

Paper I

Pajula, E., Seuranen, T., Koironen, T., Hurme, M. Synthesis of separation processes by using case-based reasoning, *Comp. Chem. Eng.* 25 (2001) 775-782.

Reprinted with permission from Elsevier.

© 2001 Elsevier Science

Synthesis of separation processes by using case-based reasoning

Elina Pajula *, Timo Seuranen, Tuomas Koironen, Markku Hurme

Laboratory of Chemical Engineering and Plant Design, Helsinki University of Technology, P.O. Box 6100, FIN-02015 Hut, Finland

Received 10 May 2000; accepted 5 January 2001

Abstract

The paper presents a new approach to chemical process synthesis, case-based reasoning (CBR). CBR is based on the reuse of proven process solutions, which are used for solving new problems. Examples are given to demonstrate the principles of conceptual reasoning of separation systems. The phases of reasoning in this application are: the search for specific separation methods for a single separation phase, search for creative new solutions by using analogies, using negative cases to exclude some solutions, adaptation by engineering formulas and comparison of cases. Two quality factors, maturity and performance, can be used to evaluate the quality of the cases. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Case-based reasoning; Separation process synthesis; Conceptual design; Object-oriented approach

1. Introduction

Typical task in process design is to determine the configuration of a separation sequence. Several approaches have been studied in separation process design, e.g. residue curve maps in liquid–liquid separation (Fien & Liu, 1994), process design based on molecular structure (Gani & Constantinou, 1996) or thermodynamics (Jaksland, Gani, & Lien, 1995). Several rule-based AI methods have been applied (Barnicki & Fair, 1990, 1992; Wahnschaft, Le Rudulier, & Westerberg, 1993; Douglas, 1995; Hurme, 1996; Siirola, 1996). Another popular approach is mixed-integer non-linear programming (Grossmann & Kravanja 1995). The problem in this approach is that the amount of alternatives leads easily to combinatorial explosion.

It can be seen from the studies that the synthesis problem is difficult to handle by rules. Also creating rules (generalisations) causes information losses. Most of these methods are suitable only for limited types of separation processes. Therefore, it is obvious that more specific information is required in process synthesis.

The objective of this paper is to introduce a new method for finding feasible separation processes and process structures by case-based reasoning. This means

finding most alike existing processes and applying the knowledge of their separation capacity and design for solving new problems in the early phases of process design. This method does not try to replace any rigorous simulations in the process design, but gives a few feasible ways to split the given feed into products. In this way, it limits the number of processes that need to be considered and gives a systematic way of utilising earlier designs in new problems.

2. CBR — general features

CBR is a method for reusing existing knowledge (Kolodner, 1993) and it consists of four parts, retrieval, adaptation, revising and retaining (Fig. 1.). The current problem is defined as a query by giving the essential parameters e.g. feed components and product purity requirements. Based on these, the similarity is calculated and a user-defined number of the most similar cases are retrieved. The user can select a case and launch adaptation routines, e.g. scale-up calculations.

The more similar cases can be found the fewer accurate simulations are needed because a large part of the design can utilise data already available in the existing cases. The use of a database also forces to a systematic documentation practice. Documentation in a standard form can be used in the database as ‘cases’ and no additional work is needed to maintain the database.

* Corresponding author.

E-mail address: elina.pajula@hut.fi (E. Pajula).

3. CBR applied in process synthesis

Hurme and Heikkilä (1999) have applied CBR to evaluate the value of safety index used as an objective function in process synthesis. To our knowledge, no complete CBR based process design methodology has been presented earlier.

3.1. Hierarchy

Dividing the process into different levels can create the hierarchy. Levels are for example concept level (e.g. solvent recovery process), process systems level (e.g. a distillation column with all its auxiliaries) and equipment level (such as selection of filter type). The use of design levels makes the use of analogies more feasible. Such hierarchy includes complex domains in which cases have different structures. Object-oriented approach is particularly suitable for representing cases of this kind. Cases are represented as collections of objects. Structures (attributes) of objects are defined in object classes. Classes are arranged in a class hierarchy in which sub-classes inherit their parent classes. Complex case structures are represented by relational attributes, which define relations between objects of arbitrary classes.

