A Hybrid Syntactic and Semantic Approach to Service Identification in Collaborative Networks

Ehsan Alirezaei, Saeed Parsa

A Hybrid Syntactic and Semantic Approach to Service Identification in Collaborative Networks

Ehsan Alirezaei[1], Saeed Parsa[2]

[1]PHD Student, Software Engineering Group, Iran University of Science and Technology, School of Computer Engineering, Tehran
Ehsan.alirezaii@gmail.com
[2]Assosiate Professor, Software Engineering Group, Iran University of Science and Technology, School of Computer Engineering, Tehran
parsa@iust.ac.ir

Abstract. In this paper a semi-automated approach for service identification, considering the requirements of partners in a collaborative network is presented. The requirements meeting the business goals are further applied as a means for building the business process to-be model. Approach begins with process and goal combination and maps action rules to tasks for specific resources for fill the gap between goal model and business model representations. Semantic analysis based on extraction of resource relations and their similarity is second part of main method to service identification. A correlation matrix from extracted weighted task's relations is used as an input of space vector machine. Semantic clustering of pre-processed vector is used for identify partner's services.

Keywords: Service Identification, Business Process Model, Goal Model, Collaborative network, Semantic clustering

1 Introduction

Collaborative network defined as "a network consisting of a variety of entities that are largely autonomous, geographically distributed, and heterogeneous in terms of operating environment, culture, social capital and goals" [1]. To achieving collaboration among partners, technical and non-technical approaches should consider. In technical point of view, Service identification is one of the approaches to achieve collaboration among partners. Several researches have suggested service-modeling approaches that can identify and specify service components [2-4]. There are a few automated and semi-automated techniques to identify services based on business tasks [5, 6 and 2] that outcomes of most of these techniques are business services that might be quite different from software services. Those techniques are not using combination of syntactic and semantic analysis to identification that this paper suggested to.

We used practical strategy based on existing business processes and goals for service identification. Identification of services by decomposing the business processes into tasks, providing reusable right-grained functionalities, proposed in some methods [7]. A major benefit of this approach is that the identified services satisfy functional needs and objectives [8]. Each task has an action rule to business process that should apply on resources with pre-conditions and post-conditions that we call them goals. Binding services to business processes with combining goals could fill the gap of top-down implementation with the bottom-up requirement engineering and implementation. Taking to the account the syntactic approaches of modeling structure it is not enough to fill the gap, so there is also a semantic approach to clustering correlated resources and joining two methods. Based on definition of goal that presented, tasks could be interrelated through the supporting goal for business processes. Therefore, functionality of a target system tracked and traced to business objectives and goals [9]. To identify proper services, combination of business process and goals should analyze [10].

In summary, we can say that to identify appropriate services, cross-organizational business requirements and business change factors should analyze to meet the collaboration objectives and agility [11]. The great benefit of goal driven approaches is that their resulting services have guaranteed fit with the organizations functional needs [12]. To facilitate accessing and manipulation of business resources one or more services could be defined as reusable modules of code. Therefore, business resources are useful to identify services and they classified based on their lifetime, handled by organization units of work [13] and resulting class considered as a candidate service. A resource that shared among business activities has a usage to identifying of activities as a service [14]. In general, a resource indicates a main stable domain abstraction of an enterprise that in certain situations, activities applied to relate a number of those resources commonly. Each service should identify through resource, resource-resource and resource-activity matrixes, those are built to measure the cohesiveness of the service [15]. However, services considered as a reusable set of non-interacting activities, which inter-related solely through shared access to one or more resources.

In the method proposed in this paper, three matrices, task-goal, task-resource and task-task applied to detect cohesive and reusable functionalities as candidate services. Services identified as cohesive, independent and reusable components. A service may be composed of set of tasks, which are interrelated through shared access to one or more entities or supporting a same goal. Goal, task and resource may lead to a cluster that shows same class that means an identified service. With considering the efficiency of the final code, some of the candidates selected as services.

The organization of paper is as follows: Service Identification Method described in section 2 in details, with a sample case study. The proposed method presented in some qualification criteria, tasks to services specified, and features has described. A summary to other methods presented in section 3. A summary of the results and future considerations of our research presented at section 4.

2 Service Identification Method

The method proposed in this article identifies services using combined business process and goal models and entities in two syntactic and semantic analysis way. The method consists of some steps that in first main step, the goals and the business process models extracted, and then combined for reaching Business Process (BP) to-be model. Next main step is going to extract the task-task matrix from the combined model. Next main step is using information retrieval based on vector space modeling techniques to building correlation matrix and building groups of tasks as services [16]. Comparing extracted tasks, entities with goals will make possible to identify services. These steps described in detail in the following subsections.

