Testing of Web Services – A Systematic Mapping

Abhishek Sharma, Theodore D. Hellmann, Frank Maurer
Department of Computer Science
University of Calgary
Calgary, Canada
{absharma, tdhellma, frank.maurer}@ucalgary.ca

Abstract— **Web services have been gaining popularity since the introduction of Service-oriented architecture and cloud computing. With more and more legacy systems migrating to service-oriented architectures and the cloud, an urgent need for proper testing techniques is becoming apparent. This paper provides an overview of the current state of research into testing of web services. To understand this subject, we conducted a systematic mapping. The results suggest that research into testing web services is still in its early stages. We provide recommendations about holes in existing research that need to be addressed and directions for future research that will have maximum novelty and potential for impact on the field.**

***Keywords*— Systematic Map, Testing, Web Services, Service-Oriented Architecture**

I. INTRODUCTION

A web service, as defined by the World Wide Web Consortium, is "a software system designed to support interoperable machine-to-machine interaction over a network" [1]. The increasing popularity of web services can be attributed to a movement towards service-oriented architecture (SOA) and cloud computing. SOAs emphasize interoperability, decoupled components, and reuse. This can only be achieved using web services because web services can be used to develop components that are independent of each other and can be accessed by multiple service users simultaneously. Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [2]. Web services provide developers with the freedom to employ light-weight clients. This makes web services an ideal choice for applications that will be accessed from handheld devices like smart phones and tablets. This is important because users currently want ubiquitous access to applications.

This increase in the use of multiple components and third-party services has called attention to several pain points in the development process for applications relying on these services. Testing these systems for correctness and reliability has always been a challenge because of the black-box nature of web services. Because the source code of these services is usually unavailable, testing must be carried out without knowledge of the internal structure of the service. This can make it difficult to understand the root cause of errors. Without proper testing frameworks, practices like refactoring and iterative development are problematic because it can be difficult to detect regression errors quickly.

In order to better understand the current possibilities and limitations of testing of web services, we performed a systematic mapping study [3]. This mapping provides an overview of the current state of research, important implications for practice, open research issues, and areas for improvement. Hence the goal of this investigation is to identify, evaluate, and synthesize state-of-the-art testing practices in order to present what has been achieved so far in testing of web services. The rest of the paper is organized as follows: Section II presents related work; Section III describes the methodology we used to create and interpret our mapping; Section IV presents an interpretation of our results; Section V presents limitations and threats to validity; and Section VI concludes the paper and describes necessary future work.

II. RELATED WORK

Both systematic mapping studies and web services are quite new to the field of software engineering. This coincidence makes it difficult to find any existing systematic mapping studies that look into testing of web services. However, a few literature reviews were found that provide details about testing of web services. Zakaria et al. [4] present a systematic literature review of unit testing approaches for Business Process Execution Language (BPEL) which provides an overview of how BPEL is tested and verified for correctness. Canfora and Penta [5] identify open research issues in testing of service-centric systems. They also provide insights about the reasons why web services are becoming a popular choice for certain development tasks.

Ladan [6] classifies web service testing approaches through a survey of various testing approaches which have been used for testing of web services. Ladan [6] classified the research literature into four different categories based testing approaches: WSDL-based, Mutation-based, Test Modeling, and XML-based approach. He also excluded approaches based on formal methods and data gathering.

