Automated Service Identification Methods: A Review

Ali Taei Zadeh, Shahnorbanun Sahran#1, Muriati Mukhtar#2

#Department of Industrial Computing, UKM, Kuala Lumpur, Malaysia
E-mail: 1alitaee@gmail.com, 2mm.shah@ukm.edu.my

Abstract—Service identification represents the first phase in service modelling, a necessary step in SOA. This research study reviewed and analyzed the issues related to automation issues of service identification. However, the importance of service identification methods’ (SIM) automation and their business alignment are emphasized in literature, reviewing existing service identification methods (SIMs) reveals the lack of business alignment, automation as challenging issues. We close the gap by proposing ASIF which relies on automating the SIMs’ steps to identify business aligned services based on business processes and business goals.

Keywords—business-IT alignment; automated tool; integration and modelling; software architectures

I. INTRODUCTION

According to different views regarding service oriented architecture (SOA), the existing service identification methods (SIMs) differ significantly in terms of delivery strategy, business-oriented (top-down) or technical-oriented (bottom-up). Services in the business-oriented methods correspond to business domain entities including business processes (as-is) and business goals (to-be), while technical-oriented SIMs merely rely on applications’ assets as their inputs such as database, interface, or other technical entities. Nevertheless, most service orientation researches have focused on technical domain instead of business domain. Establishing a link between business entities such as business processes and web services will guarantee business-IT alignment.

Hence, business-oriented SIMs are more effective, but due to descriptive nature of business domain entities, automation of these methods is not fully developed and consequently they are underutilized [1]. Automation is a challenging subject in existing SIMs due to lack of an automated business-oriented method. However, among existing SIMs few of them aim to propose automated method. The reason behind the lack of automated techniques is the complexity and fuzzy nature of top-down SIMs. Besides, the majority of automation techniques in SIMs have been proposed within software-oriented (bottom-up) SIMs, on the other side business entities such as business goals and business processes have descriptive nature, which makes their processing costly and time-consuming. Therefore, they are needed to be quantified to be involved in automating SIMs. Therefore, to reach an automated service identification method, this research aims to address the following questions:

RQ1: What are the existing service identification techniques?
RQ2: Which types of automated techniques are used by SIMs? This can be elaborated into:
RQ2.a What types of automated techniques are used in Bottom-up or Top-down techniques?
RQ2.b Which steps of service identification are fully or semi-automated?
RQ3.c What are the existing SIMs’ tools?

To address the research questions, all necessary activities and techniques of service identification that reveal the service portfolio should be considered. The literature review identifies pre and post phases of service identification, namely, scope determination, input-type selection, and service refinement. According to the state-of-art literature review, there is a lack in consensus on the type and way of automation techniques.

The rest of this paper is outlined as follows. Section II presents the review results related work focusing on issues of service identification methods; then, the review results regarding the existing automated tools are discussed in Section III. The last section deliberates the conclusion and closing remarks of the study.

II. MATERIAL AND METHOD

Automating the manual activities has been the trend of software development methodologies. Automation is a challenging subject in existing SIMs due to lack of an
automated business-oriented method. [2] State that the reason behind neglecting applying an automated technique is the complexity and the subjectivity of making decision in such methods. The SIMs can be classified into three types in terms of automation levels: fully-automated, semi-automated and manual [3]. Besides, the majority of automation techniques in SIMs have been proposed within software-oriented (bottom-up) SIMs which essentially rely on legacy systems’ transformation [4]. This section firstly presents some bottom-up SIMs that stress on automation in their methods. Then, the focus will be on the top-down SIMs.