2.1 Creating Combined Business Process and Goal Model

In step one, it assumed BP As-Is models exist, and a job to do is reengineering it from current state to to-be model based on combination of goal models. For achieving this objective and bringing agility to BPs, combination of works [10] and [17] used. BPs modeled using BPMN2.0 standard. To create the current BP model of a cross-organization collaboration, that is not part of problem domain in this paper, using experts and process mining techniques could take to account for creating as-is model. Figure 1 shows part of BPMN model as "Purchase part" process with three role and some tasks for each role. It shows how purchase program commits by company experts trough configuring the program, email and comments. Middle-out approach based on existing BPs and main work of it, is bridging the gap between goals and services.

2.1.1 Goal Analysis

In first step, we represent goal model in order to satisfaction of current state relationship. A goal is an action rule, pre-condition and post-condition to an action in BP with roles that should under considerations [16, 18]. Pre-conditions are included "AND" provided statement and Post-conditions are included "OR" provided statement that it helps traceability analysis. From requirement engineering point of view, each request for services is an objective to satisfy that we could classify them in functional and non-functional requirements. These requirements have focus on current and future enterprise goals and define desired features to achieve in the future [19].

For representing goals in informal way that could be traced in hierarchical way and semantic manner, we used KAOS approach to modeling [18] modeling to fill the gap with BPs. In this method, goals hierarchy are refine by "AND" and "OR" operators that each refinement clear the satisfaction of higher goal in hierarchy. Simplicity of KAOS method is the main reason for using its method, and a formal goal definition begins with an assertion of the goal concretion objects. Each goal has informal and formal definitions. Event to complete the process make the traceability of the goal, and satisfaction links are between goal and process. The result of the work is important to us, so we do not enter the details of doing this. As it shows in Figure 2, the action rules and their hierarchy should satisfy to specific task accomplished like for achieving supporting program's workflow, two providing program and estimating sub-goal should satisfy that each of them represents providing program and sending approved estimation tasks in Figure 1.

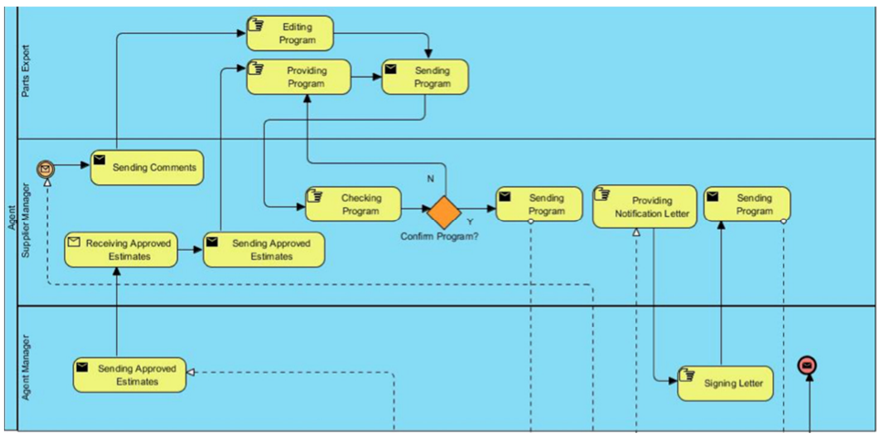

Fig. 1: Part of BP Model (As-Is) for Purchase Part Process [20]

2.1.2 Creating the BP To-Be Model Based on Goals and Tasks

In step 2, gap analysis is performed to extract needed changes in BP model. Tasks, which do not support any goal or sub-goal of organization, should be eliminated. In addition, to support reusability, tasks with a same set of goals can merged in a single task. To identify BP model tasks, which do not cover any system goal and objective, Task Goal matrix should consider with considering goal hierarchy that rows correspond to tasks and columns belongs to goals. Also, each goal analysis should consider clarifying need of task modification or introduction for task alignments. When BP designed in Visual Paradigm tools as it shown in figure 1, the tool could help to build task-task matrix and it is one it's outputs. For achieving task-goal correlation matrix, we need goal- goal matrix with weighting of related goals refinement relations it could be 1 value for "AND" and .5 value for "OR". Each correspond task in row with a goal in a column could be extract from the diagram of goal task as it presented in figure 2. For each related task and goal, the corresponding element in the matrix should have a value of 'y'. By using final goal-task matrix and goal-goal matrix task-task matrix could re-evaluate. For two tasks, all of related goals will be extracted from the final goal-task matrix and goal-goal matrix, with re-valued tasks weights.

