Meet the challenge in service identification: A ratio-based approach

Jens Weller
*TU Dresden, jens.weller@tu-dresden.de*

Werner Esswein
*Lehrstuhl für Wirtschaftsinformatik, insb. Systementwicklung, Technische Universität Dresden, werner.esswein@tu-dresden.de*

Jeannette Stark
*Technische Universität Dresden, jeannette.stark@tu-dresden.de*

Martin Juhrisch
*Lehrstuhl für Wirtschaftsinformatik, insb. Systementwicklung, Technische Universität Dresden, martin.juhrisch@tu-dresden.de*

Follow this and additional works at: [http://aisel.aisnet.org/pacis2009](http://aisel.aisnet.org/pacis2009)

Recommended Citation
[http://aisel.aisnet.org/pacis2009/52](http://aisel.aisnet.org/pacis2009/52)

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2009 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
MEET THE CHALLENGE IN SERVICE IDENTIFICATION: A RATIO-BASED APPROACH

Werner Esswein, Jens Weller, Jeannette Stark, Martin Juhrisch
University of Dresden
Chair of Information Systems, esp.
Systems Engineering
01069 Dresden, Germany
{werner.esswein|jens.weller|jeannette.stark|martin.juhrisch}@tu-dresden.de

Abstract

Business processes in todays German universities possess high reorganization potential through Information Technology (IT). In particular service-oriented architectures (SOA) gain importance and respective development projects are started, distributed with respect to personnel, place and can comprise diverse organizations. We assume that a detailed analysis of company’s business process structure is a prerequisite to define useful services. In the SOA domain, the process landscape is most likely modelled in distributed modelling projects. This leads to variations since distributed models e. g. do not necessarily share the same level of detail. Thus, the model integration is a necessary step that has to be done before core functions can be identified within the documented processes for a planned implementation of services for a SOA. As the process models are documented electronically using a modeling tool, the process of identifying the service candidates can be supported by an automatic analysis of the process models. We introduce an approach to support the identification of services by using ratios generated out of the business process models and demonstrate its applicability within the research program.

Keywords: Information model, Service Identification, SOA.
1 INTRODUCTION

Many universities in Germany are faced with the challenge to improve and standardize its administration processes (Becker, Algermissen, Falk, Pfeiffer and Fuchs, 2006). In 2005 a consortium of four German universities started a research program to meet the challenge by analyzing its core processes in order to find reorganization potential of its structure. Thereby, information technology (IT) has been identified as a key driver for increasing the process performance.

Within the project MIRO (Münster Information System for Research and Organization) at the university of Münster, processes have been analyzed and documented using business process models to identify core functions within the documented processes for a planned implementation of services for a software architecture based on the SOA paradigm. However, due to the large amount of process models and its strong relationships the task of identifying these service candidates has been realized to be very difficult.

Within the research program the creation of a university wide service-oriented architecture (SOA; Fremantle, Weerawarna and Khalaf, 2002) is planned. In contrast to previous software architectures, SOA offers the possibility to establish a closer connection of business process flow and information system infrastructure (Erl, 2005). To realize such a university wide SOA the reengineering of legacy information systems is planned in order to provide Web services supporting the university processes. To identify necessary services, the university processes has been documented during the last years. Thereby activities, data structures and necessary resources have been modelled using the architecture of integrated information systems (ARIS; Scheer, 2000). Based on the process models, core functions shall be identified acting as candidates for a Web service implementation.

Due to the large amount of process models and its strong relationships, however, the task of identifying these service candidates has been realized to be very difficult. Thus, existing methods guiding such an identification (Ivanov and Stähler, 2005) could not have been used. As the process models were documented using a modelling tool, they are available electronically. Thus, the idea grew up to support the identification of service candidates by using a formal algorithm implemented in a software system. In this paper we introduce an approach to realize such a scenario. We present ratios that guide the identification of service candidates out of process models and show the feasibility of the approach by implementing in the context of the above introduced research program.

Our research is a matter of design science (Hevner, 2004). Following the research method presented in (Verschuren and Hartog, 2005), the paper is structured as follow. In the next section we start with a requirement analysis. Based on the identified requirements, ratios are for the automatic analysis of business process models are created afterwards. Finally, the implementation of the approach within the research program is demonstrated. The paper ends with a discussion, summarizing the research results and exposing open questions regarding the identification of service candidates.

2 REQUIREMENTS ANALYSIS

Management ratios are used as prospective instrument and come up to an important information and control function (Staudt, 1985). Reichmann and Lachnit emphasize the impact of ratios as information in decision making processes in managerial surroundings (Reichmann and Lachnit, 1976). The term ratio may be understood as condensed numerically measures (indicator, measurement) that gain their meaning independently of its structure as they singly bear upon the scientific object (Staudt, 1985).

