

Khalid Mirghnee Yousif Mohammad <sup>1</sup> University of Bahri, Sudan

David Budgen<sup>2</sup>

A Review of Experience-Based **Knowledge about Web Service** Composition

A Review of Experience-Based Knowledge about Web Service Composition

# **Abstract**

**Context** Web Service technology offers a potentially powerful paradigm for creating software 'on demand'. However, to realise that potential, it is necessary to employ effective ways of dynamically composing atomic services to form composite services.

**Purpose** To identify the technologies and strategies currently being explored for organising the composition of Web Services, and to determine how extensively each of these has been demonstrated and assessed.

**Methodology** We undertook a mapping study of empirical papers that had been published over the period 2000 to 2009, using four electronic archival databases.

**Findings** We found 1172 papers that matched our search terms, from which 206 were classified as providing practical demonstration of their ideas. These were then categorised according to the approach used for composition.

**Conclusions** We have identified several approaches to service composition that are currently being investigated. No one approach seems to meet all needs, but a number merit thorough investigation.

**Keywords** Systematic literature review, Mapping study, Web service, Web service composition

# Introduction

The emergence of Web Services as a platform-independent basis for distributed systems is encouraging companies to run their business services over the internet. Many techniques and tools are being developed for building those systems or connecting business-to-business and business-to-customer applications using distributed service models. As an important element of this, researchers are exploring ways to build new services from existing services by composing services from a range of resources. In addition, the emergence of standards and languages that support Web Service composition allows many promising approaches to be explored (Milanovic, 2004).

We can classify the way that services are linked together as either static composition, which involves binding requests to specific services at construction time, or dynamic composition, where they are bound at run time (i.e. when the service is delivered). Static composition has many parallels with component-based development, in that the processes of discovery, selection and composition are performed by the developer in advance of actual use. Important as this is, since the design decisions are made by a human, static composition offers few new challenges for the researcher. In contrast, for dynamic composition to be effective, some or even all of these tasks need to be performed with the minimum of human intervention.

The research challenge this then creates is how to find the best way to perform dynamic composition of services, where these may well be provided by distributed autonomous agencies and the process of composition involves selecting from them "on the fly" in order to meet specific run-time requirements.

Our motivation for undertaking this review was to identify the approaches to service composition that are currently being proposed, and to identify the most important issues that need to be addressed in managing the process of service composition. Another goal was to obtain initial pointers to the most promising approaches and solutions from those that have been proposed as the basis for service composition. The research questions adopted were therefore:

- What are the main issues that need to be addressed if dynamic service composition is to be successfully implemented and widely adopted?
- What solutions have been proposed to deal with the issues raised?
- Which research methods have been used to investigate the proposed strategies?
- What gaps are there in current research?

To address these, a systematic literature review to collect evidence related to service composition was employed, in the form of a mapping study. Since "Web Service" is widely used as a synonym for "software services" we restricted our study to this class of service. In the rest of this paper we describe how this was organised and performed; report on the results; and provide a classification of these.

## **Defining and basic standards for Web Services**

A Web Service (WS) is a self-contained piece of software which can be viewed as being a procedure or subprogram that can be accessed over the internet by using existing internet protocols, which use a URL to access the WS (Dustdar, 2005). In addition, the World Wide Web Consortium (W3C) has approved many standards which play an essential role on WS development. Those standards were proposed to overcome the drawbacks of previous technologies for distributed systems such as the lack of support for heterogeneity, and the standards make the WS a promising solution for distributed systems. Moreover, by complying with these standards, the developers can continue using their own experience and technologies to develop distributed systems with no more worries about interaction with other heterogeneous systems (Erl, 2004; Potts, 2003). The following are the major technologies used for developing Web Services:

- XML (eXtensible Markup Language): The idea behind the design of XML is to provide generality, simplicity, and usability over the internet. It is a text format used for presenting data in a format that can be widely understood. While the essential focus of XML is centred on describing a document, it is also used to represent a user's defined data structure that can be used elsewhere, for example in Web Services (Ray, 2003).
- SOAP (Simple Object Access Protocol): is a protocol that is used to convey messages (exchange data) in the implementation of Web Services over the internet by using HTTP (Hypertext Transfer Protocol). Furthermore, it is an XML-based format, which means it uses XML

structures as the mechanisms for information exchange. This combined use of XML and HTTP (which are installed in almost all major OS) makes SOAP a widely usable solution for interaction between heterogeneous programs over the internet (SOAP, 2007).

