

SELF-ADAPTIVE CUSTOMIZING WITH DATA MINING METHODS

A Concept for the Automatic Customizing of an ERP System with Data Mining Methods

Rene Schult and Gamal Kassem

Institute of Technical and Business Information Systems

Otto-von-Guericke-University Universityplace 2, Magdeburg, Germany

Keywords: ERP systems, reference model, data mining, workflow.

Abstract: The implementation of an ERP system is a long and cost intensive process. Functions of the ERP system, which are delivered in an enterprise neutral but sector specific fashion need to be adjusted to the specific business requirements of an enterprise. Exact knowledge of the ERP system is required because each ERP system has its own technical concepts and terminologies. Therefore many enterprises employ ERP system experts in order to customise the ERP system to be introduced as well as to further enhance the customisation after its introduction. A concept for the implementation of a Self-Adaptive ERP System should allow for the automatic customisation of an ERP system on the basis of the of enterprise process models provided and analysis of the ERP system usage.

1 INTRODUCTION

Customising is the main phase of an ERP implementation process. The behaviour of the ERP applications is controlled by the customisation. Changes within the conversion of the ERP application parameters are executed at this level (Peßl, 2004). First workflow models of the enterprise are modelled. The workflow, according to Gierhake, represents a work process, which is extensively supported by technology and is based on a triggered event, leading along a defined chain of part steps towards a defined work result whereby the completion level of the work result increases with each step. A workflow is the description of a specification that defines which elementary functions or tasks, that can be put into a standard, are to be executed by a person, who has been assigned a specific role at a given process state. If necessary, required data processing application for object processing are to be provided (Gierhake, 2000), pp. 57 ff.). At the implementation level, Rautenstrauch defines workflows as "...a part of a business process ..., that consists of sequentially or parallel arranged operation sequences (activities). Therefore it describes subprocesses of the sequence organisation within enterprises. The activities themselves are implemented as functions within applications systems. The cross-applications are controlled and therefore can be broken out of the application systems ..." (Rautenstrauch

and Schulze, 2003), p. 269). The workflow model encompasses a description of all relevant aspects of the workflow, which includes all incidental tasks, their sequences and dependencies. The model of workflows, specially business processes of an enterprise, is a precondition or basis for the customisation of an ERP system. During the implementation of an ERP system enterprise neutral sectoral functions are adjusted utilising the customisation on the basis of the modelled enterprise workflows to the specific business requirements of an enterprise. ERP system reference models are available as guideposts for experts in customising. Reference models generally describe objects and relations of the business reality. ERP reference models are pre-modelled on standard enterprise models and shapes the business architecture of an ERP system. Reference models often provide "best practice" business processes in the form of a software library (Peßl, 2004). Therefore, the customisation is used for the adjustment of the ERP system to the enterprise requirements. The adjustment itself is done by the specific ERP customising experts (consultants for the ERP system to be introduced). With the help of reference processes all business operations that are supported by an ERP system are mapped. The separate processes show tasks of a business situation that belong together logically. This can include a purchase order processing with the focus set on the lowest priced material, the vendor selection and order trac-

ing or a complete sales order processing from the customer inquiry to the delivery and the billing. Through the dependency of the processes on each other an understanding of cross-enterprise processes can be created with the user and on the other hand the interaction of operational function units can be depicted. The main task of customising experts is the implementation of the enterprise specific workflow models in the ERP System. Reference models may be used as a pattern for the system adjustment. The reference models are to be considered here as supporting information that helps customising experts to try and (reality true) map the enterprise specific workflow models in the ERP system. With this experts should accomplish (resolve) two main problems: the terminology problem that can arise through the different terms used and their meaning in an enterprise on the one hand and the terminology of the ERP system on the other hand. A further problem may arise at the selection of the adequate processes from the ERP reference models. The ERP processes should be selected and adjusted so that the enterprise processes are depicted in the system without developing new ERP functions or changing the structure of the enterprise itself. In this paper a self-adaptive customising concept is introduced that leads to the solution of the above mentioned problems through the usage of different data mining methods and which automates the customising procedure. Furthermore the concept should provide for the automatic execution of further customising procedures in the future. For that purpose the structure of a self-adaptive customising middleware system (SACMS) is presented that acts as a means of mediation between the workflow models of an enterprise and the ERP system. The SACMS should process different process modelling languages and ERP systems of different vendors.