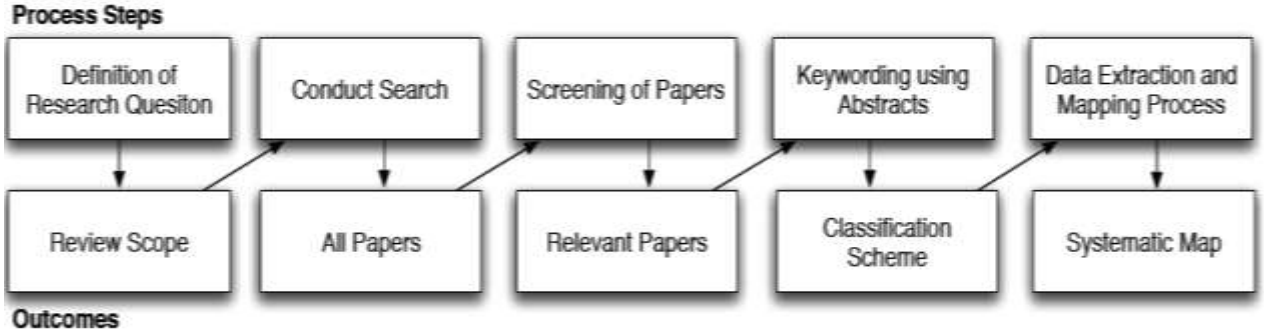

Figure 1. Systematic mapping flowchart from [3].

The issues with this study include too specific focus for a survey study and absence of formal protocol outlining the search and categorization. Further the final set included only 10 papers, which is a comparatively small set to provide any conclusions.

Rusli et al. [7] performed a similar study for web service compositions. This study provides an overview of testing of web service compositions. This paper also provided a comparison between various testing techniques that are used for testing web service compositions.

III. METHODOLOGY

Our study was conducted according to the guidelines provided by Petersen et al. [3]. Their five-step process, included here as Figure 1, can be described as follows:

A. Research Questions

Research questions define the scope and focus of a systematic mapping. This study was motivated by two research questions:

1) *What is the current state of research regarding testing of web services?* This question will provide an overview of what the research community has previously looked into in testing of web services.

2) *What techniques are used for testing of web services?* This question was included specifically to find all the testing techniques that are used for testing web services and estimate which one is more effective based on popularity.

B. Search Strategy

We performed a search using our initial search strings in well-known databases such as IEEE Xplore and the ACM Digital Library. Based on the search results from this pilot study, the scope of the study was refined. This also led to refinement of research questions and search strings. Further, the research questions were reformulated based on the preliminary results; this was done to ensure that the final results obtained are valid and reliable. Figure 2 shows the outline of the search strategy.

C. Data Sources

We used the following databases in our search for relevant publications:

1) *IEEE Xplore*: Covers electrical engineering, computer science, and electronic subject areas and provides full-text and bibliographic access to IEEE transaction, journals, magazines and conference proceedings.

2) *SciVerse Scopus:* A bibliographic database containing abstracts and citations for academic journal articles. Covers both ACM Digital Library and Springer Link digital libraries in addition to with other sources. Scopus includes approximately 18000 titles.

D. Data Retrieval

The search strings were created by combining sets of terms describing testing and web services. This can be

Figure 2. Search strategy overview.

represented as:

$(T_1 \text{ OR } T_2 \ldots..\text{OR } T_n) \text{ AND } (W_1 \text{ OR } W_2 \ldots\ldots\ldots\text{OR } W_n)$

where T covers most of the testing keywords and W includes keywords related to web services.

T: [("unit" OR "integration" OR "system") AND ("testing" OR "test")]

W: [("web service" OR "cloud service" OR "third-party service" OR "REST" OR "SOAP" OR "WSDL")

We used these terms based on the results of our pilot search, which found that using more generic search terms for testing like "test" and "testing" provided thousands of results with a high percentage of unrelated results.

The search strings were applied to the title, abstract, and keywords of papers. In order to be included in the set we used for our mapping, papers had to be peer reviewed and written in English. The inclusion and exclusion criteria were applied to papers by reading the title and abstract of each paper. For papers where these were not adequate to reach a conclusion, the introduction and/or conclusion were read. Duplicate papers were also removed during this process. Duplicates could be due to extended versions of papers submitted in journals or papers that were found in both the databases. A final list of 150 papers was acquired through this process. A list of these papers is available online at http://pages.cpsc.ucalgary.ca/~absharma/.

E. Data Extraction and Synthesis

The classification scheme was built using guidelines provided by Petersen et al. [3] which suggested that text should be explored when a paper's abstract does not provide adequate information. We used a similar approach but restricted this full-text search to the introduction and conclusion only.

Microsoft Excel and Endnote were used for data collection. The mappings were created using Excel. The keywords that were assigned to the papers can be categorized as follows: types of research approach, types of testing, and types of technology. These categories will be described in detail in results section.