- WSDL (Web Service Description Language): is a language based on XML which is used to describe Web Services. WSDL characterises a Web Service as a set of ports, or endpoints that work on messages enclosing either procedure-oriented or document-oriented information. In addition, in a WSDL file, messages and operations are described in an abstract way, and are separated from their real implementation, which allows them to be reused. WSDL is used to describe the WS interface that can publish to the public (W3C, 2001).
- UDDI (Universal Description, Discovery and Integration): Microsoft and OASIS, as providers of the public UDDI registry, have defined UDDI as "a registry service for Web Services and for other electronic and nonelectronic services. A UDDI registry service is a Web Service that manages information about service providers, service implementations and service metadata. Service providers can use UDDI to advertise the services they offer. Service consumers can use UDDI to discover services that suit their requirements and to obtain the service metadata needed to consume those services" (OASIS, 2010).

We based our study upon the guideline from Kitchenham and Charters (2007) and the reported experience from Brereton et al. (2007) who have studied those domains that are currently using systematic literature reviews, such as medicine and education, ("and particularly where the purpose of the review is to examine

the extent to which empirical evidence supports or refutes hypotheses") in order to adapt the techniques for software engineering. Since the need was to identify the scope and form of existing experience, we have used a "broad" form of systematic literature review, termed a "mapping study".

The purpose of a mapping study (sometimes termed a scoping review [Petticrew and Roberts, 2006]) is to identify the set of studies that exist for a topic and to categorise them. Such a review addresses broader research questions than a full systematic literature review, with the aim of identifying "gaps" (where primary studies are needed) and "clusters" (where a systematic review might be applicable). The outcomes of a mapping study are usually in the forms of categorisation and count of primary studies and do not normally involve aggregation of the study outcomes. The review protocol is used to specify the process that will be used to perform the review. In conducting our mapping study, the review protocol addressed a number of activities associated with planning a review such as formulation of the research question, identification of the need for a review, research strategy, classification of papers, study selection criteria, study quality assessment and data extraction strategy and process. The following sections describe some key elements from the protocol.

## Identification of the need for a review

Service-based technologies have only a relatively short history (Brereton et al., 2007). They are still developing, and experience about what might be effective is scattered across many sources. Since dynamic composition is an important (and technically challenging) feature of service technology, we wanted to find out what

was known about performing this. The aim of the systematic review was to find as many primary studies relating to service composition as possible by using an unbiased and objective search strategy. The rigour of the search process is one factor that distinguishes systematic reviews from traditional reviews.

#### **Search strategy**

To identify papers of interest, we needed to use keywords (index terms) that would capture the essence of the research topic. The research questions were aimed at providing the initial indicators to possible approaches and solutions that have been proposed. The scope of the review was defined by:

- Population: is the published scientific literature reporting on Web Service composition.
- Intervention: are the studies demonstrating service composition practices; techniques and processes.
- Outcomes of relevance: the quantity and type of evidence relating to various service composition techniques and processes.
- Experimental design: any form of empirical study.

The search strategy consisted of using a set of relevant keywords to search for primary studies and to identify the population, intervention and outcomes. The research strategy was to identify a set of electronic databases as principal sources of primary studies, and to determine the relevance of papers found on the basis of the presence of particular key terms in the title, keywords list or abstract of a paper. To search for related papers we prototyped the use of four strings ("Web Service", "service composition", "Web Service composition" and

"dynamic Web Service composition"). We then decided to use the string "Web Service composition" as our search string, which encloses what we were looking for to answer the research questions. The advance search provided by search engines provides text boxes to enter our keywords and link them with the 'AND' operator to form the search string. In addition, we choose to search for this string in the "Title" and "Abstract" fields. To check the validity of our search string, we had identified a number of "known" papers beforehand and checked that we found them with the chosen string. We adapted our search strings to the interface provided by each search engine to seek papers that fit our interests. The start date for our search was restricted to 2000, as this was when the first draft of SOAP v. 1.1 was released. Our initial search covered papers published up to the end of 2009.