Process design is far too complex field to be presented by using a treelike structure alone. For instance linking equipment to special separation operation and describing the related equipment properties in a feasible way is not simple. Our solution to this problem is the use of relational attributes. Relational attributes represent a direct binary relation (e.g. a part of relation) between object that defines the relational attribute and the object to which it refers.

3.2. Similarity

The nearest case is selected by comparing different features between the user input and the existing cases. These features are the phases present, kind of substances to be separated (e.g. hydrocarbons), chemical properties (e.g. relative volatility), and product purity requirements. At this stage, the 'design quality factors' has to be taken into account to make sure that old mistakes won't be repeated. This factor consists of technical maturity and performance as discussed later. When the best case is selected, the new design is created based on adaptation rules and heuristic knowledge included in the adaptation part of this application.

3.3. Target

Another important point to be considered is how to describe the target of the separation process. The target could be e.g. removal of the solvent as waste, when the amount of water in the solvent is far less important than purity of water leaving the separation section. On the other hand, if we are recovering dilute valuable component to be recycled, also the purity of the recovered can be very important. Therefore, the primary target of separation has to be defined.

3.4. Creativity

An important application of the computerised institutional memory is the reuse of the existing experience in the synthesis of new processes. Process synthesis is usually done in an industrial practice without computer support by copying and applying existing designs, and when those do not exist by using analogies. For example, a hydrogenation of a new compound can be probably done in an analogous way, as the hydrogenation of a related compound. This kind of design process has already been computerised in synthesis of new equipment configurations (Koiranen, 1998; Pajula, Koiranen, Seuranen, & Hurme, 1999).

Another important point in process synthesis is creativity. The system should not only be capable of modifying old designs included in the database but also capable of creating new designs. One possible way of including creativity into synthesis is to use analogies. Analogies can be included by using 'generalisations' and by using other structural features such as proper hierarchy. The generalisations introduced may include general level categories in database such as type of separation and phases present in separation (e.g. gas/liquid), physical properties of components (relative volatility etc.). See the example in Section 5.

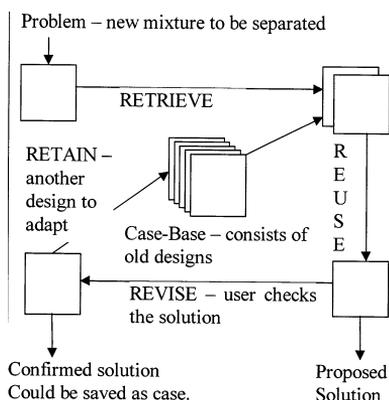


Fig. 1. CBR cycle (modified from Watson & Marie, 1994).

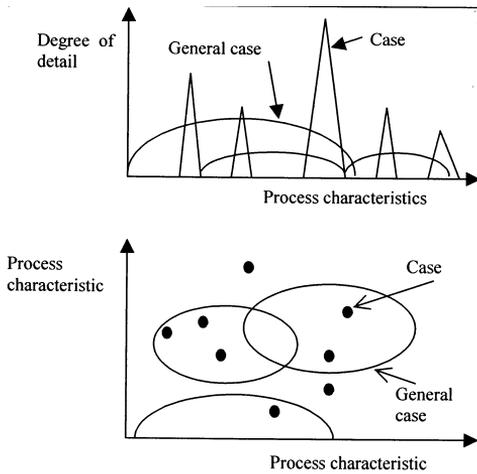


Fig. 2. Structure of the case-base (detailed and general cases).