A. Bottom-Up Automated SIMs

[5] Proposed a SIM called MOGA-WSI which is considered as the first automated approach belonging to bottom-up methods. It proposes hierarchical grouping of classes in object models and then extracts the web services according to strength degree of relationships between objects via spanning tree algorithm. However, since it is dependent on experts’ decision in terms of determining the relationship strength between two classes, it is regarded to be a partially automated method. [6] Introduced a bottom-up service identification method to identify services from user interfaces automatically. Its main contribution is developing the XML-based representation of user interfaces called Unified UI Design Specification (UUIDS). The method firstly creates a specific format for each task in the user interfaces manually, then, the interaction points are represented in the UUIDS format to be generated, finally, are transformed to WSDL automatically. Although it is claimed to be an automated SIM, it is applicable only when the user interface has been designed based on their specific tool that is human-based tasks. In addition, it has not considered the service quality factors measurement within service identification. Zhang et al. [4] Present a bottom-up service identification method that applies Options Analysis for re-engineering technique for revealing the status and architecture of legacy systems. Their service identification method emphasizes on business functionalities of legacy systems to be reused in the identified services. The method includes service packing and service registration under Universal Description Discovery and Integration (UDDI). In addition, it contains domain analysis to discover new requirements in comparison with identified components in the architecture of the legacy systems which can be realized as services via new service list. Regarding service quality factors, it prioritizes the reusability of the legacy systems’ components as well as a loose coupling by decreasing the inter-relationships. However, the proposed method provides automated method based on Simple Architecture Description Language (SADL) that uses clustering technique via an automated tool to analyze the recovered architecture information for transforming the legacy systems components to service domain. The complementary steps regarding business domain analysis was lack of guidelines and automated techniques. [7] Proposes a bottom-up method that aims to reduce the difficulty and high cost of building and maintaining big software from scratch via Reverse engineering technique to extract design specification of an existing system from its database tables. It uses schema transformation pattern for service identification. He also applies CRUD matrix for service specification and CASE tool as an application to facilitate the conversion process from database tables to web services.

B. Top-Down Automated SIMs

LABAS (introduced by [8]) is a top-down service identification method from business models and business reference models to identify business process patterns via graph matching method by the search algorithm. The method firstly transforms the BPMN to the business process patterns manually, and then a search algorithm is applied to the business process patterns. Hence, an automated tool is needed for transforming the BPMN to business process patterns. In addition, although it uses BPMN as a standard business process model, the proposed method has not been measured the service quality factors quantitatively to determine the suitability of its identified services. [9] Emphasize on the lack of automatic service identification analysis, sufficient guidelines and the inability of manual SIMs to address service identification in enterprises due to the size and volume of processes. They propose an automated method for service identification based on parsing the labels of the business models’ activities to identify the activities as ‘verb’ or ‘noun’ by using the lexical database. The method provides algorithms for recognition of the labels’ structures. It is also based on assumption that process models are available in medium-sized and large enterprises. However, the method aims to solve the service identification automatically but, it was not considered to quantitatively calculate cohesion and coupling as service quality factors. The reusability is considered by counting the repetition of a label among activities via parser without consideration of relations between activities as a base of reusability calculation. In addition, it did not determine specific process model and how the activity labels can be derived automatically as well as it is not applicable to activity labels with non-English-languages. ASIM (proposed by [3]) is an automated service identification method from business domain supported via a tool that uses business entity as input and a set of services as output. The elementary business process, as a granular asset, represents lowest level of the enterprise process model. The first phase of ASIM is the creation of CRUD (create, read, update, delete) matrix of elementary business processes that determine the interactions between lowest level entities, and then the CRUD matrix entities are grouped in order to increase the abstraction level. ASIM applies Simulated Annealing as a generic meta-heuristic algorithm for service identification, and then the quality of services are measured by calculating service quality factors to cover reusability, cohesion, coupling, maintainability, and granularity, calculated based on CRUD matrix. ASIM relies on business entities as input but the selection of the lowest level of business tasks which belongs to most granular tasks decrease the clarity, and also indetermination of business process models lead to ambiguity of ‘How’ the high level business models can be decomposed to elementary business processes. In addition, the CRUD operations for each low level entity are not available as they are proposed by ASIM. Therefore, the preparation of CRUD matrix and determination of the
CRUD operations for each granular entity for a large volume of low level entities is a time consuming and costly activity. Their ASIM states that the CRUD preparation phase should be manually created via business architecture. Thus, the ASOM as ASIM automated tool only automates the quantification of service quality factors from low level business entities, and does not cover the whole service identification phase, as a heuristic-based algorithm has been proposed to be applied to the CRUD matrix, so that necessity of human involvement during the ASOM automated tool utilization is inevitable. [10] Introduce a framework using UML to identify services via an automated tool known as SQUID. SQUID identifies services from UML based on decomposing business processes hierarchically with emphasis on mapping represented in XMI. Then, the XMI is converted to MOF meta model, through XMI reader. However, it relies on UML which is not the most desirable model in the business domain. In addition, the four-step decomposition process which is done hierarchically is a human-based activity and is not supported by the automation tool. Also, there is a lack of guidelines for the four-step decomposition process. Furthermore, the importance of reusable and loosely coupled services has been highlighted, but SQUID does not provide how these factors can be satisfied with automatable formula. Moreover, no validation has been conducted through examples or case studies, and nor executable application has been presented. [11] Proposed a method for evaluating the quality of identified services quantitatively. The method models the business activities into a set of services, then the quality of service portfolio is measured based on cohesion, coupling and granularity factors. The results of measurement are evaluated based on normalization to set a tradeoff between values of each factor. Besides, an automated measurement tool has been presented which relies on metrics and weights of metrics that should be set by the user. The proposed method has not presented a clear process for identifying the service portfolio, while it emphasizes on providing guidelines for service quality factors measurement of the service portfolio via a set of formalized equations based on identified relationships in the previous phase. Therefore, its identification of service portfolio phase relies on human expertise for estimating the complexity of each entity that hinders the automation of the service identification process.