## Study selection criteria

The guidelines to conducting systematic literature review in software engineering recommend searching at least four electronic databases (Kitchenham and Charters, 2007). We used four major archival databases (IEEE Xplore, ACM, ScienceDirect and CiteSeer) on the pragmatic basis that they encompass a wide range of sources for published papers relevant to service composition. In particular IEEE and ACM are major publishers of conference proceedings, which is particularly relevant for an emerging topic such as software service technology.

A systematic review of the literature should cover all potentially relevant resources; therefore we chose to include papers from conferences, journals and workshops. To avoid irrelevant papers we excluded literature that was only available as

abstracts or in a slideshow format, and we also excluded studies that do not describe or make use of services composition and those that do not have an experience element, as well as excluding duplicates of papers that were found in more than one database.

## Guideline to classifying the collected papers

The number of papers retrieved was moderately large (1172 papers) and after applying the inclusion and exclusion criteria, the remaining 206 "experience" papers were classified by using the following attributes:

- The type of study (see Appendix A).
- Whether composition was dynamic / static / semi-dynamic.
- Publication type (conference / journal / workshop paper).
- The research method used (see appendix A).

For the research method used here, the technical reports by Kitchenham and Charters (2007) and Brereton et al. (2007) provided valuable practical guidance for conducting the SLR. The basic analysis phase consisted of putting each paper into one specific category, depending on which form of knowledge they presented. To perform this, we had to analyse each paper carefully, as some papers presented many kinds of knowledge (e.g. a new technique combined with an experiment) which meant a paper might be classified in different categories. For the purpose of this study we sought to put each paper in just one category. Categorisation is useful for studying the characteristics of Web Service Composition (WSC), as it offers a practical way to learn about each characteristic of WSC individually.

Wherever possible, we sought to classify papers by using the information provided in the title and abstract. Most papers explicitly mentioned the form of composition they were addressing (static, dynamic or semidynamic) in their title or abstract. Many also provided information about the form of study, making it possible to classify them by research method too. However, for a proportion of papers, the poor quality of the abstracts did require us to consult the full paper in order to classify them. Classification was undertaken by one of us (K Y-M), consulting the second author when anything was unclear.

## Study quality assessment

There was no assessment of study quality, as we wanted to ensure maximum coverage at this early stage of the mapping study. We decided to include in the review all those papers from selected sources that were considered relevant. However, given that all sources were ones that involve a process of peer review, good enough quality was assumed.

#### **Data extraction strategy**

The mapping study exercise was designed to allow for the categorisation of published literature; and so the data extraction strategy was restricted to simply obtaining the information required for classification. For our review we needed to extract data relevant to the research questions. We sought to identify and record:

- What were the issues raised or addressed.
- Whether the paper presented a potential solution and, if so, the nature of that solution.
- The research method used.

## The conduct of the mapping study

The conduct of the mapping study was the process of implementation of the study design, as described previously. This involved recording the outcome of the searches and noting any divergence from the design that occurred. Before starting to collect papers from each database, we familiarised ourselves with the relevant search engine's interface when we prototyped the search string that we used to query that specific database. The interfaces are somewhat different, but they provide the same capabilities that are needed to run our query. Each interface has a box to enter the search string. We looked to find the search string in the title or the abstract of the paper. There is also an option about the publication year that we utilised where possible to limit the search period to 2009-2000.

We created a form to record information about each paper we selected. The information included, beside the standard citation data, was a description of the models and strategies used, and the type of the composition (static or dynamic). We followed the pre-defined inclusion/ exclusion criteria while searching the four publishers' databases. To include a paper into our set, it should be presenting some kind of practical experience about WSC. After that, we went through the outcomes again to ensure that all remaining papers met our criteria and also to remove any duplicate entries out of the final list.

## Results of the mapping study

After applying the inclusion and exclusion criteria, the 206 used papers were classified using different sets of categories. At the same time, some classification criteria with regard to mechanisms commonly used in service composition were identified.