IV. RESULTS

This section will provide detailed insights about the results of our systematic mapping as well as an interpretation of their significance. Our initial search was conducted in October 2011 on IEEE Xplore and Scopus and provided close to 2000 results. After separately applying first filter on results obtained from Scopus and IEEE Xplore, the number of results were reduced to 300. After filtering out duplicate entries, this reduced the number of results to 150 papers. The number of duplicate enteries was quite large and this can be attributed to papers being revised from conferences publications into journal articles, being extended and submitted in later conferences, and overlapping results from databases. Figure 3 shows the results obtained after each step.

After this, results were categorized according to the year they were published in order to understand the distribution of research over the time. Figure 4 shows the distribution of results. It is evident that an increase in the number of publications in this area occurred in 2009. This could be attributed in part to a deluge of light-weight clients like the iPhone or to a general interest in topics such as cloud computing. The results indicate that testing of web services has gained substantial attention in the last four years.

A. Classification by Research Type

The classification scheme provided by Wieringa et al. [8] was used as a basis for determining the research type for the set of papers. As described by Wieringa et al. [8], this classification is supposed to be a generalized classification and can be applied to any type of research. The categories are described in detail below:

1) *Validation Research*: Techniques used are novel, but are only implemented in a lab environment.
2) *Evaluation Research*: Techniques or solutions are implemented in practice, evaluations are conducted and consequences are investigated in terms of benefits and drawbacks.
3) *Solution Proposal*: A solution for a problem was proposed. This solution can either be novel or a significant extension of an existing technique. The potential benefits and applicability of the solution are shown by a small example or a good line of argument.

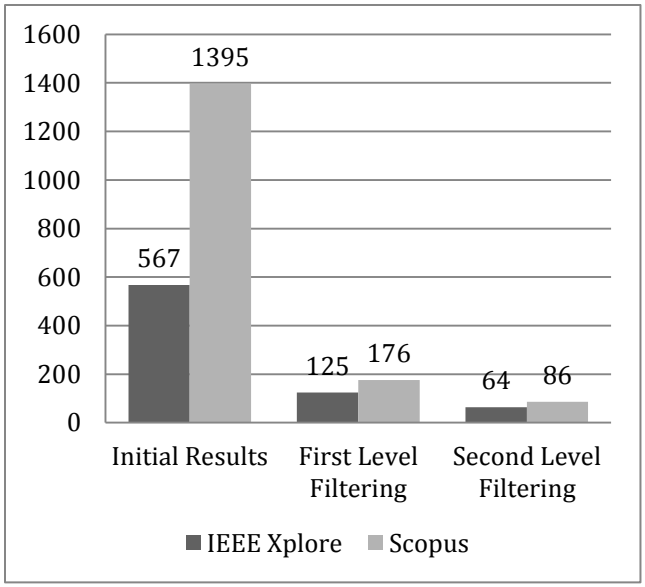

Figure 3. Search results.