A general ratio requirement is to map quantifiable data onto a concentrated form (Reichmann, 1995). Hence, we can derive two central characteristics. On the one hand it is presumed that every ratio implies informational value since its purpose is the condensation of data volumes to a single measurand. That measurand is the foundation to evaluate a certain situation subsequently. In addition, it is presumed that the data is quantitative measurable on a metric scale (Jäger-Goy, 2001).
The construction of ratios is bound to the following requirements:

- Fitness for a particular purpose. The informational value of a ratio should correspond with the information needs in the rough.
- Exactness. The exactness is conditioned by its reliability and validity.
- Actuality. The space of time between measurement and evaluation ought to be minimal.
- Cost-benefit ration. The effort to survey the ration should not cause costs above the value of the findings (Haufs, 1989).
- Simplicity and traceability. A result of a measurement must be simply interpretable.

The informational value of a ratio grows if it is connected into a comprehensive ratio system (Ester, 1997). This conclusion rests upon the assumption that single rather a few ratios are not able to exhaustively reflect the complexity of a system. Furthermore a multitude of single ratios impede the view to the essential issues (Wissenbach, 1967). Thus, with an ordered set of ratios we mean a ratio system. The ratios are correlated and provide as a whole information about a certain issue (Frank, 2001). Hence, the existence and catenation of at least two or more single ratios is condition precedent to create a ratio system.

A ratio can either be derived from superior ratios or developed concurrently and related with each other in a quantitative model. In the absence of such quantitative correlation, relations can be derived from empirical coherences as well (Ester, 1997). The creation of ratios and their coherency should always base on a comprehensive theoretically founding. The purpose should not be to respond to a specific question exempt from theory. For further information concerning the development of ratio systems we refer to (Ester, 1997). As a general rule, the measurement of ratios in business process models is related to single subprocesses or parts of the entire model. Seeing that, ratios are comparatively easy to implement in process models due to the fact of the straightforwardness to decompose processes into subprocesses. Considering a ratio system we have to assure to put measurement results of individual subprocesses into perspective. Thus, this is mandatory since we have to avoid the formation of suboptima (Engelke and Rausch, 2002). Besides, the examination of various subsystems associated with a multitude of ratios necessitates a concentration to a few significant ratios that combine collaboratively a maximized informational value.

The coverage of exclusive quantitative resp. quantifiable issues appears to be problematic (Ester 1997). In the case of additional required qualitative statements that are not reproduceable as a ratio, this information stay to a large extent outside of the examination. The adoption of ratios in business process models is a broadly discussed issue in literature. Though, on most cases existing approaches focus on the measurement of complexity in business process models. This means to survey in the first place the comprehensibility and (Gruhn and Laue, 2006; Cardoso, Mendling, Neumann and Reijers, 2006) and secondly the maintainability and correctness of models (Cardoso et al., 2006). We use ratios to identify eligible service candidates, a survey that bases not only on the measurement of complexity of process models. In fact criteria for the design of appropriate ratios have to be outlined in the following section.

Grounded on the concepts of contract, service, and interface, the SOA paradigm aims the service relation to a semi- or fully automated activity in business processes. This happens following the contract terms in which the characteristics of the activity’s implementation are defined (Dietzsch and Goetz, 2005). The service functions – meaning differentiated and autonomously working functions of a service also usable by other services – are utilised by the interface of an application.

Schwemm et al. deduce five design principles from literature: business orientation, self-containedness, modularity, interface orientation and interoperability (Schwemm, Heutschi, Vogel, Wende and Legner, 2006). Services are business oriented if their functional scope is geared to the required objects. Services are modular and self-contained if resources with high dependency to each other are combined in one service.

The design principles interface orientation and interoperability base upon the assumption that services represent stabil interfaces that are entirely specified using technical and business metadata (Schwemm wt al., 2006). As a complete and formal specification of processes in business models can not be
presumed, we constrain the deduction of ratios to the principles of business orientation, self-containedness and modularity.

2.1 Business orientation

This design principle refers to the granularity of a service function. The granularity equates to the scope of functionality that is provided with the service function (Griffel, 1998). A service is business oriented if it contains these business objects that are essential to perform a certain business activity (Schwemml et al., 2006). The objects could be modeled and interconnected as information objects using a conceptual data modeling language and be assigned to processes in business process models. If so, service candidates are a process rather than a bulk of processes that perform a common business task and access similar information objects. Hence, the information objects of the processes, which constitute a service, must show a high coherence. A measure of the coherence of a system is the cohesion (McCabe, 1997). A high grade of cohesion describes a high coherence of the elements of a service. The contrary implies a low coherence.