3.5. Case-based vs. rule-based system

Rule-based systems have been the most commonly used approach to process synthesis. This kind of knowledge can also be included in a case-based reasoning system by defining ‘general cases’. For instance, if relative volatility is larger than 1.5 and the decomposition temperature is high for all components, the proposed separation method is distillation. The benefit of the general cases is that it gives always suggestion even if the specific application area is not well known before-

hand or no near cases have been stored in the database (see Fig. 2).

4. Structure of CBR Application

We have used a commercial CBR development tool, CBR-Works 4.0, for building this application. The hierarchical structure of the database is based on classification of separations presented by Wankat (1990). In the database structure, each process document contains general data for separation (feed composition, purification requirements, capacity etc.) and the separations that exists in the separation process train.

Every separation process consists of several pieces of process equipment, which have to be selected. Due to this, also a lot of equipment specific information needs to be stored as well. Link between the process description and equipment specifications has been created by using relational attributes. The tree structure and some relations of this application have been presented in Fig. 3.

The closer the separations are to one another in the tree structure more similar they are. The subclasses inherit attributes from their parent classes, and the amount of attributes increases the deeper the class is in the hierarchy.

The retrieval parameters can be selected and weighted by the user. The similarities of the attributes are calculated based on the type definitions. For in-

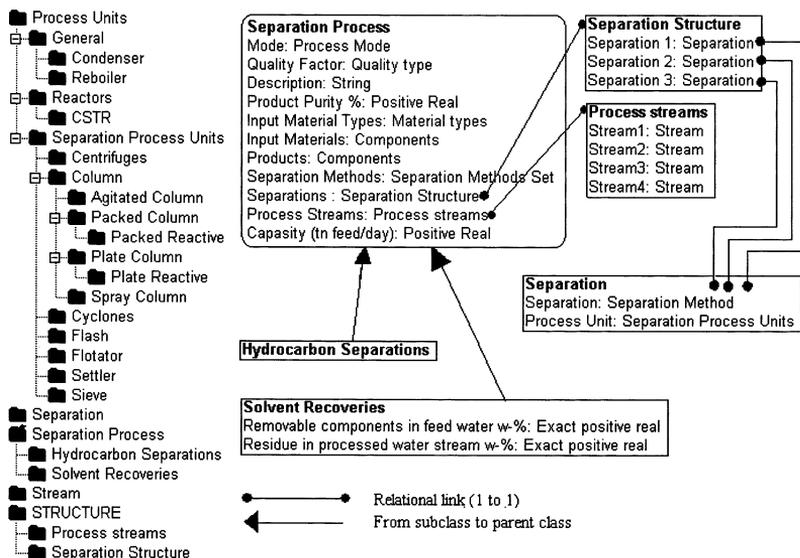


Fig. 3. A part of the structure and some relations of the CBR application.

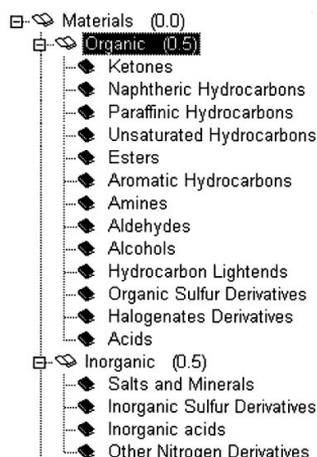


Fig. 4. A part of the taxonomy tree for material types.

Table 1
Technical maturity factor

Factor values	Description of the design
1	Process idea or concept exists
2	Process with basic engineering package exists
3	Plant in demonstration scale exists
4	Operating plant exists
5	Process is in wide use

Table 2
Technical performance factor

Factor values	Description of the design
0	Failure/unsafe
1	Out of date
2	Modest efficiency
3	Average efficiency
4	Proven good efficiency
5	Best available technology (BAT)

stance the similarity between percent values a and b is calculated as

$$\text{similarity} = 1 - \frac{|a - b|}{100}. \quad (1)$$

For taxonomy types, the similarity is based on the level where the nearest common node can be found. Part of material type taxonomy tree is presented in Fig. 4.

