifier, whose semantics is proposed in Section 4. We conclude with stating the need for efficient DL systems providing the n -ary existential quantifier (Section 5).

2 Representation of mathematical models by individuals

A simple TBox that provides concepts to describe mathematical models on a coarse level is shown in a UML class diagram-like representation in Figure 1. Note that, for simplification, we do not distinguish explicitly between models and the modelled systems. Each model is assumed to describe either an **Apparatus** or a **Plant** comprising several **Apparatuses**. In each **Apparatus**, certain **Phenomena** can occur, e.g. chemical **Reactions**.

In the following example, we assume that the individuals **main-reaction**, an **Esterification**, and **side-reaction**, an **Alkylation**, contain information about specific reactions, i.e. the reactants and the products of the reaction. For instance, the reactants of the **main-reaction** are the **Substances** **butanol** and **acetic-acid**:

`has-reactant(main-reaction, butanol) , has-reactant(main-reaction, acetic-acid) .`

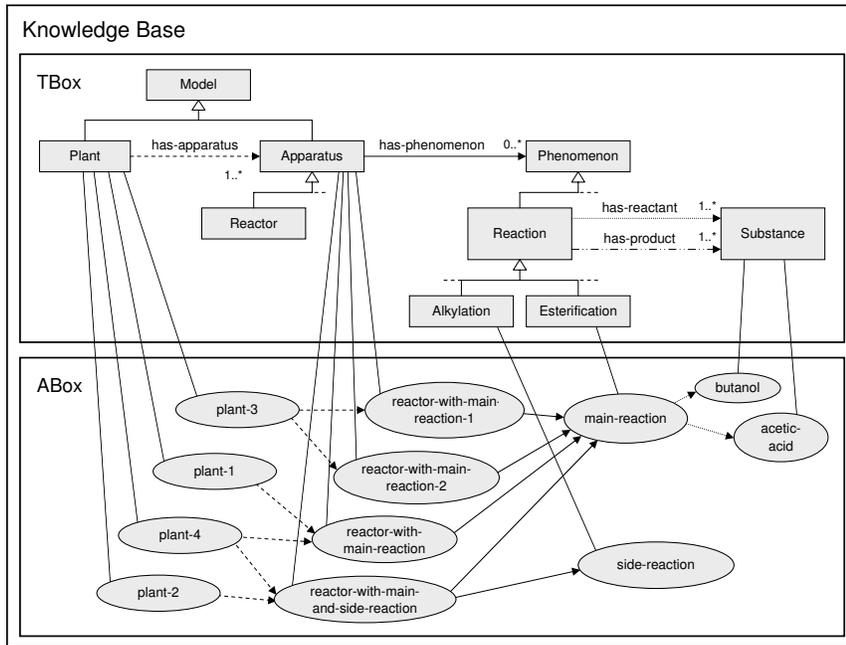


Figure 1: Representing mathematical models as individuals in an ABox

An **Apparatus**, in which only the **main-reaction** occurs, can be described as follows:

$$(\text{Apparatus} \sqcap \leq 1 \text{ has-phenomenon})(\text{reactor-with-main-reaction}) , \\ \text{has-phenomenon}(\text{reactor-with-main-reaction}, \text{main-reaction}) .$$

For a **Plant**, whose only **Apparatus** is a **reactor-with-main-reaction**, we can write

$$(\text{Plant} \sqcap \leq 1 \text{ has-apparatus})(\text{plant-1}) , \\ \text{has-apparatus}(\text{plant-1}, \text{reactor-with-main-reaction}) .$$

In a similar way, the ABox can be extended by assertions that describe the models **plant-2**, **plant-3** and **plant-4**, each of them characterized by a certain number of certain **Apparatuses** (Figure 1). By the help of the resulting knowledge base, we can do a search for models that comply with certain criteria. For instance, to determine the models that describe an **Apparatus** in which an **Alkylation** occurs, we can define the concept

$$\text{Query} \equiv \text{Apparatus} \sqcap \exists \text{ has-phenomenon}.\text{Alkylation} .$$