2 THE SAC CONCEPT

The idea of the concept consists of the creation of an "intelligent" layer between the workflow models of an enterprise, which should be mapped into a particular ERP system, and the ERP system itself. This layer provides for the automatic customisation of the ERP system in which two customising types are to be distinguished: the initial customising will be done immediately at the implementation of an ERP system, while the "post-customising" is continually performed within the scope of a Continuous Improvement Process (CIP) and provides persistent improvement of the ERP customising.

2.1 Initial Customising

At the time of the implementation of an ERP system the enterprise neutral and sectoral specific functionality is adjusted to the specific business requirements of an enterprise during the initial customising. The initial customising covers three phases: conversion, evaluation and implementation (see figure 1).

2.1.1 Conversion Phase

In this phase enterprise workflow models as well as ERP reference models are converted in formal workflow model formats such as e.g. workflow nets (van der Aalst et al., 2003). Workflow nets stem from Petri nets (Baumgarten, 1990). The enterprise specific terms used in enterprise workflow models are translated in the terminology of the ERP system in the conversion phase. With this special text mining methods are used to recognise the meaning of a term and translate it into a terminology of the ERP system. Details to those text mining methods are shown at the end of this section. As the sequence of activities or functions of a workflow define its structure the main task of the text mining is to recognise which workflow activity is equivalent to an ERP transaction. Therefore not only the activity term is to be considered, the business objects involved in the transaction, the so called "workflow reference objects" are to be considered as well. The workflow reference object encompasses all business objects processed by a function of a workflow. The workflow reference objects can be divided into following:

Output Objects. The result of a function is a transformation of data for the processing of one or more workflow objects. This object or these objects are denoted as output objects.

Input Objects. For the processing of output objects by a function, data from other objects may be required. These objects are denoted as input objects. For example during creating a sales order (an output object), data about articles and the customer is required.

Workflow Activity Object. A workflow activity object is a workflow reference object that uniquely denotes a workflow. For example at the execution of the workflow "Production order processing" the functions "Create a production order", "Confirm a production order", "Posting goods receipt for production order" and "Release a production request" are performed. The workflow activity object in this workflow is the object "Production order", as it is part of all functions of

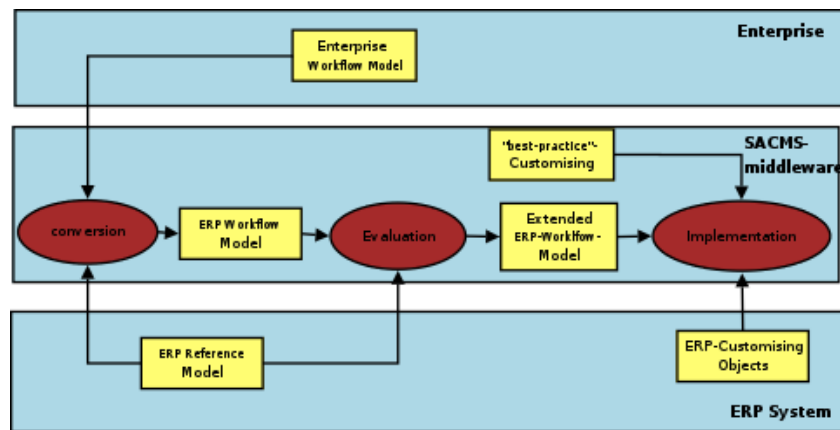


Figure 1: Self-Adapting Customising plan for the first customising.

the workflow: as output object in "Creating production order", "Confirm a production order" and "Release a production order" and as an input object in "Posting goods receipt for production order".

Workflow reference objects can be classified into the following object types:

Organisation Objects. are business objects that represent the organisation units of an enterprise, e.g. purchasing department, warehouse, company code, plant, etc.

Resource Objects. are all business objects other than the organisation objects. Such as customer, vendor, order, invoice, purchase order, article, etc.

A function of an enterprise workflow model can be converted via text mining methods if the function term (notation) and the involved workflow reference objects are equivalent to the ERP transaction term and its workflow reference objects.

Text mining use syntactical rules for the recognition of term structures and mining methods for the detection of the semantic meaning of terms. During the usage of text mining large data sets are used which describe the terms from the real enterprise world as well as from the ERP technology world. One example for such large dataset could be an ontology with the content of words of the real world and the mappings to the words of the ERP technology world. But it should not be limited to the vocabulary of only one ERP system. It is also possible that an ontology for both workflow models exists, the enterprise and the ERP specific workflow model, or one for each model. If we have both ontologies, it is possible to match both ontologies to one ontology for the formal workflow model (Euzenat and Shvaiko, 2007).