The total number of papers demonstrating the proposed techniques in some way is shown in Table 1. The figures show that most experience papers were found in IEEE and ACM electronic archives database, and CiteSeer had nine papers after removing the duplicated papers that were also found in former electronic databases.

## **Classifying the papers**

We extracted the experience papers from the papers found during our search, and for each paper, identified the research method used (Brereton et al., 2006). The papers were then grouped by the approaches used for composition.

## Classification by research method

Each paper was classified by the form of the research method and analysis used in the paper. This, combined with the Web Service Composition strategy, was used to identify the data required to answer the third research question (see Table 2).

In the following sections, we have chosen three of the categories from Table 3 (2, 3 and 4) to discuss in detail as well as some of the papers from category 1 (Framework). Category 2 addresses the implementation of some techniques that developers have presented as solutions for services composition. The third and fourth categories from Table 2 were found to be very useful as mechanisms for testing the efficiency of proposed frameworks. They help to study most of the factors involved in building a potential solution for service composition.

Electronic database	Total papers per database	No of experience papers
IEEE Xplore	398	91
ACM	470	76
Science Direct	86	30
CiteSeer	218	9
Total	1172	206

**Table 1.** Total papers per database: the numbers obtained by using our keywords

No	Electronic	Static	Dynamic	Semi- Dynamic	Total
1	Framework/ Conceptual analysis	32	19	2	53
2	MIS tool/ Conceptual Implementation	45	52	7	104
3	Simulation	5	5	-	10
4	Experiment	14	10	-	24
5	Case Study	8	7	-	15
	Total	104	93	9	206

**Table 2.** Papers classified by research method

Classification by composition method For this, the experience papers were classified into three categories, depending on the service composition models (how these models address different application areas and requirements). These were:

- Static approach: this category contained papers that describe or provide approaches that Web Service providers might utilise to offer their services to others. It also contained approaches that need intervention from the user to develop or compose the service. Most proposed approaches fall under this category.
- Dynamic approach: forms of composition in this category involve the use of semantic descriptions of Web Services. The process of selecting and controlling the final set of services (binding) takes place at "run-time" without intervention of the users.
- Semi-Dynamic Approach: in this approach the Web Service developer will intervene in some stages of composition, but binding still takes place at run-time.

The classification of the selected papers with regard to the different approaches is detailed in Table 3.

Electronic	Static	Dynamic	Semi- Dynamic	Total
IEEE Xplore	47	39	5	91
АСМ	37	37	2	76
ScienceDirect	18	11	1	30
CiteSeer	2	6	1	9
Total	104	93	9	206

**Table 3.** Papers classified by composition form

## Other classification strategies

Although the practice that we adopted for classification (determining the set of categories through a mix of theoretical models and observation) is widely used for mapping studies, we should note here that other strategies have been suggested. In particular, Wieringa and Heerklens (2006) propose classifying papers according to a conceptual scheme that partitions papers into design or research categories, according to the type of problem that they address (world problems for which we are seeking a solution, and knowledge problems for which we are seeking information). They then use different criteria for the detailed classification of each type of paper.

Most of the papers selected for this survey are design papers, proposing ways of addressing the problems of composition. Although we could have adopted the scheme of Weiringa and Heerklens, we concluded that a more conventional categorisation was better suited to the set of research questions that we had identified.

# **Discussion**

In this section we first examine the likely threats to validity and then discuss the issues presented by our categorisation of the available experiences of service composition.

## Threats to validity

Given that the practices of systematic reviews are generally well established, we would argue that construct validity is not an issue here. Also, since we are not aiming to generalise our results beyond the rather specific topic of service composition, we do not need to consider external validity.

For internal validity, there are two issues related to the processes that we followed that we need to discuss.

- Did we find all of the relevant papers?
- Did we correctly classify those we found?

Previous experience from conducting mapping studies and systematic reviews suggests that the use of four electronic databases is usually sufficient to ensure that our search is likely to find the majority of papers (anecdotal experience suggests that up to %10 more can be found by snow-balling from the references of those found—although this is usually most effective when the search terms are less well established than is the case here). In addition, to reflect the relatively recent nature of the topic, we did take care to include databases that would access major conferences. We also prototyped our search strings and ensured that these included the main terms being used in this field. So, while we cannot be confident that we have found all possible papers that are relevant, it is highly likely that we have found the vast majority.