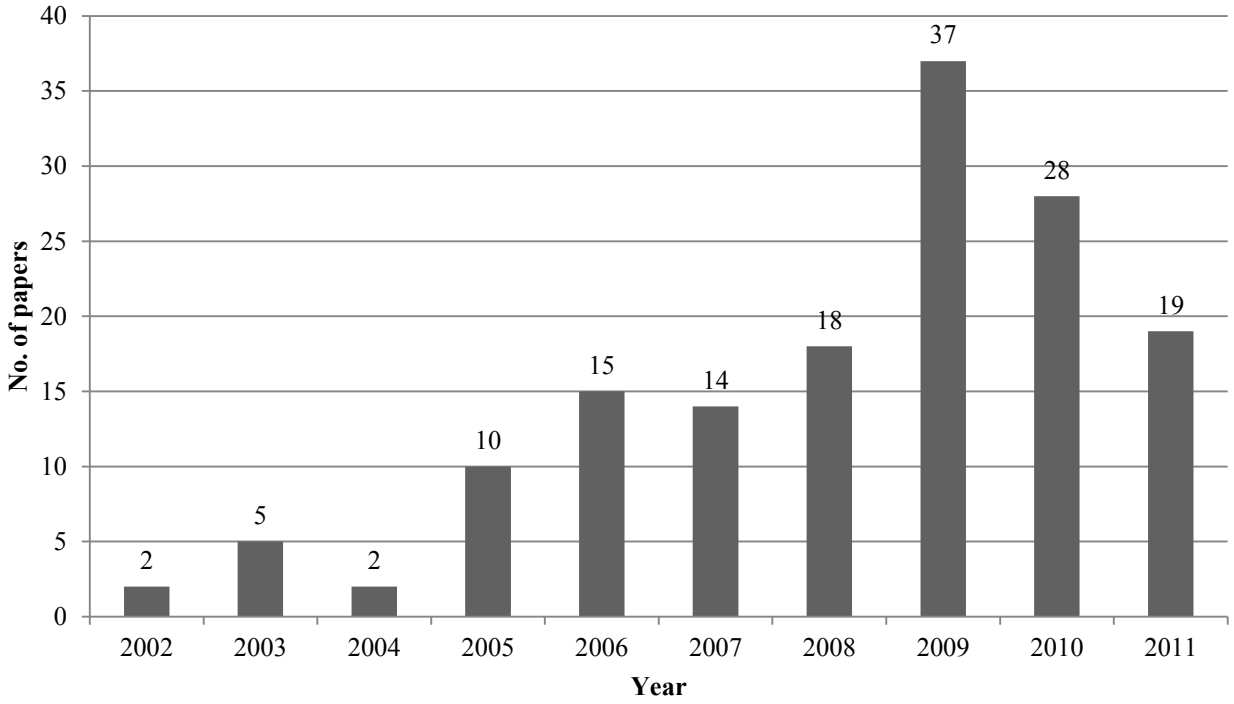

Figure 4. Distribution of literature across years.

4) *Philosophical Paper*: Structures the research area in the form of a taxonomy or conceptual framework and sketches a new way of looking at existing topics.

5) *Experience Papers*: Reports the personal experiences of the authors and explains what and how something has been done in practice.

6) *Opinion Papers*: Reports the personal opinion of the authors without relying on related work or research methodologies.

The results of this categorization are presented in Figure 5, which shows that the majority of the literature in the last four years is focused around validation and solution research. This distribution identifies a lack of philosophical and experience papers, which is a clear opportunity for future work. Experience papers can only be written when the techniques are actually implemented in industry and are used. We believe that the lack of this type of paper indicates that this field is still in an early stage and needs a significant amount of attention. Another trend that is evident is a focus on validation research; this shows that, with the introduction of this new technology, experiments are being performed to find efficient techniques for testing web services.

B. Testing Techniques Used

The second research question was aimed at investigating which testing techniques were reported in literature. A total of 40 testing techniques were reported by the papers, these are the terms that used by various authors. This seems to be a large number; however, it is possible that because of the newness of the field and a lack of standard terminology it could have been difficult to identify some relevant papers. Also, some of the testing techniques are studied as a part of validation research and are still not introduced in industry. Figure 6 shows the testing techniques we identified along with the number of times they were mentioned or used in research studies. It is clear that research community is interested in testing techniques based on formal specification. A reason behind this might be the ease of representation of web services as formal specifications. Model-based and State-based testing techniques are the next most frequently-used. This could be because the representation of the internal structure of web service as a model due to the black-box nature of web service encourages the use of testing techniques that are based on models and state diagrams.

One disturbing trend is the confusion between the concepts of unit and integration testing. A fair number of papers refer to techniques that clearly work on the logic of integration testing as unit testing. However, it is unclear from our mapping whether this is due to the misuse of these terms or to the inadequacy of these terms when applied to testing of web services.

C. Focus

There seem to be four focal points to which the majority of papers on testing of web services related. These are briefly described below:

1) *BPEL* – BPEL, or Web Services BPEL (WS-BPEL), is a standard executable language for actions within business processes of web services. The processes in BPEL export and import information by using web service interfaces exclusively. Testing of BPEL seems to be an important concern for the research community.