4.1. Quality factors

One of the most important points to be considered is definition of the quality factors. These factors describe the value and reliability of the design case. In our

application, two factors are used: technical maturity as presented in Table 1 and performance (goodness) factor in Table 2. Performance factor describes the proven efficiency of the design. The factors cannot be combined since an existing very proven process can become obsolete and therefore a low performer.

5. Example on binary separation

To present the principle of case-based process synthesis the separation of dimethyl formamide (DMF), water and some light and heavy impurities is discussed. This is first discussed as a subtask of selection of a binary separation and later in the last case extended to a synthesis of a complete separation sequence.

5.1. Query for specific cases

In a CBR system the user can make different types of queries in the database. The first query is made on DMF and water separation. Because of this query, specific cases on distillation and extraction of these components are found. See Fig. 5.

5.2. Creative querying of solutions

If new or more creative solutions are required, one can apply analogies or more general application cases in the CBR search. In this case for instance one can query for separations for similar component types: Query for nearly similar boiling point can be done as a b.p. interval (e.g. 150–160°C). Query for similar polarity can be done based on accentric factor etc. The different chemical types can be selected for query as seen in Fig. 4.

5.3. General cases

General application cases are general experience based application guidelines, which can be represented as cases in the case base. For instance air stripping can be applied when $\gamma^\infty P > 35\,000$ kPa in 1000 ppm concentration range. The general application guidelines give a more complete but shallower coverage of the search space than specific cases as seen in Fig. 2.

5.4. Negative cases

Negative cases can sometimes cancel the solutions proposed by general application guidelines: Based on general application guidelines it might be possible to apply pervaporation for the separation but a negative case found in the database lists DMF as a component which cannot be separated by pervaporation due to membrane problems. The current situation on this kind

of restrictions should be checked due to development of technologies.

5.5. Adaptation

Since the cases found in the query (Fig. 5.) have different operating conditions (e.g. feed concentration) an adaptation of the case has to be done. For instance, a shortcut way to adapt distillation for different conditions has been given by Douglas, Jafarey, and McAvoyn, (1979):

Separation factor S is defined as the ratio of the light (L) to heavy key (H) in the distillate (y) divided by the same ratio in the bottom (x) product:

$$S = \frac{y_L/y_H}{x_L/x_H} \quad (2)$$

Relation of separation S to average relative volatility α , number of real stages N , tray efficiency η , distillate rate D , reflux rate L and feed composition z of components lighter than the heavy key is given as

$$S = \left(\frac{\bar{\alpha}}{\sqrt{1 + D/Lz}} \right)^{\eta N} \quad (3)$$

5.6. Comparison

Comparison of the found cases can be done in principle by:

1. Costing (requires dimensioning of equipment).
2. Shortcut comparisons. For instance, the method of Porter and Momoh (1991) uses column vapour flow for comparison. This can be combined with the method of Souders (1964) that gives a generic method of comparison between normal and extractive distillation and extraction.

If the latter is applied to this case it seems that extraction is more feasible than distillation since for

distillation $\alpha = 3$, whereas for extraction $\alpha > 100$ in dilute DMF solutions. According to Souders (1964) $\alpha > 60$ is required for extraction to be more feasible than distillation in this case. Therefore, extraction is selected.

6. Examples on azeotropic separation

Separation of pyridine from water is studied as an example of azeotropic separation. The basic principle here is that the separation concept is determined by the concentration of product streams required, the composition of the azeotropic point and the solubility of the mixture (i.e. if there is a phase split). If these criteria are analogous to the case found in database, the process concepts are similar too and the concept found can be reused.