DL reasoners can identify **reactor-with-main-and-side-reaction** as an individual of the **Query** concept. Furthermore, the architecture of the knowledge base presented so far supports the exploration of a hierarchy. We can introduce special classes of **Models** and define corresponding concepts, e.g. a **Reactor**,

$$\text{Reactor} \equiv \text{Apparatus} \sqcap \exists \text{ has-phenomenon}.\text{Reaction} .$$

A DL reasoner can calculate the subsumption hierarchy of the primitive and

defined concepts, and for each concept, the associated individuals can be determined (Figure 1). Hence, a user searching for **Reactor** models can start the exploration of the model library at the general **Model** concept, and, by gradually refining his search, he will choose the **Apparatus** and then the **Reactor** concept. The individuals of the latter, e.g. **reactor-with-main-reaction**, represent models with the desired properties. In particular, this approach supports the automatic assignment of models to model classes as it is required above. There is no need to specify the affiliation of models to certain classes; this assignment is rather performed only on the basis of the concept definitions and the ABox assertions.

This approach suffers from several deficiencies. First, one might oppose the necessity to introduce rather identical individuals like **reactor-with-main-reaction-1** and **reactor-with-main-reaction-2**. Even if these two individuals represent two distinguishable objects in the real world, the only information we are interested in here is the occurrence of two **reactors-with-main-reaction** within **plant-3**.

Apart from this practical point of view, the representation of mathematical models by individuals involves some fundamental difficulties. Mathematical models in chemical engineering are abstract descriptions of real world systems in the sense that only those aspects are factored in which are regarded as essential for a certain task, e.g. the dimensioning of a chemical reactor. Frequently, the abstraction in the modelling process goes even behind that scope: We consider the mathematical model of an esterification reactor in which all necessary equations are given, but in which substance dependent parameters are not defined; these parameters are rather to be chosen depending on the specific esterification reactions which are to be examined by the help of the model. Such a general esterification model is no model building block in the sense of the definition by Marquardt [2] because not all its characteristic properties, that is to say the substances within the model, can be given. Nevertheless, today's modelling tools like gPROMS or Aspen Plus provide object oriented techniques that allow the implementation of such general models, their multiple instantiation, and the enrichment of the instances with the necessary information. Thus, we require a knowledge management system that can handle these models adequately. By means of ABox assertions, the general esterification reactor could be represented in the following way:

**Esterification(esterification) , Apparatus(esterification-reactor) ,
has-phenomenon(esterification-reactor, esterification) .**

Here, the individual **esterification-reactor** does not denote a specific reactor model, but the models of all possible esterification reactors. Consequently, the representation of the generic esterification reactor by a concept is more appropriate:

Esterification-Reactor \equiv Apparatus \sqcap \exists has-phenomenon.Esterification .

Obviously, the **reactor-with-main-reaction** is an individual of the **Esterification-Reactor** concept. This fact can also be interpreted in the sense that the **reactor-**

with-main-reaction is more special than the Esterification-Reactor: Whereas the generic model Esterification-Reactor is indeterminate with respect to a certain esterification, the model reactor-with-main-reaction describes a Reactor for the esterification of butanol and acetic-acid. A consequent development of this train of thought leads to the idea to represent all models in the knowledge base by the help of concepts. This approach enables the determination of specialization relationships between models via the calculation of subsumptions. In the following section, we will discuss the advantages and problems of this method.