2.2 Self-containedness

To what extent a service can be evaluated as self-contained determines its maintainability. Following Simon, self-contained systems are better to maintain compared to dependent systems since modifications just imply marginal modifications at neighboring systems (Simon, 1962; Wand and Weber, 1990). Simon operationalizes self-containedness by dint of coupling (Simon, 1962). Coupling is a measure for the pairwise coherence between several subsystems (Wand and Weber, 1990). A single or an amount of processes could be identified as a service candidate if this process resp. this amount is independent of other processes. A process is independent of other processes if other processes do firstly not use its business objects and secondly the objects transferred to other processes are of little complexity (Yourdon, 1979). Accordingly this process could be automated as service without to hazard the other processes.

2.3 Modularity

By complying with the modularity principle during the design of a service, the complexity of the service could be reduced, parallel execution of services realized and uncertainty eliminated (Baldwin, Clark, 2000).

The fundamental idea of modular design can be traced back to the work of Parnas (Parnas, 1971). Self-contained functional units are merged and provided with a defined interface (Balzert, 1998). Balzert defines a module as representation of a functional unit or a semantically related functional group that is self-contained; possesses defined interfaces for external access and is in matters of its scope qualitatively and quantitatively manageable and understandable (Balzert, 1998). Analogical to business orientation, modularity is operationalized using the criteria of cohesion. Efforts are being made to quantify the distance dimension for miscellaneous decompositions and correlations.

3 DESIGN

Single processes as well as an amount of processes using the same information objects can be identified as a service candidate according to previously observed design principles autonomy, self-containedness and business orientation. Single processes can be identified as a service candidate if low coupling and high cohesion characterize these processes. An amount of processes can be identified as service candidates if they are characterized by high cohesion among the processes under consideration and low coupling towards outside processes. In this section coupling and cohesion will be operationalized to identify single processes and an amount of processes as service candidates. Furthermore requirements are derived for modeling languages to create process models that can be used to identify service candidates out of these models. To enhance understanding of the ratios
developed within this paper, we introduce an example, which will be used for demonstration for each of the ratios presented in this paper. We use the process view, event-driven process chain (EPC) and the data view of the Architecture of Integrated Information Systems (ARIS) to exemplify the ratios as this modeling language allows to combine both views as is demanded to identify an amount of processes as a service candidate (see Figure 1).

**Figure 1. Example for service candidate identification**

Single processes can be identified as a service candidate if its information objects are coupled as little as possible with information objects used by other processes. In case there are any information objects shared with other processes, these objects have to be as little complex as possible (Yourdon, 1979). To identify the number of information the process under examination \( P_i \) shares with other processes, we identify the intersection \( I_{i,t} \) of the information objects of \( P_i \) and the other processes. Process 1 \( (P_1) \) for example shares the information object A with \( P_2 \) and the information object B and C with \( P_4 \). The total intersection of \( P_i \) and all other Processes \( (I_{i,t}) \) is composed of A, B, and C. Its modulus \( (|I_{i,t}|) \) results three.

To identify the complexity of information objects shared with other processes, we derive the amount of relationships that is necessary to combine all information objects shared with another process \( j \) within the data view of the model. The total amount of all relationships of information objects within
the intersection of the process under examination with the other processes composes the total complexity of the process \( \sum C(I_{ij}) \). For example, the total complexity of \( P_1 \) is composed of the number of relationships necessary to combine the information objects within \( I_{12}, I_{13}, I_{14} \). \( I_{12} \) contains the information object \( A \), which is a single object and therefore can not be combined with other objects. \( I_{13} \) does not contain any information objects while \( I_{14} \) contains \( B \) and \( C \) as information objects. To combine \( B \) and \( C \) there is only one relationship required. Therefore the total complexity \( \sum C(I_{1t}) \) is one. Coupling can be derived with the following ratio:

\[
\text{Coupling} = |I_{it}| + \sum C(I_{ij})
\]