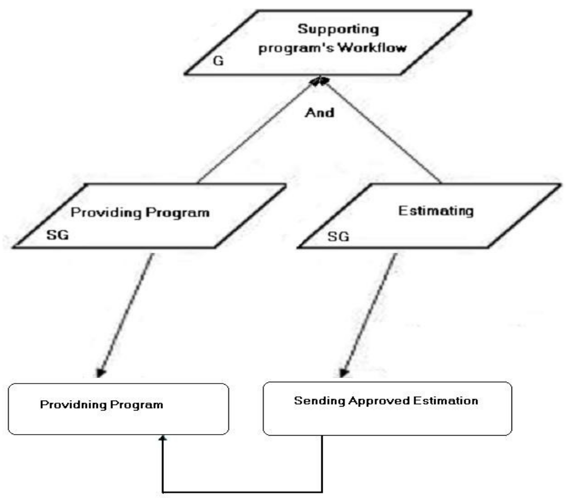

Fig. 2: Goal Model map to process task items for creating Supplying automation Systems

2.2 Creating Task-Resource Matrix

We assumed each goal as a rule action for tasks with resources. So resources are third criteria with importance weight that should apply on pervious built task-task matrix. To determine the importance of each resource for weighting, we assumed factors such as tasks operation on resource and access costs on resource. By consideration of these criteria, the importance of each resource is determined by value between 0 and 1. To perform each task, some resources will be accessed, so the relation between tasks and resources should be considered. In this step, by creating a matrix with rows that show the tasks and columns that show the resources the importance of the relevant entity for each task should be put in the corresponding element of the matrix. In our case study the “Approved Estimate” resource is related with these tasks: editing program, checking program, providing program, and receiving estimate [20].

Each resource may use some tasks and each task may access to some resources. This relation could have effect on task-task generated matrix by changing the values. The final values could be sum of the previous weights of task-task matrixes. In this step we are providing a correlation matrix with normalized weights that could help to identify services. Finding correlation and mutual effect or relation of task in same group and cluster that it helps to clarify services by them does Service identification by this matrix. Table 1 is a representation of correlation matrix that shows purchase program related to purchase process, which is valued by the combination of the first, second, and third matrix.

Table 1: correlated Task-Task Matrix for Purchase Part Process [20]

	Editing Program	*Checking Program*	*Notifying Program*	*Providing Program*	*Sending Comments*
Editing Program	*0*	*5.8*	*1.5*	*2.8*	*4*
Checking Program	*5.8*	*0*	*2*	*3.8*	*4*
Notifying Program	*1.5*	*2*	*0*	*1.5*	*1.5*
Providing Program	*2.8*	*3.8*	*1.5*	*0*	*2*
Sending Comments	*4*	*4*	*1.5*	*2*	*0*

To cluster tasks in vector space machine we used WordNet and aim is finding Pearson correlation coefficient. Similarity function between two tasks “a” and “b” with predefined weight is defined as:

$$\mathrm{Sim}(a_m,b_n)=\mathrm{W}(\mathrm{R}(a_m,b_n))$$

The optimum cluster occurs when cohesion between is minimum tasks as a service is the maximum possible value and coupling is minimum value between tasks in different clusters. For services reusability the optimum way is maximum value of similarity. For vector "V" there is frequency "f" of repeat patterns "p" and in each cluster "c" , for "W" the formula is as follows:

$$W\ (\mathrm{R}(a,b))=\sum_{\mathrm{ri}\in R(a,b)} f(r,a,b)/(\sum_{\alpha\in \mathrm{cj}} \sum_{(a,b)\in W} f(a,b,p)) * \sum_{(a,b)\in W} f(a,b,\mathrm{pi})$$

For clarifying the distinction between clusters and their relation, we can use the next formula for cluster "A" and task "t":

$$rel(t,A) = Sim(\mathrm{t},A) - \frac{1}{A} * \sum_{a\in A} Sim(t,a)$$

So there are two results, first for computing similarity between two tasks in vector space, and second for determining relation between a task and cluster. So after distinction between relation and similarities, there will be clusters with some elements that are our tasks, and districted cluster will be known as services. Identified services for the experiment are four clusters:

- ***Providing-program with similarity 0.75*** service with receiving estimation and preparing program methods
- ***Program-notification with similarity 0.62*** service with preparing program, email evaluation, email signing, and email evaluation methods
- ***Program-evaluation with similarity 0.58*** service with program correction, writing comments, sending comments, and comments evaluation methods
- ***Email- sending with similarity 0.7*** service with email signing, email notification

In the implementing the clustering subject and applying it to the current tasks, we found services with degree of similarity for each one that indicates that each cluster with similar concept and relation to its components.

3 Conclusion

For service identification in a collaborative environment, we presented a syntactic analysis based on hierarchical decomposition task to action rules, task-task matrix and considering the resources that goals act to. For part of semantic analysis we used clustering of identified relations in a task-task matrix by using the appropriate formula. Base of work is on documented existing Businesses processes between partners that help in considering structural relations between tasks and plays the main role in identifying services. In the future work our aim is to find cluster of services base on requests of users and their contexts in a dynamic collaborative environment processes are not documented.