3 Representation of mathematical models by concepts

For most of the individuals introduced in the previous section, their translation into concepts is quite straightforward. We assume that the concept Main-Reaction represents a specific Esterification and that Side-Reaction describes a certain Alkylation. Then, for the two reactors, we can define the concepts

$$\begin{aligned} \text{Reactor-With-Main-Reaction} &\equiv \text{Apparatus} \sqcap \exists \text{ has-phenomenon.Main-Reaction} , \\ \text{Reactor-With-Main-And-Side-Reaction} &\equiv \text{Apparatus} \\ &\sqcap \exists \text{ has-phenomenon.Main-Reaction} \sqcap \exists \text{ has-phenomenon.Side-Reaction} . \end{aligned}$$

If the unique name assumption is adopted, then reactor-with-main-and-side-reaction is an individual of the concept

$$\begin{aligned} \text{Reactor-With-At-Least-Two-Reactions} &\equiv \text{Apparatus} \\ &\sqcap \geq 2 \text{ has-phenomenon.Reaction} , \end{aligned}$$

whereas the concept Reactor-With-Main-And-Side-Reaction is not subsumed by Reactor-With-At-Least-Two-Reactions. From an application point of view, this subsumption is desirable. It can be achieved by formulating the disjointness of the generic concepts Esterification and Alkylation by the help of an appropriate axiom. Furthermore, we observe the subsumption

$$\text{Reactor-With-Main-And-Side-Reaction} \sqsubseteq \text{Reactor-With-Main-Reaction} ,$$

that expresses the idea that the first reactor is more special than the second because an additional reaction, the Side-Reaction is taken into account. In analogy with the individuals plant-1, plant-2, and plant-3, the concepts

$$\begin{aligned} \text{Plant-1} &\equiv \text{Plant} \sqcap \exists \text{ has-apparatus.Reactor-With-Main-Reaction} , \\ \text{Plant-2} &\equiv \text{Plant} \sqcap \exists \text{ has-apparatus.Reactor-With-Main-And-Side-Reaction} , \\ \text{Plant-3} &\equiv \text{Plant} \sqcap \geq 2 \text{ has-apparatus.Reactor-With-Main-Reaction} \end{aligned}$$

represent specific Plants, each of them comprising of one or two of the Reactors.

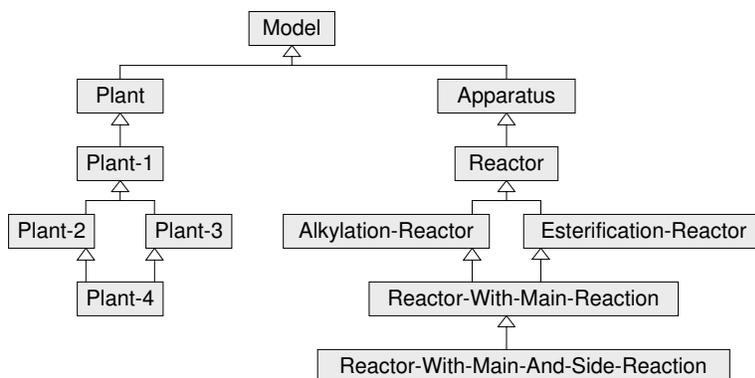


Figure 2: Specialization hierarchy of mathematical models

In the previous section, the concept **Esterification-Reactor** has been introduced as a representation of a general esterification reactor. In the corresponding definition, the existence of an individual of **Esterification** as a role filler for **has-phenomenon** is required, but no further restrictions, as for example ≤ 1 **has-phenomenon**, are involved in the definition. The conformance of the semantics of the concept **Esterification-Reactor** with the idea of a general esterification reactor as an apparatus with an esterification reaction and possibly other phenomena is apparent.

The concept **Plant-1** is a representation of the same plant which is described by the individual **plant-1**, i.e. it consists of an reactor, in which the main reaction occurs, but there are no further apparatuses in the plant. Nevertheless, in the concept definition above, an existential restriction is used rather than an *exactly* restriction. When specific models are described, our approach results in a certain inconsistency between the semantics of the introduced concepts and the objects that are represented by the concepts. But, by accepting this inconsistency, we gain the possibility to determine specialization relationships between mathematical models via the determination of subsumptions between concepts. The concept definitions lead to the hierarchy in Figure 2 that expresses intuitively comprehensible specialization relationships.