Case 1. Dilute pyridine water solution needs to be separated into products containing 40 and 1 wt% of pyridine. Search criteria:

1. Azeotropic composition of pyridine with water: 94°C and 57 wt%
2. Solubility with water: total
3. Feed composition: 15 wt% pyridine
4. Product compositions: 40 and 1 wt% pyridine

Using these parameters the nearest case found is:

Components	Tetrahydrofuran (THF) and water
Azeotrope	At 64°C 96 wt% THF
Solubility	Total (at 70°C)
Feed	20 wt% THF
Products	55 wt% and 3 wt% THF
Separation	Distillation in a single column without entrainer

Attributes	Query (Solvent Recoveries)	14 Liquid-liquid extraction	16 Distillation
Description	?	'DMF recovery'	'DMF recovery'
Distillate HK	0.0005	0.0005	0.0005
Feed HK	0.11	0.125	0.2
Input Materials	'DMF, Water'	'DMF, Water'	'DMF, Water'
Input Material Types	{Amines, Water}	{Amines, Water}	{Amines, Water}
Mode	Continuous	Continuous	Continuous
> Process Streams	?	?	?
Quality Factor	?	?	?
Separation Methods	?	{Extraction}	{Distillation}
v Separations	?	Separation Structure	Separation Structure
v Separation 1		Separation	Separation
v Process Unit		Plate Column	Plate Column
> Key Component - heavy		DMF	DMF
> Key Component - light		Water	Water
MSA		'chloroform'	?
N(Real)		?	50.0
Quality Factor		?	?
Number of Cases found (max. 10): 10		Similarity: 0.955	Similarity: 0.95

Fig. 5. Query and results on the case study.

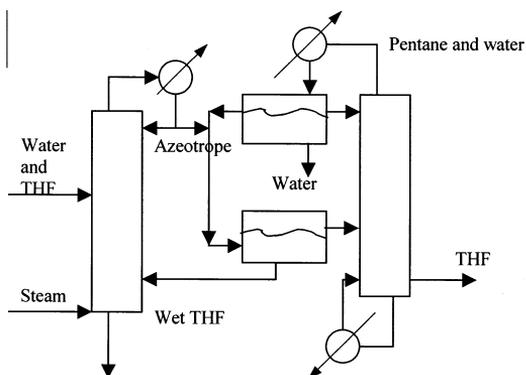


Fig. 6. Separation strategy found for azeotropic distillation with phase split.

The found case is analogous in the azeotropic behaviour and relative stream concentrations. Based on this it can be reasoned that distillation in a single column is applicable also in the pyridine case searched and no entrainer is required. Note; in both cases the azeotropic concentration is not crossed.

Case 2. To extend the reasoning into a more complicated case the following design problem is studied:

Dilute pyridine water solution needs to be separated to products containing 99 and 1 wt% of pyridine.

Search criteria:

1. Azeotropic composition with water: at 94°C and 57 wt%
 2. Solubility: total
 3. Feed composition: 15 wt% pyridine
 4. Product compositions: 99 and 1 wt% pyridine
- Using these parameters the nearest case found is:

Components	Tetrahydrofuran (THF) and water
Azeotrope	64°C and 96 wt% THF
Solubility	Total
Feed	20 wt% THF
Products	99 and 3 wt% THF
Separation	Distillation with entrainer (pentane) in a two column system (Fig. 6) with water decantation

The two column separation strategy found (Fig. 6) can be applied as the process concept of pyridine separation, if a suitable entrainer with similar behaviour (water solubility, binary azeotrope formation with water, etc.) can be found. The search of the entrainer can be made based on the criteria mentioned either on a computerised system by the CBR principle,

Table 3
Components in the case study

Component	Amount (%)	Boiling point (°C)
Heavy	3	165
DMF	9	153
Water	83	100
Light	5	25

or on a VLLE property data book manually. In fact cyclohexane and benzene are found to be analogous solvents for pyridine. Therefore, the separation concept of Fig. 6 can be directly applied to the pyridine separation (case 2) using these solvents.