The second question is rather harder to address, since it implicitly includes both the inclusion/exclusion phase, and also the categorisation. Both of these tasks were performed mainly by one of us (KYM) with reference to the second author when appropriate. Where the quality of the abstract for a paper made it difficult to decide about inclusion/exclusion, we consulted the main paper. The subsequent task of categorisation involved reading the complete paper, and we are reasonably confident that after some discussion, we have categorised these correctly - if not always exactly as the original authors might have done!

#### **Classification issues**

We originally began this mapping study with the expectation that we would find a corpus of empirical work in the form of experiments, case studies and surveys that could form the basis for a fuller review, as occurs with some other topics in software engineering. However, we came to realise that the relative immaturity (and incompleteness) of the service paradigm, the variety of possible approaches to service composition, and the difficulty of performing comparative studies across these, constrained the scope to perform systematic evaluations of the form we had originally expected to find. We therefore broadened our searching to include all forms of "experience" related to service composition. A consequence of the very limited scale of systematic study through empirical forms was that the papers found offer little in the way of "comparative" evaluation of the ideas being investigated, in part because of the lack of any real baseline for comparison. (More generally, the

services community also seems to lack any really good and widely accepted "service scenarios" that can be used as benchmarks—which in itself must be a matter of concern.) Indeed, if we return to the framework proposed by Wieringa and Heerklens (2006) discussed earlier, then we can certainly confirm that the studies that we found were very largely world solutions rather than knowledge ones.

Perhaps the approach that offers the most scope for addressing the knowledge aspect is that of simulation, since it offers some scope for making comparison between different techniques, and hence providing the exploratory framework needed to provide a focus for more systematic knowledge gathering. However, this has not been the role for which it has been used in the papers that are reviewed here.

# Conclusion

In this section we review how well we are able to answer our initial questions, and what approaches might be appropriate for improving these answers.

Answers to the research questions As emphasised at the start of the paper, we set out to conduct a mapping study for which the aim was to find out how much of the published literature about WSC is actually supported by experience of some form, where this might be based upon implementation, experiments, prototyping etc. Approximately %25 of the studies were of this form, which suggests that the literature regarding composition is probably still largely based upon modelling rather than experience. With regard to the specific research questions that we asked:

• What are the main issues that need to be addressed if dynamic service composition is to be successfully implemented and widely adopted? Just over a third of papers addressed dynamic service composition, and used a wide variety of approaches. As identified previously, there is an urgent

need for an agreed set of benchmark scenarios that describe dynamic composition of a set of services in order to help with assessing how well different composition strategies work, and in what situations they work best. Clearly, the incorporation of semantic knowledge into the composition process is another key issue too, but beyond this, it is simply not possible to identify any clear patterns related to ideas about composition.

- What solutions have been proposed to deal with the issues raised? The solutions we found in our survey have been outlined in the body of the paper. Arguably, this preoccupation with world problems rather than with knowledge problems is actually an impediment to progress. At this point in time, we need a deeper understanding of the nature of service composition (and some good examples), rather than any additional design solutions.
- Which research methods have been used to investigate the proposed solutions? Table 3 shows that the largest category is that of conceptual implementation (which is commonly

the case for software engineering). The second largest category was that of frameworks, with more empirical forms making up the rest.

• What gaps are there in current research? The very scattered nature of research into this topic makes it hard to really identify gaps. Indeed, we might argue that there are too few clusters of studies to allow us to classify any gaps! When we posed this question, we did expect rather more experience to be found than proved to be the case, and so we cannot really answer this question on the basis of the outcomes from this study.

While this is perhaps not a surprising set of answers, there is at least one obvious conclusion that we can draw from it, namely that conferences and journals do need to emphasise the need for authors to include evidence about the effectiveness of a strategy wherever possible. The problem we face is not a lack of models, but a shortage of studies that seek to apply those models in some way and to assess how effective this is.

#### **Future work**

Our motivation for this mapping study was to identify where we could most usefully make a contribution to understanding the strengths and weaknesses of particular composition strategies. The service concept and its realisation through Web Services are still relatively immature. So it is not entirely surprising that we did not find a rich set of experiences from which to draw any conclusions.