Another example for such large data set is the usage of synonym lists. In which each word of the

ERP technology world and the enterprise world are mapped to the equivalent of the formal model terminology. An algorithm search for each word the equivalent synonym from such a list to create an uniform vocabulary for the enterprise and the ERP workflow model.

2.1.2 Evaluation Phase

In the evaluation phase an converted ERP reference model is compared with the converted ERP workflow model and a suitable workflow or a workflow from the workflow combinations form. The reference model is constructed for the implementation phase as an "extended workflow model". The workflow selection is based on the comparison of model objects as ERP workflow structures and the functions and objects involved on the one side, and how they are mapped in the ERP reference model and whether they are linked to other model objects on the other side. The linking of objects in the ERP reference model can lead to the cascading of further objects that have to be considered in the implementation phase. An example is the selection of a workflow from the ERP reference model for "Processing Sales Order". Here it is considered whether the sales order is linked to further sub-workflows, for example for economic and/or technical checks, and whether these sub-workflows are linked to further sub-sub-workflows. Through the consideration of the linked objects all ERP transactions necessary for the execution of a workflow are provided and adjusted. Attention should be paid to the fact that not all transactions can actually be used in the future. In the post-customising phases the ERP system can be further improved and customised after the actual usage of the system.

2.1.3 Implementation Phase

At the implementation phase ERP customising object data are adjusted based on the selected workflows (the extended workflow models) from the ERP reference model. ERP customising objects are ERP repository objects which are necessary for the adjustment of an ERP system. For this purpose historical "best practice" information, stored for the customising of the ERP reference models, are used. This information allow that user roles, authorisation profiles, organisation units, currency unit and all other customising objects are automatically implemented in this phase based on the extended workflow models.

2.2 The Post-Customising

During the implementation of an ERP system the functionality provided within the initial customising is adjusted to the specific business requirements of an enterprise. With the provision of the linked model objects or customising objects, the system parameter settings are dimensioned in a large-scale fashion, so that all possible functionalities of the enterprise can be supported by the ERP system. Parts of the ERP functionality set are not used in the daily operational work which may lead to greater system complexity.

The underlying Self-Customising System takes place in the post-customising phases. The system is modified, respectively improved and adjusted to the needs of the enterprise on the basis of the system usage over time. In this process methods of Application Usage Mining (AUM) are used. AUM is a system technically realisable procedure for the reconstruction of actual workflow models of an ERP system. AUM should be an auxiliary means for the automatic modelling of a detailed actual workflow model of an enterprise.

ERP systems support enterprise-wide business processes and implement them as workflows. During the work with an ERP system, the ERP users leave traces in the form of ERP trace data that reflect the user interactions with the system. A supporting tool for this purpose is the methodology of the Application Usage Mining which automatically reconstruct actual workflow models from ERP traces that vulnerabilities in the system usage can be detected and removed through customising adjustments.

The post-customising is constantly executed in the frame of a Continuous Improvement Process (CIP) in order to provide a constant improvement of the ERP customising. The post-customising covers, similar to the initial customising, three phases: provision of the actually executed ERP workflow models (actual

workflow models), evaluation phase and implementation phase.

2.3 Provision of ERP Workflow Models

Instead of using the conversion phase of the initial customising in order to provide ERP workflow models, in the post-customising phase the actual workflow models are automatically extracted using the AUM methodology from the ERP system and are provided for the evaluation phase (see figure 2).

The existing workflow models represent the actual ERP workflow models used within the system. The AUM itself is a method that uses a KDD process based on ERP trace data. The AUM process covers three phases (Kassem, 2004):

In the **preparation phase** interaction data from different sources of the ERP system are integrated in a database (data pool). This are mainly trace data that protocols the user interactions with the ERP system. Additionally the data pool contains meta interaction data that describes the trace data.

In the **pattern detection phase** the workflow cases of a workflow are detected using special mining algorithms (Kassem and Rautenstrauch, 2006). A workflow case displays the sequence of the performance of ERP functions of a workflow instance. The algorithms address conflicts that may arise at the assignment of functions to the workflow cases. Based on these algorithms task steps (function cases) can also be detected as a screen sequence at the execution of a function. It displays which work steps of a task were performed.

In the **pattern analysis phase** the workflow cases detected in the pattern detection phase are analysed with the help of workflow algorithms (Agrawal, 1998; van der Aalst and van Dongen, 2002; van der Aalst et al., 2002) and are extracted formally as well as graphically using actual workflow models. The extracted actual workflow models can now be evaluated in the evaluation phase.