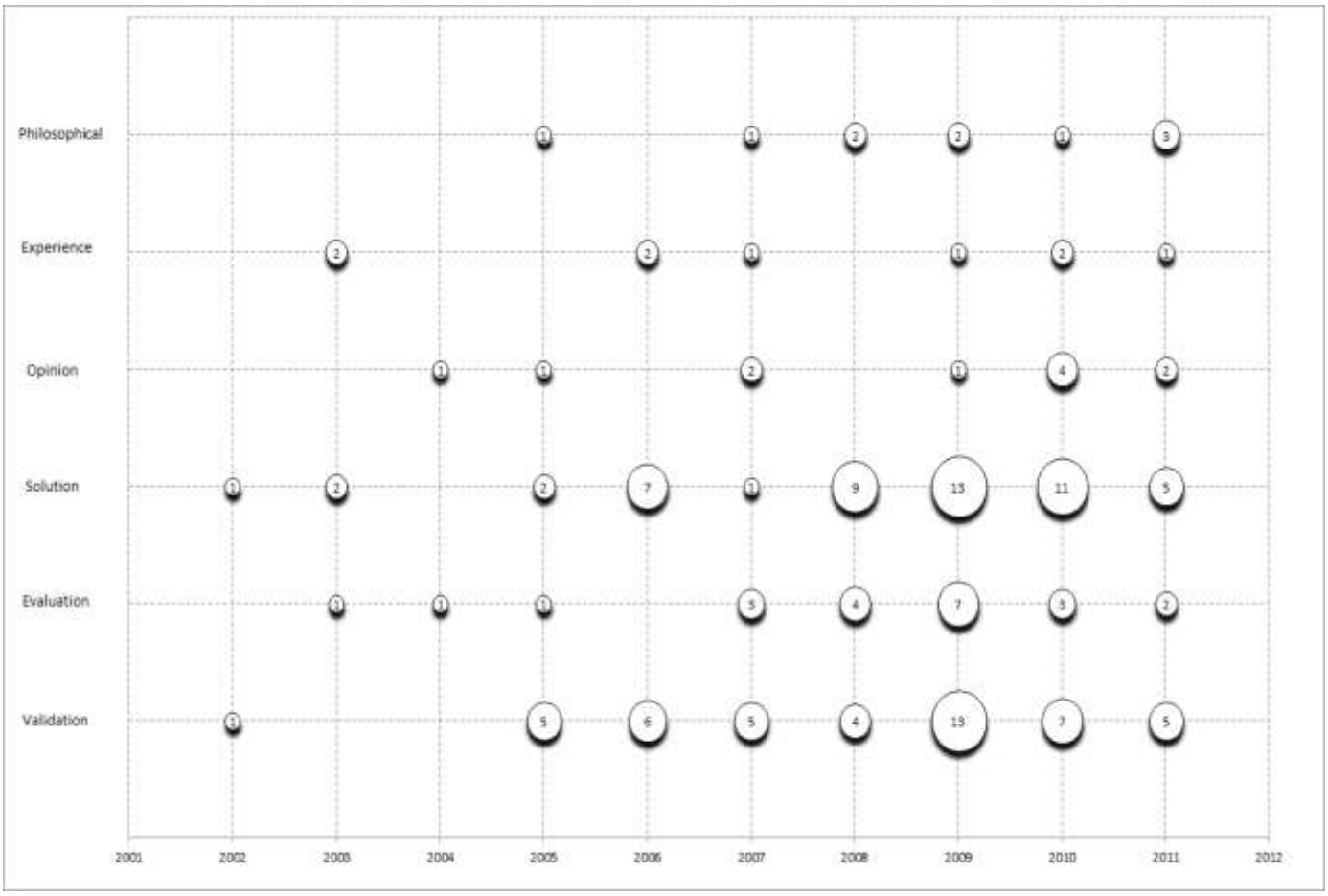

Figure 5. Research type distribution.