7. Example on synthesis of a separation sequence

The earlier binary separation problem on separating DMF and water is now extended into a multicomponent process synthesis problem. The composition of the mixture to be separated is presented in Table 3.

7.1. Synthesis strategies for sequences

Synthesis of a separation sequence starts from the selection of a single separation as shown before. There are several alternatives to proceed to the selection of a sequence:

1. Finding all possible separation combinations. This is feasible only in small cases. Normally the number of combinations is large. For example for four components and ten separation methods there are 5000 different sequences. The combinatorial explosion takes place quickly when the number of products to be separated is increasing.
2. The use of an optimisation algorithm to find the most feasible separation sequence is a feasible approach as shown by Hurme (1996) by using a genetic optimisation algorithm. Another possible approach is MINLP.
3. Using an upper level CBR to find out the sequence. It is possible to store into the case base feasible existing cases of feasible separation sequences with the characteristics of the components separated etc. These cases can be retrieved based on the analogies and used for reasoning about the sequence of the current case. In this approach, there would be two levels of reasoning by CBR. Upper level, which reasons on the sequence, and the lower level, which concludes the separation method for, single separations in the sequence.

7.2. Combined operations

Another point, which should be covered, is the possibility of combined operations. Example a single column

can separate several products using side streams etc. The approach for this is first to consider single separations and then try to combine them one by one. Another way would be to conclude the possible combination operations from the retrieved cases.

If we use distillation for the water/DMF separation in the case study problem of Table 3, we could conclude for example the separation sequence in Fig. 7.

Searching for possible combination operations would give us e.g. a case for hydrocarbon separations with uncondensable gases where uncondensables are taken

out from the condenser as a third stream. Applying this case would give system shown in Fig. 8.

If the first separation method is not distillation but extraction, there is no combined operation for this but the result would be as in Fig. 9.

8. Conclusions

The paper has presented a new process synthesis method; case-based reasoning (CBR). CBR requires a database of existing design solutions, which is at least partly available in companies utilising databases for their engineering information management. The benefit of applying case-based approach is the systematic reuse and storing of the accumulated knowledge. The evaluation of the quality of stored cases prevents repeating the earlier design mistakes.

The phases of reasoning for a separation system are: the search for specific separation methods for a single separation phase, search for creative new solutions by using analogies, using negative cases to exclude some solutions, adaptation of the found cases to correspond to the problem to be solved, and the comparison of cases.

A separation sequence can be synthesised by using the single separations e.g. by optimisation methods or an upper level CBR.

The approach can also be applied in creative process synthesis. One possible way of including creativity into synthesis is to use analogies by using 'generalisations' as discussed.

The approach presented speeds up process design by defining in the early phase the process alternatives to be studied further by rigorous methods. This is more and more important as early design decisions are required in implementing new design paradigms such as process integration, inherent safety and clean technology. The major design decisions have to be made as early as possible in the process design.

References

- Barnicki, S. D., & Fair, J. R. (1990). Separation system synthesis: a knowledge-based approach. 1. liquid-mixture separations. *Industrial and Engineering Chemistry Research*, 29(3), 421–432.
- Barnicki, S. D., & Fair, J. R. (1992). Separation system synthesis: a knowledge-based approach. 2. gas/vapor mixtures. *Industrial and Engineering Chemistry Research*, 31(7), 1680–1694.
- Douglas, J. M. (1995). Synthesis of separation system flowsheets. *American Institute of Chemical Engineers Journal*, 41(12), 2522–2536.
- Douglas, J. M., Jafarey, A., & McAvoy, T. J. (1979). Short-cut techniques for distillation column design and control, part 1. Column design. *Industrial Engineering Chemistry Process Design and Development*, 18(2), 197–202.
- Fien, G. A. F., & Liu, Y. A. (1994). Heuristic synthesis and shortcut design of separation processes using residue curve maps: a review.