References

1. Camarinha-Matos, Luis M., H. Afsarmanesh, and M. Ollus. "ECOLEAD and CNO base concepts." In *Methods and tools for collaborative networked organizations*, pp. 3-32. Springer US, 2008.
2. Rosen, Michael, Boris Lublinsky, Kevin T. Smith, and Marc J. Balcer. *Applied SOA: service-oriented architecture and design strategies*. Wiley, 2008.
3. Arsanjani, Ali, Shuvanker Ghosh, Abdul Allam, Tina Abdollah, Sella Ganapathy, and Kerrie Holley. "SOMA: A method for developing service-oriented solutions."*IBM systems Journal* 47, no. 3 (2008): 377-396.
4. Fensel, Dieter, and Christoph Bussler. "The web service modeling framework WSMF." *Electronic Commerce Research and Applications* 1, no. 2 (2002): 113-137.
5. Jamshidi, Pooyan, Mohsen Sharifi, and Saeed Mansour. "To establish enterprise service model from enterprise business model." In *Services Computing, 2008. SCC'08. IEEE International Conference on*, vol. 1, pp. 93-100. IEEE, 2008.
6. Jain, Hemant, Huimin Zhao, and Nageswara R. Chinta. "A spanning tree based approach to identifying web services." *International Journal of Web Services Research (IJWSR)* 1, no. 1 (2004): 1-20.
7. Mani, Senthil, Vibha S. Sinha, Noi Sukaviriya, and Thejaswini Ramachandra. "Using user interface design to enhance service identification." In *Web Services, 2008. ICWS'08. IEEE International Conference on*, pp. 78-87. IEEE, 2008.
8. Hubbers, Jan-willem, Art Ligthart, and Linda Terlouw. "Ten ways to identify services." *The SOA Magazine* 13 (2007): 1-7.
9. Levi, Keith, and Ali Arsanjani. "A goal-driven approach to enterprise component identification and specification." *Communications of the ACM* 45, no. 10 (2002): 45-52.
10. Krogstie, John. "Using eeml for combined goal and process oriented modeling: A case study." In Proceedings of EMMSAD, p. 113. 2008.

11. Iocola, P. "When Legacy Meets SOA: Achieving Business Agility by Integrating New Technology with Existing Software Asset." In *Systems Conference, 2007 1st Annual IEEE*, pp. 1-8. IEEE, 2007.
12. Fareghzadeh, Nafise. "Service identification approach to SOA development." In *Proceedings of World Academy of Science, Engineering and Technology*, vol. 35, pp. 258-266. 2008.
13. Rostampour, A., A. Kazemi, F. Shams, A. Zamiri, and P. Jamshidi. "A metric for measuring the degree of entity-centric service cohesion." In *Service-Oriented Computing and Applications (SOCA), 2010 IEEE International Conference on*, pp. 1-5. IEEE, 2010.
14. Daghaghzadeh, Mohammad, Ahmad Baraani Dastjerdi, and Hossein Daghaghzadeh. "A Metric for Measuring Degree of Service Cohesion in Service Oriented Designs." *International Journal of Computer Science* 8.
15. Yousef, Rana, and Mohammed Odeh. "A New Service Identification Approach Based on the Riva Business Process Architecture." *book chapter: chapter 8 of "FRONTIERS IN INFORMATION TECHNOLOGY, 2012.*
16. Kuhn, Adrian, Stéphane Ducasse, and Tudor Gírba. "Semantic clustering: Identifying topics in source code." Information and Software Technology 49, no. 3 (2007): 230-243.
17. Koliadis, George, and Aditya Ghose. "Relating business process models to goal-oriented requirements models in KAOS." In Advances in knowledge acquisition and management, pp. 25-39. Springer Berlin Heidelberg, 2006.
18. Van Lamsweerde, Axel. "Goal-oriented requirements engineering: A guided tour." In *Requirements Engineering, 2001. Proceedings. Fifth IEEE International Symposium on*, pp. 249-262. IEEE, 2001.
19. Kavakli, Evangelia, and Pericles Loucopoulos. "Goal driven requirements engineering: evaluation of current methods." In *Proceedings of the 8th CAiSE/IFIP8*, vol. 1, pp. 16-17. 2003.
20. Saeed Parsa, Mohammad Javad Amiri and Amir Mohammad Zadeh Lajevardi, Semi-Automated Entity, Goals Model, and Business Process Model Driven Service Identification (pp. 40-55) JCSE Vol. 11, No. 3, Fall 2013