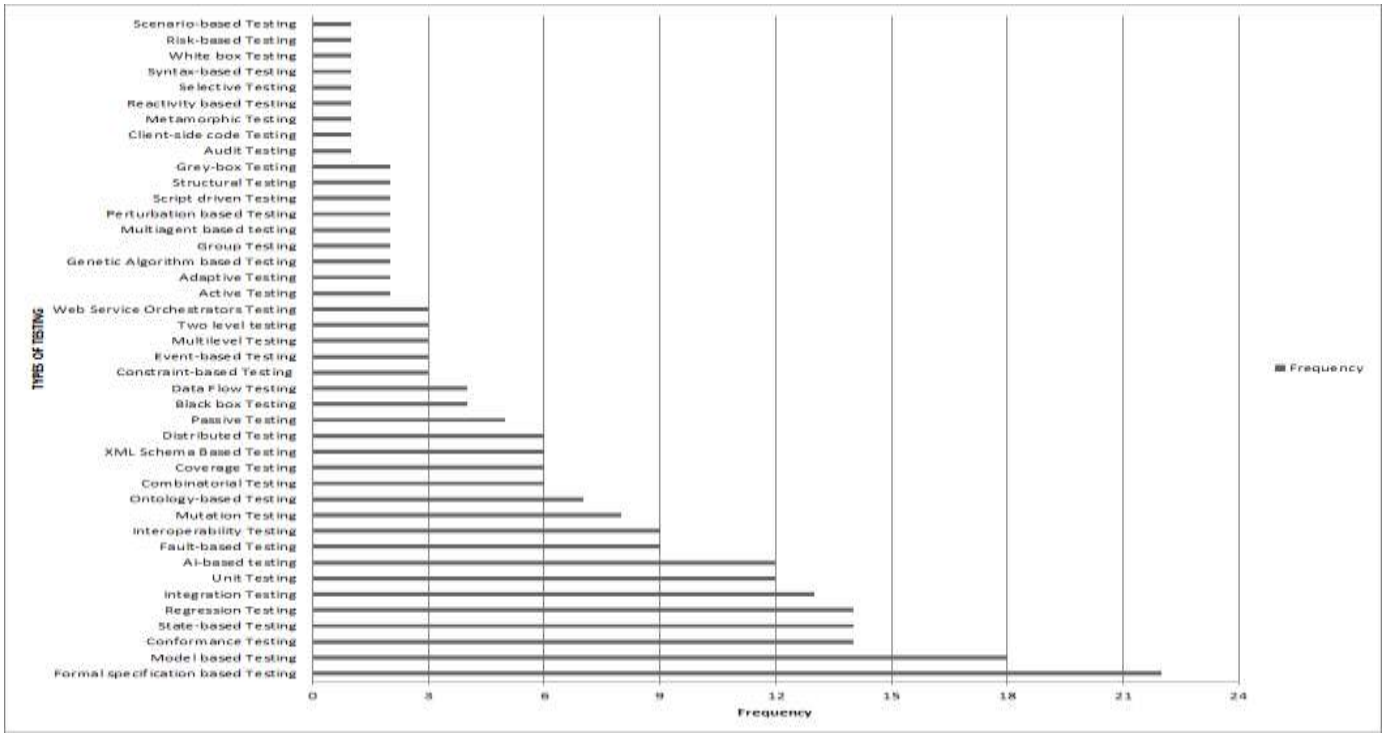

Figure 6. Types of testing.