The less a process is coupled with other processes the more independent is the process and the less an automatisation of the process will influence other processes. Thus, processes with low coupling are possible service candidates. According to our example \( P_3 \) will constitute an adequate service candidate as can be seen in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>P1 (A,B,C)</th>
<th>P2(A)</th>
<th>P3(D,E)</th>
<th>P4(B,C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 (A,B,C)</td>
<td>-</td>
<td>I_{22}={A}</td>
<td>I_{31}=\emptyset</td>
<td>I_{41}={B,C}</td>
</tr>
<tr>
<td>P2 (A)</td>
<td>I_{22}={A}</td>
<td>C(I_{22})=0</td>
<td>I_{23}=\emptyset</td>
<td>C(I_{23})=0</td>
</tr>
<tr>
<td>P3 (D,E)</td>
<td>I_{23}=\emptyset</td>
<td>C(I_{23})=0</td>
<td>I_{24}=\emptyset</td>
<td>C(I_{24})=0</td>
</tr>
<tr>
<td>P4 (B,C)</td>
<td>I_{24}={B,C}</td>
<td>C(I_{24})=1</td>
<td>I_{33}=\emptyset</td>
<td>C(I_{33})=0</td>
</tr>
<tr>
<td>I_{t}</td>
<td>I_{t}={A,B,C}</td>
<td>I_{32}={A}</td>
<td>I_{43}=\emptyset</td>
<td>I_{46}={B,C,E}</td>
</tr>
<tr>
<td>(</td>
<td>I_{it}</td>
<td>) )</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>( \sum C(I_{ij}) )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Coupling</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Legend:
- \( I_{it} \) Intersection of shared information objects between Process \( i \) (\( P_i \)) and the other processes.
- \( |I_{it}| \) Modulus of intersection of shared information objects between Process \( i \) (\( P_i \)) and the other processes.
- \( \sum C(I_{ij}) \) Accumulated number of relationships of the intersections of information objects used by process \( i \) and process \( j \), whereas \( j \) symbolizes each of the other processes (Total complexity).

**Table 1. Example of coupling (single processes)**

A second ratio to identify single processes as service candidate can be derived from cohesion. We derive cohesion of a process from the complexity of its information objects. Thus, a process is characterized by high cohesion if its information objects can be combined with as little relationships as possible. The lower the number of relationships is required to combine the information objects the higher results cohesion of the information objects and thus, the more suitable the process serves as a service candidate. To norm processes we introduce \( N \) as the number of information objects a process uses and obtain the following ratio:

\[
\text{Cohesion}_i = N - C_i
\]

We apply this ratio to our example within Table 2. That way, \( P_1 \) uses the information objects \( A, B \) and \( C \). These information objects can be combined within the data view using two relationships. Therefore, the complexity of the information objects process \( 1 \) uses results two. After the norming process the cohesion of \( P_1 \) results one, as \( P_1 \) uses three information objects that have a complexity of two. The higher cohesion results the more suitable the process results as a service candidate. Within our example \( P_1 \) and \( P_2 \) form suitable service candidates.
Table 2. Example of cohesion (single processes)

An amount of processes forms a service candidate if its processes are characterized by high cohesion among the processes and low coupling to outside processes. As introduced in the requirements analysis section services should provide adequate granularity. While a high granularity of services reduces reuse low granularity constrains comprehension of services as context is lacking (Aier and Schönherr, 2004; Schwinn and Winter, 2005). Therefore, we limit our study to the granularity of two processes that form one total process. This limitation is chosen as the procedure to assess suitability as a service candidate for two processes can be easily adopted for more than two processes by successively adding one process after another. To identify two processes as a service candidate the number of information objects between two processes i and j are identified. The combination of processes that have the most information objects in common are selected. The corresponding ratio is:

\[
\text{Coupling} = \max |I_{ij}|
\]

As demonstrated in Table 3 P1 and P2 as well as P1 and P4 share information objects. P1 and P4 exhibit the biggest intersection of commonly used information objects.

Table 3. Example of coupling (an amount of processes)

Processes that can be combined to a single service candidate are furthermore characterized by low coupling to outside processes. For our example combined processes are P12 (a combination of P1 and P2) and P14 (a combination of P1 and P4). By combining information objects of P1 and P2 to P12 and P1 and P4 to P14 the ratio for identifying single processes as a service candidate can now be applied (Coupling = \(|I_{ij}| + \cdot C(I_i)\)). Table 4 demonstrates that P12 exhibits a coupling of three and P14 a coupling of 1. Therefore P14 is preferred to P12.
To identify the above mentioned ratios, information about relationships between information objects and relationships between processes and its information objects has to be included within the model. Modeling relationships between information objects allows to derive the complexity of information objects under examination. Modeling relationships between information objects and processes allows firstly to identify shared information objects by different processes, and secondly, allocating complexity of information objects to processes.

The event-driven process chain itself does not allow allocating information objects with its processes and to model relationships between information objects. However if we use the expanded event-driven process chain and the data view of the architecture of integrated information systems a connection between information objects can be established and information objects can be allocated to processes. Thus, the event-driven process chain in its ARIS context fulfils all requirements to use the ratios presented above.