2.4 The Evaluation and Implementation Phases

At the evaluation phase only the actual used functionalities are considered in order to assure that the workflow structures are customised by the requirements of the enterprise and therefore the model is being improved consequently. Here analysis methods can be used for the evaluation of the generated actual workflows and the interaction data, collected in the data pool. The analysis should address the following items:

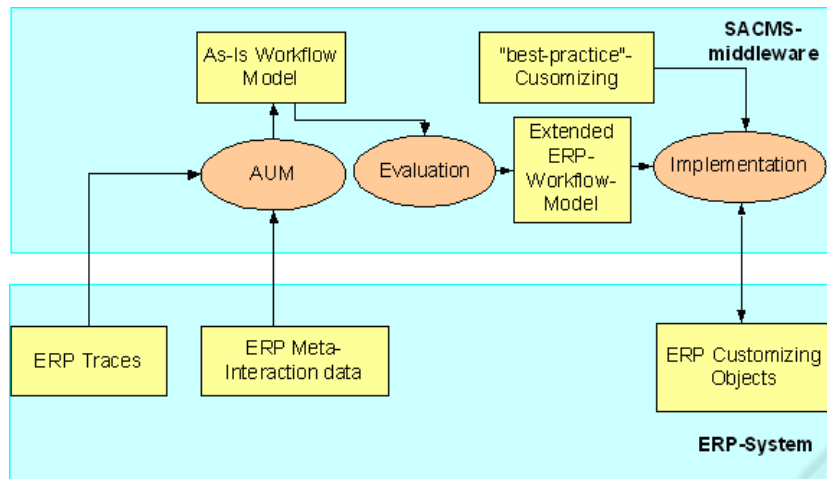


Figure 2: Self-Adapting Customising plan for the "post" customising.

The connection between the workflow process (its structure) and the workflow attributes (i.e. costs, runtime or attributes of workflow used business objects like customer name or material weight) can contribute to the detection of vulnerabilities at a workflow or, amongst others, illegal processes. Such kind of information can support the process management of an enterprise. Therefore mining methods can be developed to classify the workflow structures (like parallelism, branches, sequencibility or iterations/constellations of workflow structures) and for comparison with workflow attribute values for detection of connections as a type of patterns.

An the evaluation phase user behaviour can be analysed. The access paths of the users during the workflow processing are represented as graphs (Kassem et al., 2003). Different mining methods can be used to find answers to the following questions:

Path Analysis:

- At which step of equal user access paths an error message appears?
- At which step of user access paths transactions are often aborted?
- At which step the access paths reach dead ends?

To answer such questions from the workflow model different graph mining methods like classification of graphs (Borgelt and Kruse, 2002) can be used.

Association Analysis: The connection between events (e.g. input of datafields and the pressing of a button) and the screen sequence should be recognised through the employment of association rule methods. Here the customising of the system is checked according to its correctness. The deviation of the user access paths from the ideal access paths can be recognised. The business process should be analysed by

such methods. The well known apriori algorithm from Agrawal and Srikant in 1994 (Ester and Sander, 2000), for instance, can be used in order to create association rules from the screen sequences.

Temporal Analysis: Time referenced data is needed for the temporal analysis of the data. One example for the goal of such analysis is to examine the effectivity of the system usage of a user. Following questions can be asked (and hopefully answered) after the analysis process:

- How quickly new users are enabled to use the system?
- How much time is needed for a user to process a task?
- Which text fields always contained the same values and which values precisely were used?
- Which screens of a transaction were not used?
- From which transactions and from which screen other transactions were called?
- From which transaction screen help programs were necessary and which topic did they address?

The temporal mining methods like sequence analysis, detection of pattern evolutions or cluster monitoring gives us the possibility to find changes and movements at the workflow model during runtime and so to adjust the system over time to customise the system according to the new requirements of the enterprise (Aggarwal, 2007).

The **User behaviour Analysis** can be used for personalisation and for checking the correctness of the system customising and for checking the efficiency of the system usage at user tasks. As one result of such analysis it can be seen, how familiar a user is with the system and if a user or a user group needs some training or if the system should be customised or personalised for some users. to their needs through personalization. The extended ERP workflow models are provided for the implementation phase. The analysis of

the workflow structures and the user behaviour forms the basis for the models. These extended workflow models are after the system usage customised workflow models.

The ERP customising objects are modified by using the extended ERP workflow models at the implementation phase.