A naive attempt to describe **plant-4** by a concept could look as follows,

$$\text{Plant-4-a} \equiv \text{Plant} \sqcap \geq 1 \text{ has-apparatus.Reactor-With-Main-Reaction} \\ \sqcap \geq 1 \text{ has-apparatus.Reactor-With-Main-And-Side-Reaction} ,$$

but as $\text{Reactor-With-Main-And-Side-Reaction} \sqsubseteq \text{Reactor-With-Main-Reaction}$, the first existential restriction subsumes the second one. Therefore, we get the equivalence

$$\text{Plant-4-a} \equiv \text{Plant} \sqcap \geq 1 \text{ has-apparatus.Reactor-With-Main-And-Side-Reaction} .$$

Differently from the approach in the definition of **Reactor-With-Main-And-Side-**

Reaction, here the formulation of an axiom postulating the disjunction of the two filling concepts of the role `has-apparatus` is not feasible: It would result in the unsatisfiability of `Reactor-With-Main-And-Side-Reaction`. Nevertheless, the definition of a concept

$$\begin{aligned} \text{Plant-4} \equiv \text{Plant } \sqcap \geq 2 \text{ has-apparatus.Reactor-With-Main-Reaction} \\ \sqcap \geq 1 \text{ has-apparatus.Reactor-With-Main-And-Side-Reaction} \end{aligned}$$

guarantees an appropriate classification of `Plant-4` with respect to `Plant-2` and `Plant-3` (Figure 2): The model represented by `Plant-4` is more special than the model `Plant-2` because of the existence of an additional reactor, and it is more special than the model described by `Plant-3`, because the two plants have one reactor type in common and the second reactor of `Plant-4` is more special than the remaining reactor in `Plant-3`.

In natural language, the concept definition above can be read: "Plant-4 has two `Reactors-With-Main-Reaction` and one `Reactor-With-Main-And-Side-Reaction`, the latter possibly coinciding with one of the former two." But, as this description of `Plant-4` requires further knowledge about the filling concepts of the role `has-apparatus`, that is to say the existence of a subsumption relationship between them, the proposed solution is not satisfying. More desirable is a direct formalization of the description "Plant-4 has a `Reactor-With-Main-Reaction`, and, in addition, a `Reactor-With-Main-And-Side-Reaction`." We propose the notation

$$\begin{aligned} \text{Plant-4} \equiv \text{Plant } \sqcap \exists \text{ has-apparatus.} (\text{Reactor-With-Main-Reaction,} \\ \text{Reactor-With-Main-And-Side-Reaction}) . \end{aligned}$$

Clearly, the representation of mathematical models by the help of concepts overcomes the difficulties that are involved by an approach making use of individuals. The subsumption hierarchy in Figure 2 exhibits another advantage of this method: Whereas for the first approach only a limited number of concepts, which are to be given explicitly by the user, is the basis for the classification of models, here, all models in the knowledge base are potential candidates as superclasses for more detailed models. The hierarchical organization of the knowledge base is induced by its own content. We hypothesize that such model hierarchies, ranging from generic concepts to concrete implemented models, are a worthwhile help for a user in search of a certain model.

4 An n -ary existential quantifier

A practical realization of the envisaged knowledge base could not be achieved due to the difficulties we hint at above: Current DL implementations do not provide a concept constructor that describes role relationships between the individuals of the constructed concept on the one hand and the distinct individuals of several, possibly different concepts on the other hand.