Furthermore, selecting UML leads to lack in supporting the other business process models such as BPMN. [12] Presented a semi-automated method for service identification focusing on business processes as input. The business processes in a BPMN standard are analyzed complicatedly creating a task dependency matrix. Service identification is performed based on pseudo code. The proposed service identification algorithm needs i) list of business processes and tasks and ii) the task dependency matrix that weights the relations between tasks. The candidate services are refined based on aggregation of identified services and involvement of designer to select appropriate aggregation of services based on expertise and manual operations. They support the method with a tool, named P2SStool, which addresses the identification and refinement step in combination with designer’s interference. They categorize their method as semi-automated due to task dependency establishment which is a manual and costly process that requires expertise. In addition, the P2S tool specifically does not consider the people with low-level expertise due to the prerequisite skills for the proposed tool. Therefore, there is a clear emphasis on service automation to decrease required expertise as well as the human involvement. Furthermore, in recent years, the trend towards automated techniques by top-down SIMs has been increased. The majority of the mentioned top-down methods did not provide a solution for automating the preparation phase of their input types to be involved in their automated technique. Hence, this activity remains manual although the volume of the input types imposes much cost. Next section, highlights the automated tools among SIMs that try to provide automation for their methods.

C. Quantification of Service Quality Factors

Some SIMs depend on the descriptive reasoning of quality factor satisfaction, while other SIMs formulate the situation of those factors to achieve clear numerical-based values for each factor as their weights [3], [13]. The evaluation of service quality factors based on formal methods is reasonable because they provide a numerical measurement of the situation of each quality factor [3], [11]. The possibility of utilizing numerical metrics of quality factors depends on the techniques selected in a SIM. Obviously, guideline-based service identification techniques descriptively assess the quality factors, whereas formal-based techniques have the necessary infrastructure to show each factor’s condition numerically.

Cohesion and coupling are subject of close consideration in service modeling domain as well as O.O or component-oriented domains. The goal is to achieve high cohesive and loosely coupled services. These factors are subject to frequent discussion in existing SIMs. However, lack of comprehensive metrics for assessing these two factors is obvious [14].

In some SIMs, achieving cohesive and loosely-coupled services is performed through mathematical methods that concentrate on relations between service’ entities to calculate the numerical value of each factor [5], [14]. Apparently, formalization of each service quality factor depends on the definition offered in each method. [3] calculate the cohesion based on relation degree between CRUD of entities in a service via the ratio of the maximum strength of CRUD relations to existing relations weight. [11] measure the coupling based on formalizing two factors, first, aggregation of relations inside a service and second, the strength of the relations. To ensure that relations are focusing on one single task, the relations’ number is counted and this strength is estimated based on the designer’s experience.

The assessment of the reusability of entities was considered a difficult task in previous software methodologies [5]. Reusability of services has effects on development and maintenance costs depending on the number of single-use services as well as the number of consumers that reuse a service [15]. The top-down methods differ in terms of service reusability measurement calculating the reusability based on relations and calls between business-aligned entities such business tasks, while
bottom-up SIMs calculate the reusability based on relations and calls between software components. Service granularity is another crucial factor that its optimal condition depends on other factors specifically cohesion and coupling. Cohesive services can lead to granular services and increase the number of services, and consequently, raise the management costs [16]. However, due to lack of detailed guidelines and validation through case studies, this idea appears to be immature. Also, most of the works covering the service granularity level focus on tightly coupled activities. Some studies use CRUD matrix to find the level of relationships between tasks or processes [17]. Granularity determines the functional scope of a service. Making a tradeoff between factors that affect granularity is a multi-criteria domain. Coarse-grained services can support more number of tasks, but they are less flexible and more difficult to reuse. In addition, reusability level can take a role to make a tradeoff in the granularity of services. What is more, the flexibility to mix with other functions or domain areas, as well as complexity, affects the service granularity. The quantification of service quality factors is an increasing trend in SIMs due to its crucial role in the clear measurement of quality of services. There is few top-down SIMs that adopt quantitative measurement of quality factors; however, till now there is no top-down SIM to cover quantitatively the measurement of the service quality factors based on the specification of high-level business entities. In addition, among top-down quantitative methods, those with automated tools including service quality factors calculation are not fully developed.