Our own conclusion from the study is that a useful contribution would be the development of a simulation framework that would allow us to turn service composition into more of a knowledge problem, and hence to be able to make comparisons between different strategies—while accepting that the wide variation in these does impose some limitation upon the measures that we can use.

# References

Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M. and Khalil, M. (2007) "Lessons from applying the systematic literature review process within the software engineering domain", Journal of Systems and Software, Vol. 80 No. 4, pp. 583-571. Brereton, P. Gold, N.E., Budgen, D., Bennett, K.H. and Mehandjiev N.D. (2006), "Systematic literature review: a pilot study of service-based systems", Technical Report, Keele University, 2006.

**Dustdar, S. and Schreiner, W.** (2005) "A survey on web services composition", International Journal of Web and Grid Services, Vol. 1 No. 1, pp. 30-1.

**Erl, T. (2004)** Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services. Prentice Hall PTR, Upper Saddle River, NJ, USA.

**Kitchenham, B. and Charters S. (2007)** "Guidelines for performing systematic literature reviews in software engineering", Technical report, Keele University and Durham University.

Milanovic, N. and Malek, M. (2004) "Current solutions for web service composition", Internet Computing, IEEE, Vol. 8 No. 6, pp. 59-51.Petticrew, M. and Roberts, H. Systematic reviews in the social sciences: A practical guide, Malden, MA: Blackwell, 2006.

**OASIS. (2010)** Oasis UDDI specification tc. Accessed -21Oct2010-. Potts, S. and Kopack, M. (2003), Sam's teach yourself web services in 24 hours. Sam's, Indianapolis, IN, USA.

Ray E.T. Learning XML. O'Reilly & Associates, Inc., Sebastopol, CA, USA, 2003.

Soap version 1.2 part 0: Primer (second edition) *April 2007.* 

**W3C** Web services description language (WSDL) 1.1, March 2001.

Wieringa, R.J. and Heerkens, J.M.G. (2006) "The methodological soundness of requirements engineering papers: a conceptual framework and two case studies", Requirements Engineering, Vol. 11 No. 4, pp. -295 307.

# Appendix a (research method)

RESEARCH METHOD	DESCRIPTION
AR-Action Research	Researcher participates in the action studied. Differs from participant observation as researcher is aware that their presence will affect the situation they are researching
CA-Conceptual Analysis	Basic assumptions behind constructs are first analysed; theories, models and frameworks used in previous empirical studies are identified and logical reasoning is thereafter applied
CAM-Conceptual Analysis/ mathematical	CA using mathematical techniques
CI-Concept/ proof of concept Implementation / proof of concept.	Self-explanatory - 'We built it and it worked'
CS-Case Study	A method, tool or procedure under investigation is tried out on a real project using the (otherwise) standard methods/tools/procedures of the organisation
DA-Data Analysis Implementation	Analysis of data generated or published elsewhere
ET- Ethnography	Researcher participates in the action studied.  Differs from AR in that researcher has no intent of interfering in phenomenon, does not relate findings to generalisable theory and does not interpret from researcher's point of view
FE – Field Experiment	Extensions of laboratory experiments into the 'real world' of organisations/society. Independent variables are controlled
FS-Field Study	Examines data collected from several projects (or subjects) simultaneously. Less intrusive than case study. Usually less detailed than case study because aim is not to perturb the subject under study
ID-Instrument Development	Development of MIS (or other) instrument (e.g. user satisfaction questionnaire)

**Table A (summary):** Categories of Research Method defined in Brereton et al. (2006)

# **About the authors**

**Dr. Khalid Mirghnee Yousif Mohammad** is a University lecturer (Assistant professor), systems analyst and software engineer with 12 years experiences in deferent computer disciplines. Khalid is based at the College of Computer Sciences, University of Bahri (former University of Juba) – Sudan. His research interests include systematic literature review and Web Services.

**Prof. David Budgen's** research interests include: evidence-based software engineering (EBSE); evidence-informed software development (EISD); and software design. A long-term research interest has been the study of software design practices, and David is the author of the textbook Software Design, now in its second edition (2003), and published by Pearson Addison Wesley.