Definition

We introduce an n -ary existential quantifier $\exists r.(C_1, \dots, C_n)$, where r is a role and the C_i , $i \in \{1, \dots, n\}$, are concepts, and define its semantics as follows:

$$\begin{aligned} (\exists r.(C_1, \dots, C_n))^{\mathcal{I}} = \{ a \in \Delta^{\mathcal{I}} \mid \exists b_1, \dots, b_n \in \Delta^{\mathcal{I}} : \\ (\forall i \in \{1, \dots, n\} : b_i \in C_i^{\mathcal{I}} \wedge (a, b_i) \in r^{\mathcal{I}}) \wedge \\ (\forall i, j \in \{1, \dots, n\}, i \neq j : b_i \neq b_j) \} . \end{aligned}$$

Obviously, if $C \equiv C_i \forall i \in \{1, \dots, n\}$, one has $\exists r.(C_1, \dots, C_n) \equiv \geq n r.C$, and in particular $\exists r.(C) \equiv \exists r.C$. The examples discussed in the previous section suggest that even in the general case, the n -ary existential quantifier can be paraphrased in a DL that provides conjunction and disjunction of concepts as well as qualified number restrictions.

Theorem

Let n be a positive integer, r a role, and C_i for $i \in \{1, \dots, n\}$ concepts. Then

$$\exists r.(C_1, \dots, C_n) \equiv \prod_{\substack{\mathcal{J} \subseteq \{1, \dots, n\} \\ \mathcal{J} \neq \emptyset}} \left[\geq \|\mathcal{J}\| r. \left(\bigsqcup_{j \in \mathcal{J}} C_j \right) \right] .$$

Proof

For an interpretation \mathcal{I} on a domain $\Delta^{\mathcal{I}}$ and $a \in \Delta^{\mathcal{I}}$, we define $R_i^{\mathcal{I}}(a) = \{ b \in \Delta^{\mathcal{I}} \mid (a, b) \in r^{\mathcal{I}} \}$. Then, to prove the theorem, we have to show that for all interpretations \mathcal{I} and for all $a \in \Delta^{\mathcal{I}}$

$$\begin{aligned} \exists b_1, \dots, b_n \in \Delta^{\mathcal{I}} : (\forall i \in \{1, \dots, n\} : b_i \in R_i^{\mathcal{I}}(a)) \wedge \\ (\forall i, j \in \{1, \dots, n\}, i \neq j : b_i \neq b_j) , \end{aligned}$$

if and only if

$$\forall \mathcal{J} \subseteq \{1, \dots, n\} : \left\| \bigcup_{j \in \mathcal{J}} R_j^{\mathcal{I}}(a) \right\| \geq \|\mathcal{J}\| . \quad (*)$$

For a given interpretation \mathcal{I} and $a \in \Delta^{\mathcal{I}}$, this is the statement of Hall's Theorem [3, 4], which, in its formulation by Cameron [5], says that a family $R_i^{\mathcal{I}}(a)$ of subsets of a set $\Delta^{\mathcal{I}}$ has a system of distinct representatives, if and only if it satisfies Hall's condition (*).

Remarks

The theorem shows that the proposed existential quantifier does not extend the expressiveness of a DL with concept conjunctions and disjunctions and qualified

number restrictions. For instance, to describe **Plant-4**, comprising at least one **Reactor-With-Main-Reaction** and, in addition, at least one **Reactor-With-Main-And-Side-Reaction**, we can write

$$\begin{aligned}
\text{Plant-4} &\equiv \text{Plant} \sqcap \exists \text{ has-apparatus.} (\text{Reactor-With-Main-Reaction,} \\
&\quad \text{Reactor-With-Main-And-Side-Reaction}) \\
&\equiv \text{Plant} \sqcap \geq 1 \text{ has-apparatus.Reactor-With-Main-Reaction} \\
&\quad \sqcap \geq 1 \text{ has-apparatus.Reactor-With-Main-And-Side-Reaction} \\
&\quad \sqcap \geq 2 \text{ has-apparatus.} (\text{Reactor-With-Main-Reaction} \sqcup \\
&\quad \text{Reactor-With-Main-And-Side-Reaction}) ,
\end{aligned}$$

and because of the subsumption $\text{Reactor-With-Main-And-Side-Reaction} \sqsubseteq \text{Reactor-With-Main-Reaction}$, we get the same expression for **Plant-4** as above:

$$\begin{aligned}
\text{Plant-4} &\equiv \text{Plant} \sqcap \geq 2 \text{ has-apparatus.Reactor-With-Main-Reaction} \\
&\quad \sqcap \geq 1 \text{ has-apparatus.Reactor-With-Main-And-Side-Reaction}
\end{aligned}$$