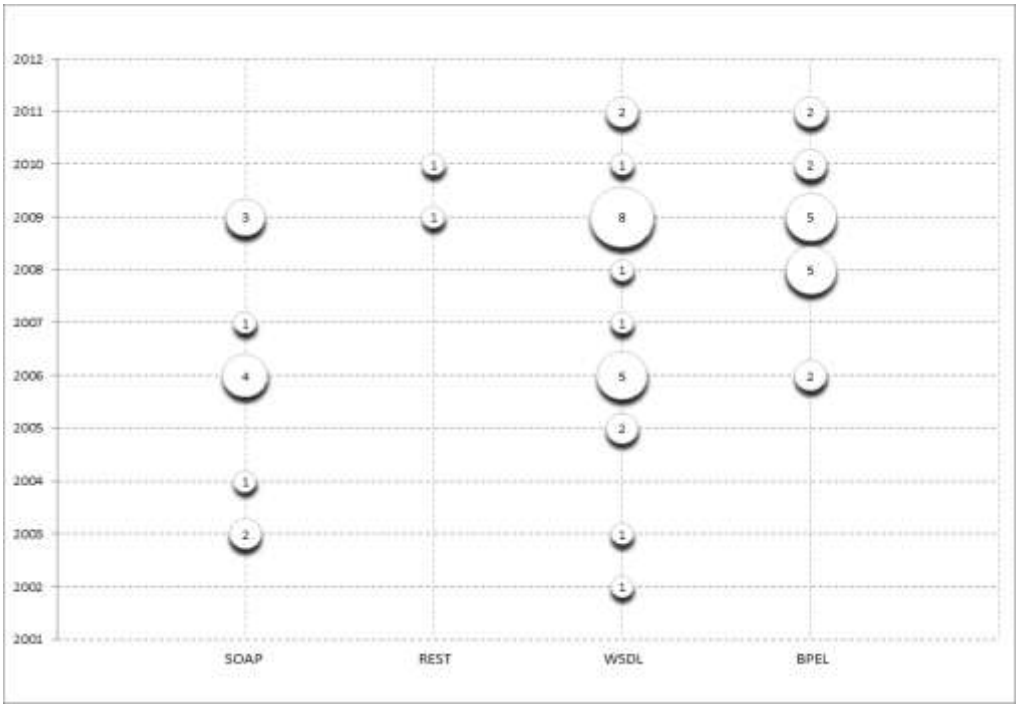

Figure 7. Distribution of focus.

2) *Web Services Description Language (WSDL):* WSDL is an XML-based language that is used for describing the functionality offered by a Web service. This is specified using a machine-readable description of the expected parameters are return values.

3) *Representational State Transfer (REST):* REST is a style of software architecture for distributed systems which was introduced and defined in the year 2000. A REST-style architecture is quite similar to client-server architecture – in both systems, a client initiates a request and the server processes the request and returns a response. The request and response in REST are built around the transfer of states of resources.

Simple Object Access Protocol (SOAP): SOAP is a specification for exchanging structured information in the implementation of web services. SOAP relies on XML for its message format and application layer protocols for message negotiation and transmission. SOAP has three major characteristics: Extensibility, Neutrality, and Independence

Figure 7 shows the distribution of focus over the years. From this diagram, we can see that testing techniques for WSDL have been steadily investigated over the years. BPEL testing is also being investigated enthusiastically in the last four years. Interest in testing of SOAP-based web service seems to be diminishing since no new literature investigated SOAP in the last two years. Finally, the research community has invested very little effort into testing of Web Services that are based on REST architecture. Figure 8 provides details about contribution and focus. From this figure we can see that researchers have more inclination towards developing frameworks than tool support.

V. LIMITATIONS

A major concern with any type of research is reliability. Since in this study only one researcher was involved in the analysis of results, the analysis is subjective and this might have introduced substantial bias. Further, since the full text off results was not analysed, this study could be biased due to misleading or inadequate abstracts.

Further, relevant results may have been missed because of the use of only 2 databases (IEEE Xplore and Scopus) in our search for papers. However, we are unaware of any papers that were missed.

Systematic mapping studies can provide a quick overview of research field; these studies can also provide preliminary answers for research questions. But systematic mapping studies are shallow in nature and should only be used to get a quick overview of the research area. One cannot rely on systematic mapping studies when a thorough and detailed analysis of a research area is required. Classifying literature based on abstracts may produce some misleading results based on quality of abstracts. This problem is quite common in systematic mapping studies; as stated earlier mapping studies should be done to get a very broad overview of the field and claims from these studies should be validated by extending them to literature surveys.