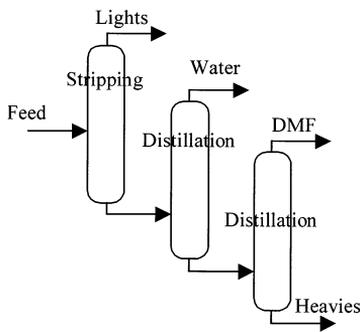


Fig. 7. Train of single distillation separations.

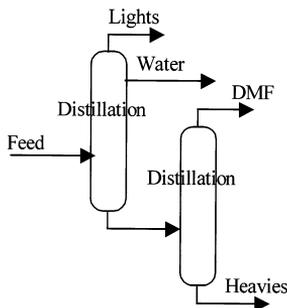


Fig. 8. The train with a combined operation.

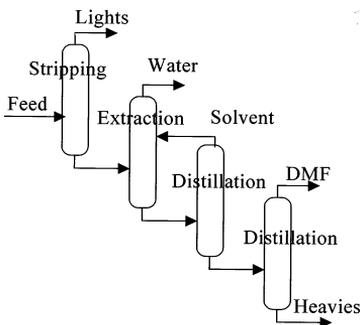


Fig. 9. The train with extraction separation.

- Industrial and Engineering Chemistry Research*, 33(11), 2505–2522.
- Gani, R., Constantinou, L. (1996). Molecular structure based estimation of properties for process design, *Fluid Phase Equilibria*, 116 1–2 pt 1 March 15.
- Grossmann, I. E., & Kravanja, Z. (1995). Mixed-integer nonlinear programming techniques for process systems engineering. *Computers and Chemical Engineering*, 19(Suppl.), 189–204.
- Hurme, M. (1996). Separation process synthesis with genetic algorithm. In: Alander, J.T. (Ed.). *Proceedings of the second nordic workshop on genetic algorithms and their applications*, pp. 219–224, Vaasa.
- Hurme, M., Heikkilä, A. -M. (1999). Conceptual design of inherently safer processes by genetic algorithms and case-based reasoning, pp. 341–346, *Proceedings of PRES'99*, Budapest.
- Jakslund, C. A., Gani, R., & Lien, K. M. (1995). Separation process design and synthesis based on thermodynamic insights. *Chemical Engineering Science*, 50(3), 530–551.
- Koiranen, T. (1998). Computer aided process equipment selection and design using knowledge engineering techniques, *Acta Polytechnica Scandinavica* No. 251, Espoo.
- Kolodner, J. (1993). *Case-Based Reasoning*. California: Morgan Kaufmann Publishers.
- Pajula, E., Koiranen, T., Seuranen, T., & Hurme, M. (1999). Computer aided process equipment design from equipment parts. *Computers and Chemical Engineering*, 23(Suppl.), 683–686.
- Porter, K. E., & Momoh, S. O. (1991). Finding the optimum sequence of distillation columns- an equation to replace the rules of thumb (heuristics). *Chemical Engineering Journal and the Biochemical Engineering Journal*, 46(3), 97–108.
- Siirola, J. J. (1996). Strategic process synthesis: advances in the hierarchical approach. *Computers and Chemical Engineering*, 20(Suppl.), 1637–1645 Part B.
- Souders, M. (1964). The counter current separation process. *Chemical Engineering Progress*, 60(2), 75–82.
- Wahnschaft, O. M., Le Rudulier, J.-P., & Westerberg, A. W. (1993). A problem decomposition approach for the synthesis of complex separation processes with recycles. *Industrial and Engineering Chemistry Research*, 32(6), 1121–1141.
- Wankat, P. C. (1990). *Rate-Controlled Separations*. Barking, England: Elsevier.
- Watson, I., & Marie, F. (1994). *Case-Based Reasoning. A Review*. *The Knowledge Engineering Review*, 9(4), 50–60.