4 IMPLEMENTATION

For the documentation of the processes within the university project, we used the modeling tool “Cubetto Toolset” (Cubetto, 2009). The software was given to us free of charge. It offers three major advantages that were relevant within the project: First, the tool allows the adjustment of the underlaying modeling language. Thus, it was possible to add concepts necessary for the creation of rations. Second, distributed modeling of processes is supported by an integrated configuration management (CM) system, including CM server. And thirdly, it is possible to increase the functionality of the tool by using so called plugins. Thus, the creation of ratios could be implemented easily.

How the process of automatically identifying single processes as service candidates is facilitated within the plugin we exemplarily show in figure 2 taking the example of selecting a single service candidate with cohesion. At first information objects are determined. These objects serve to calculate the number of information objects as well as the complexity of processes. The number of information objects as well as complexity are then used to draw the difference. This procedure is redone as many times as there are still processes. If there is no more process, one process will be selected as a service candidate that provides the biggest difference.
Within the project the architecture of integrated information systems (ARIS) has been used. Thereby, the business processes were documented directly by the several departments (university and state library, student affairs office, computing center) and integrated afterwards into a large integrated model (containing of lots of sub-models), which was supported by the CM system. Due to the integration it was possible to identify Web services that are valid for the whole university. For implementing the approach, a plugin for the Cubetto Toolset was created. It analysis the integrated model and provides functionality to generate the introduced ratios (see Figure 2).

As submodels of the University are already integrated to an overall model during the process analysis, we were able to detect also complex linkages with the help of the plugins. This allowed us to identify service candidates within the project MIRO, which may be used all over the university. Thus, the total number of services is kept down, without sacrificing functionality. With the help of automated analysis already 68 processes were identified as service candidates from the present business process models. Thereby, the necessary analysis time in comparison to a manual solution could be significantly shortened.

Figure 2. Exemplary process for generating ratios
Since ratios always only support decision processes, a subsequent assessment of the candidate service was necessary before they could step in the implementation phase. This showed that not all service candidates could actually be implemented as a service because they were either too coarse or too fine-grained. For example, the process "Send email to recipient" was considered being too fine-grained.

From a business point of view, this function is a result of a meaningless process decomposition. The business process is fragmented in nearly atomic functions. Inevitably “Mail” as an object was modeled, too. The only function of “Send email to recipient” is to send an email with password information to a given person. This provides a degree of granularity, which is to question having regard to the design principle of the need for business orientation (see Section 2.2.1). Accordingly, the granularity should be chosen in a way that it equate to business functions, and not mere technical basic functions. Hence, an implementation of the service was not performed.

We could, however, also identify the other extreme. Due to the automated analysis the process "examination processing and grading" was identified. Though, this is too coarse-grained, so that even here a direct implementation was abandoned.

Instead, a further analysis is necessary from the perspective of software development. In the present case, the service function could have serious impacts in the respective systems (e.g. removal from student registry by reasons of finally not passing the examination) and must be in a special way secured against abuse and manipulation.

5 CONCLUSION

In the previous sections we introduced an approach that support the identification of service candidates out of business process models. Using it within the university project has showed the feasability of the approach. Thereby candidates for Web services have been identified in spite of the
large amount of models. As our evaluation has shown, the generated ratios, however, sometimes pointed to processes that were too small for a meaningful implementation or that were too general and needs further process documentation. Reasons for this are the different levels of detail of the created process models. These problems, however, cannot really be avoided in a distributed modeling environment. Rules can be given when start the modeling process, but it is not possible to find formal rules concerning the level of detail.

To validate our ratios and further refinement, they must be tested through evaluation. Services that were found without methodological assistance are repeatedly compared to services that have been found with the help of our approach. From this test conclusions are drawn about the usefulness of the presented ratios. As we stated, our approach currently focus only three of the presented principles of service identification. Thus, our future research focuses the design and implementation of ratios regarding the principles not being considered for the current approach. Furthermore, we will transfer the general approach of generating ratios automatically out of business process models to other questions of process analysis. Thus, the efforts necessary within the process analysis stage of business process projects can be decreased step-by-step.

References
Engelke, M.; Rausch, A. (2002), “Supply Chain Management mit Hilfe von Key Performance Indikatoren”. In (Stölzle, W.; Gareis, K. Hrsg.): “Integrative Management- und Logistikkonzepte”. Festschrift für Prof. Dr. Dr. h.c. H.-Chr. Pfohl, Wiesbaden: 205-235