VI. CONCLUSION AND FUTURE WORK

This research study provided some important information about testing of web services. The results of this study yielded answers to our research questions:

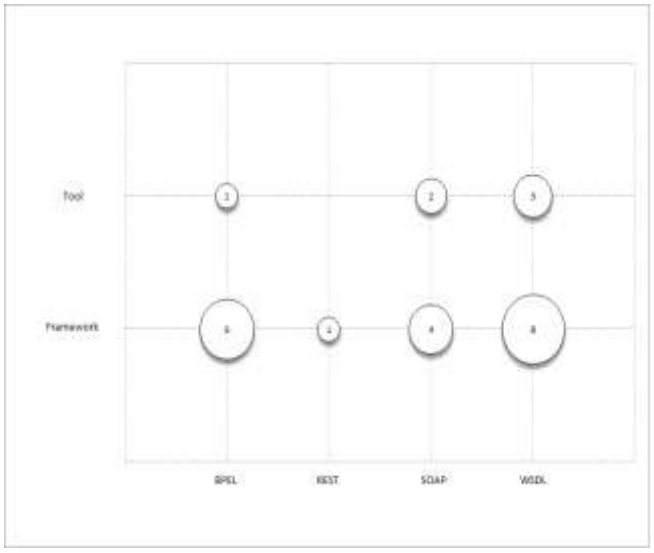

Figure 8. Contribution.

1) ***What is the current state of research regarding testing of web services?*** Figure 5 depicts the distribution of research across years. It can be deduced that, when it comes to testing of web services, more emphasis is given to practical research leading to solutions than is given to theoretical research like literature surveys. Because of large number of validation research studies, which correspond to the experiments implemented in lab and are not yet used in real industry environment we can conclude that this field has a lot of open research issues that are still being explored.

2) ***What techniques are used for testing web services?*** Figure 6 describes testing techniques that are being researched. There seems to be an overflow of testing techniques, due to a lot of testing approaches which might be quite similar at the detailed level but are reported with different names. This might be due to freshness of research in this field and due to lack of established terminologies. Identifying techniques that follow similar approaches was not possible by details provided at the abstract, introduction and conclusion levels. This is one of the prominent reasons that encourage us to extend this study to a systematic literature review.

Results from this mapping study provided a quick overview of the current state of research in testing of web services, but for detailed insights this study should be extended into a systematic literature review. In future work we will do so in order to generate a detailed taxonomy of testing techniques for web services.

REFERENCES

[1] W. W. W. Consortium. (2004, February) Web Services Glossary. [Online]. www.w3.org

[2] (2011, January) National Institute of Standard and US Department of Commerce Technology. [Online]. http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf

[3] Kai Petersen, Robert Feldt, Shahid Mujtaba, and Michael Mattsson, "Systematic Mapping Studies in Software Engineering," in *12th International Conference on Evaluation and Assessment in Software Engineering*, 2008, pp. 71-80.

[4] Z. Zakaria, R. Atan, A. A. A. Ghani, and N. F. M. Sani, "Unit Testing Approaches for BPEL: A Systematic Review," in *Asia-Pacific Software Engineering Conference*, 2009, pp. 316-322.

[5] G. Canfora and M. Di Penta, "Testing services and service-centric systems: Challenges and opportunities," *IT Professional* , vol. 8, no. 2, pp. 10-17, March-April 2006.

[6] M. I. Ladan, "Web services testing approaches: A survey and a classification," in *International Conference on Networked Digital Technologies*, Beirut, 2010, pp. 70-79.

[7] H. M. Rusli, M. Puteg, S. Ibrahim, and S. G. H. Tabatabaei, "A comparative evaluation of state-of-the-art web service composition testing approaches," in *International Workshop on Automation of Software Test*, 2011, pp. 29-35.

[8] R. Wieringa, N.A.M. Maiden, N.R. Mead, and C. Rolland, "Requirements Engineering Paper Classification and Evaluation Criteria: a Proposal and a Discussion," *Journal of Requirement Engineering*, vol. 1, no. 11, pp. 102-107, 2006.

[9] N. Laranjeiro and M. Vieira, "Extending test-driven development for robust web services," in *International Conference on Dependability*, 2009, pp. 122-127.

[10] P. Hamill, "Unit Testing Web Services," *Dr. Dobb's Journal*, vol. 33, pp. 53-58, 2008.