Nevertheless, the periphrasis of the quantifier does not allow for practical applications: $\exists r.(C_1, \dots, C_n)$ is expressed as a conjunction of $2^n - 1$ qualified number restrictions, for most of which the filling concept is a disjunction of several concepts. Apart from its sheer length – an algorithm for the generation of the translation taking into account formal simplifications due to subsumptions, equivalencies, and disjointness of some of the C_i is given in [6] – the high cost of the constructors in the periphrasis impedes any efficient reasoning. For example, if the existential quantifiers in the TBox

$$A \equiv \exists(C_1, C_2, C_3, C_4) , \quad B \equiv \exists(D_1, D_2, D_3, D_4) , \quad C_i \sqsubseteq D_i \forall i \in \{1, 2, 3, 4\}$$

are paraphrased, it takes RACER approximately 30 minutes to classify the concepts¹. Thus, for an efficient implementation of a DL providing the proposed existential quantifier, an extension of current DL algorithms or even a new algorithmic approach is needed.

5 Conclusion

A formalization of mathematical process models in terms of concepts as defined in Description Logics has been successfully applied. Based on concept definitions that approximately describe a model to be developed, suitable candidates for reuse can be determined by means of a simple subsumption calculation.

¹This test was performed with RACER V1.7.21 on a Windows 2000 system with an AMD Athlon XP 1700+ processor and 1.5 GByte RAM.

However, in addition to common Description Logics, an n -ary existential quantifier is needed in order to successfully represent models and their properties as described in [2]. Although this quantifier has been paraphrased in [6], this approach is not promising due to the time- and space-consuming behavior of reasoning with paraphrased concept definitions.

Hence we propose the extension of Description Logics by this quantifier so that more efficient implementations of it can be directly realized in description logic systems. Note that, for the application presented in this paper, such implementations do not need to allow for conjunctions of n -ary existential quantifications referring to the same role or cyclic concept definitions.

Acknowledgements

This work has been supported in part by the DFG (Deutsche Forschungsgemeinschaft) in the Collaborative Research Center IMPROVE (SFB 476).

References

- [1] Lars von Wedel. *An environment for heterogeneous model management in chemical process engineering*. PhD thesis, Lehrstuhl für Prozesstechnik, RWTH Aachen University, 2004. Online available at <http://www.bth.rwth-aachen.de/job/disslist.pl> (accessed August 12th, 2004).
- [2] Wolfgang Marquardt. Modellbildung als Grundlage der Prozesssimulation. In: H. Schuler (ed.), *Prozesssimulation*, Verlag Chemie Weinheim, 1994.
- [3] Personal communication of the authors with Prof. Franz Baader, Chair for Automata Theory, Technical University Dresden.
- [4] Philip Hall. On representatives of subsets. *The Journal of the London Mathematical Society*, 10:26–30, 1935.
- [5] Peter J. Cameron. *Systems of distinct representatives*. School of Mathematical Sciences, University of London, 2003. Online available at <http://designtheory.org/library/encyc/topics/sdr.pdf> (accessed August 9th, 2004).
- [6] Manfred Theißen. *Semantic analysis of mathematical models in process systems engineering* (in German). Diploma Project, Lehrstuhl für Prozesstechnik, RWTH Aachen University, 2004. Online available at <http://www.lpt.rwth-aachen.de/Publication/Sada.php> (accessed August 12th, 2004).