D. Availability of Tools Support SIMs

In the analysis and design phases of object orientation, the CASE tools are involved to generate the source code of a system. Similarly, service solutions should emphasize on automating the code generation by service discovery, selection, and composition mechanisms [7]. Utilizing automated tools will result in a decrease in the complexity and implementation costs of service modeling [7]. In addition, effective implementation of a SIM depends on the involvement of automated tools [13]. The most automation is realized in bottom-up methods since their input types can be executed and taken effortlessly. However, regarding the importance of business processes and necessity of involving them in the business-aligned SIM, their automated transformation to services through top-down SIMs has attracted recent researchers [3], [9], [12]. Therefore, some of top-down SIMs that have prepared tools to support their methods will be highlighted in this sub section. Table 1 lists the service identification automated tools existing in some of automated SIMs in which the emphasis is on top-down SIMs that were discussed in the previous section. Table 1 shows the automated tools of the automated and top-down SIMs discussed in the previous sub-section with their main contributions. In addition, in order to represent the comprehensiveness of each automated tool, the automated parts in SIM is added indicating the automation degree based on automation of preparation (P), technique (T), service quality factors (S) and service refinement (R) as popular service identification phases.

<table>
<thead>
<tr>
<th>No</th>
<th>Reference</th>
<th>Tool</th>
<th>Description</th>
<th>Automated parts in SIM*</th>
</tr>
</thead>
<tbody>
<tr>
<td>[10]</td>
<td>SQUID</td>
<td>Service discovery tool to identify services via business process, UML to MOF conversion.</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>[11]</td>
<td>Clustering Tool</td>
<td>To transform software components in legacy systems to service based on applying clustering technique on obtaining the legacy systems architecture.</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>[3]</td>
<td>ASIM</td>
<td>Transforming entity business process models to service set based on CRUD matrix and formalization of service quality factors.</td>
<td>T,S</td>
<td></td>
</tr>
<tr>
<td>[9]</td>
<td>Parser Technique</td>
<td>Based on parsing the labels of the business models’ activities.</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>[12]</td>
<td>P2S tool</td>
<td>The tool implements the service identification based on value of each tasks’ relations.</td>
<td>T,R</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows the lack of comprehensive coverage of the service identification in the SIMs’ automation tools. In addition, the preparation phase that should transform the input types to an automatable domain has lack of automation efforts. Obviously, considering the volume of input types, the preparation phase remains un-automated which imposes the expertise involvement. Additionally, majority of SIMs’ tools does not support quantifying the service quality factors to enable the automation. Furthermore, the service refinement in all five automated tools in Table 1 is based on experts’ involvement.

The current issues in top-down automated tools are

- Lack of a comprehensive automated SIM that its automation covers as much as possible of the service identification activities.
- The proposed automated tools have not considered employees with low-level expertise in their tool.
- Necessity for expertise in current top-down SIMs within the service identification process increase the cost of service identification.
III. RESULTS AND DISCUSSION

Referring to literature review results, besides the automation issues, the quality of identified service is crucial. The current SIMs’ challenges were mentioned in previous sections that can be summarized in the necessity for an automated method, satisfying the service quality factors, realizing business-IT aligned services. This section presents ASIF as a top down and fully automated SIM which has distinguished by firstly, including of all necessary service identification activities as complementary phases of a SIM, scope determination and service refinement, secondly, service quality factors that include cohesion, coupling, reusability and granularity, thirdly, supporting the business dimension within service identification to evaluate if a SIM satisfies business-IT alignment via supporting the business processes and business goals [1].

ASIF consists of five distinguished phases, namely ‘scope determination’, ‘goal modelling and business process modelling’, ‘weighting’, ‘clustering’ and ‘refining candidate service list’. Whenever there is a need for identification of a service list or an existing service list needs to be updated, ASIF, as a rotational framework, can be applied. To establish the study framework, ASIF, the proposed phases were created and supported with related techniques.

The first phase is scope determination. To assess the necessity for the scope determination for service identification, a comprehensive literature review and survey were carried out. Regarding existing SIMs, only [13] put forth the idea of scope determination and stress that it should be consistent with business goals; however, they did not present details to address this issue. As a result of literature review, a technique called AHP was located and applied in this study. On the other hand, ASIF provides clear scope determination method based on AHP which supports its applicability via guidelines. In addition, the criteria set for AHP is proposed to fit the AHP functionality in line with service identification requirements.