The post-customising process can be seen as a recurring process which customises the business process of an enterprise dynamically at any given time. The degree of customisation of an ERP system depends on the flexibility potential of the system. Here the flexibility of an ERP system means how customisable the ERP system is to the business process changes of an enterprise and how effectively the business process activities can be processed by the users or user groups depending on their roles and knowledge.

3 CONCLUSIONS AND OUTLOOK

The concept of a self-adaptive ERP system is an idea for an automatic implementation and a permanent customisation of an ERP system to the changes of business processes and user behaviours within the enterprise. To implement this idea, we have shown a structure of a system infrastructure, where the enterprise and the ERP-system layer are separated by an adaptive middleware. The adaptive middleware should be the interface between both layers. The function of the adaptive middleware SACMS is to ensure the adaption of the system to the requirements of the enterprise and users through intelligent methods like data mining methods.

A condition for the successful implementation of this concept is to have ERP-repository-data from the ERP system with customised objects like business-objects, transactions, screens screen-elements. Standardisation problems of workflow models, reference models and ERP-tracedata-formats inhibit general rules for handling ERP-systems of different providers. One possible solutions of this standardisation problem can be a layer specific interface between the SAC-layers which enables the data transfer between the different layers depending on the layer standards. Therefore it is necessary to solve the standardisation problems of workflow models and ERP-tracedata formats in the future.

REFERENCES

- Aggarwal, C. C. (2007). *Data Streams, Models and Algorithms*. Springer Verlag, Berlin, Germany.
- Agrawal, R.; Gunopulos, D. L. F. (1998). Mining process models from workflow logs. In *Advances in Database Technology - EDBT'98, 6th International Conference on Extending Database Technology*, pages 469–483, Valencia, Spain.
- Baumgarten, B. (1990). *Petri-Netze: Grundlagen und Anwendungen*. BI-Wiss-Berlag, Heidelberg, 2. edition.
- Borgelt, C. and Kruse, R. (2002). *Graphical Models, Methods for Data Analysis and Mining*. John Wiley and Sons, England.
- Ester, M. and Sander, J. (2000). *Knowledge Discovery in Databases*. Springer Verlag, Berlin, Germany.
- Euzenat, J. and Shvaiko, P. (2007). *Ontology Matching*. Springer Verlag Berlin, Berlin, Germany.
- Gierhake, O. (2000). *Integriertes Geschäftsprozessmanagement, effektive Organisationsgestaltung mit Workflow-, Workgroup- und Dokumentenmanagement-Systemen*. Vieweg, Braunschweig, 3. edition.
- Kassem, G. (2004). Application usage mining as an instrument of business process management. In *Proceedings of the IV International Conference of Applied Enterprise Science (CICE' 2004)*, Santa Clara, Cuba.
- Kassem, G., Marx Gomez, J., Rautenstrauch, C., and Melato, M. (2003). Analysis of the user's behaviour in very large business application systems with methods of the web usage mining - case study sap r/3. In *Advances in Web Intelligence (AWIC 2003)*, pages 329–338, Madrid. Springer Verlag Berlin.
- Kassem, G. and Rautenstrauch, C. (2006). The ability to generate as-is-model from an erp-system based on system user traces. In *Proceedings of the 7th Annual Global Information Technology Management Conference (GITMA)*, pages 133–136, Orlando, FL, USA.
- Peßl, E. (2004). Customizing von erp-systemen - ereignisgesteuerte prozessketten und das erp-referenzmodell. Technical report, University of Applied Science, Kapsfenberg.
- Rautenstrauch, C. and Schulze, T. (2003). *Informatik für Wirtschaftswissenschaftler und Wirtschaftsinformatiker*. Springer-Verlag.
- van der Aalst, W. M. P. and van Dongen, B. F. (2002). Discovering workflow performance models from timed logs. In Han, Y., Tai, S., and Wikarski, D., editors, *International Conference on Engineering and Deployment of Cooperative Information Systems (ED-CIS 2002)*, pages 45–63. Springer-Verlag Berlin.
- van der Aalst, W. M. P., van Dongen, B. F., Herbst, J., Maruster, L., Schimm, G., and Weijters, A. (2003). Workflow mining: A survey of issues and approaches. *Data and Knowledge Engineering*, 47(2):237–267.
- van der Aalst, W. M. P., Weijters, A. J. M. M., and Maruster, L. (2002). Workflow mining: Which processes can be rediscovered? Technical Report BETA Working Paper Series, WP 74, Eindhoven University of Technology.