The second phase deals with as-is to-be elicitation. To elicit as-is and to-be situations. In addition, the related technique known as BPMN was spotted as a standard notation. Another related technique called UML AD was utilized alternatively as a result of the literature review. The goal modelling technique known as KAOS was determined as consequence of comprehensive literature review. The result of this phase is an integrated model of business processes and goals indicating the relativity and contribution of as-is situation to the business goals. It provides the prerequisites of subsequent phases and guarantees the business-aligned services identified according to these inputs.

The third phase which copes with weighting quality factors (Reusability, Cohesion, and Coupling) was constructed. Accordingly, a related technique called Weighting via Quantification was created by conducting thorough literature review.

The fourth phase of the ASIF, which deals with clustering, was conducted by carrying out a comprehensive review of state-of-the-art articles. ASIF proposes quantitative calculation of the cohesion, coupling, and reusability based on relations between business processes’ tasks as business aligned entities. According to literature review results, the granularity can be satisfied via a trade-off between cohesion and coupling values. Then, the customized version of Bunch clustering algorithm is applied for finding the integration of relationships or interactions between entities. The weighting and clustering phases of ASIF support automatic quantification of service quality factors and utilizing the clustering algorithm.

The last phase which is known as Refinement (Fig. 1). In addition, according to [9], the service refinement is addressed based on behavioral constraints of each identified service that facilitates the service refinement using human based activity, but they do not provide illustrations about the service prioritization phase. Furthermore, [12] stress on cohesion and coupling status of service candidates in order to refine service candidates automatically. Similarly, [13] conduct the refinement step to decide between realizing the candidate service as service or implementing it based on previous techniques such as Sun Enterprise JavaBeans that rely on experts’ experience. They utilize Litmus Test that consists of a set of questions used for service refinement. However, the questions and the process of refinement have not been declared. In addition, the answering of questions is based on experts’ knowledge and it has not been facilitated to address the questions quantitatively to make the answering process clear. Alternatively, ASIF utilized the goals affinity factor, a set of questions are proposed to help the service refinement based on Litmus Test. Consequently, the Litmus Test was selected. The litmus test’s questions proposed by ASIF mainly focus on refining the identified services based on ‘goals affinity factor’ of each candidate service.

Hence, based on the above justifications and descriptions, the framework is proposed graphically shows the sequence of ASIF phases in a recursive form. Obviously, the image shows the rotational characteristic of the ASIF that starts from ‘scope determination’ phase and ends with the ‘refinement of candidates’ services. Therefore, when there is a demand for new services or upgrading the existing ones, the presented cycle should be executed.
IV. CONCLUSION

There is a clear emphasis on service automation to decrease required expertise as well as the human involvement. Besides, in recent years, the top-down SIMs have increasing trends towards automated techniques. The majority of such top-down methods did not provide a solution for automating the preparation phase of their input types to be involved in their automated technique such as clustering. Hence, this activity remains manual although the great volume of the input types imposes much cost. Table 1 shows the lack of comprehensive coverage of the service identification by the existing SIMs’ automation tools.

The current issues in top-down automated tools are: lack of a comprehensive automated SIM that its automation covers as much as possible of the service identification activities, the proposed automated tools have not taken into account employees with low-level expertise in their tool development, and necessity for expertise in current top-down SIMs within the service identification process increases the cost of service identification.

Likewise, some of SIMs that provide quantitative methods to assess the quality of their services presents automated tools. In order to realize a top-down SIM, it is important to calculate the service quality factors based on business-aligned input types such as business process models, while existing SIMs calculate the cohesion and coupling according to technically oriented input types or atomic elements of business aligned input types such as CRUD matrices extracted from high abstract input types. Consequently, none of the existing SIMs encompasses all automatable steps in a comprehensive method with sufficient applicable details and guidelines. In fact, they have not covered all of the necessary steps of service identification.

This paper has presented ASIF to support service identification with regard to their goal affinity so as we can visually recognize the importance of services and also the probable gap in identifying services to be considered as required services. In addition, addresses the automation of service identification by preparing its prerequisites based on utilizing input types and techniques that support the automation. In the same vein, the quantification of effective elements in service identification is highlighted.

Future research should consider more quality factors. Indeed. ASIF tool can cover service specification phase to expose services in WSDL form.
REFERENCES


