

# A Study on Third Party Component Reuse in Java Enterprise Open Source Software

Widura Schwittek, Stefan Eicker

paluno – The Ruhr Institute for Software Technology  
University of Duisburg-Essen  
Universitätsstr. 9  
45141 Essen

widura.schwittek@paluno.uni-due.de  
stefan.eicker@paluno.uni-due.de

**Abstract:** Recent studies give empirical evidence that much of today’s software is to a large extent built on preexisting software, such as commercial-off-the-shelf (COTS) and open source software components. In this exploratory study we want to contribute to this small but increasing body of knowledge by investigating third party component reuse in 36 Java web applications that are open source and are meant to be used in an enterprise context. Our goal is to get a better understanding on how third party components are reused in web applications and how to better support it. The results are in line with existing research in this field. 70 third party components are being reused on average. 50 percent of the 40 most reused third party components are maintained by the Apache Foundation. Further research questions based on the study results were generated and are presented at the end of this paper.

## Summary

Reusing third party components has become a key success factor in software development projects [Ga08] leading to reduced costs, faster time-to-market and better software quality [Li09]. Recent studies ([He11], [RvV12]) give empirical evidence that much of today’s software is indeed to a large extent build on preexisting software, such as commercial-off-the-shelf (COTS) and open source software components. This especially holds true for web applications, where architects and developers are faced with the steady proliferation of new technologies and standards [Ro08].

In this exploratory study which is described in more detail in [SE13] we want to contribute to this small but increasing body of knowledge of third party component reuse as part of the larger field of software reuse. For this we have analyzed 36 end-user products such as content management (e.g. Liferay, Magnolia) or business intelligence systems (e.g. Pentaho). We focused on black-box reuse in Open Source Java web applications that are meant to be used in an enterprise context. As such, all identified third party components (e.g. Apache Commons, dom4j, log4j, Lucene) also follow an open source license which excludes COTS components from this study. In this first step we did not consider references to components which are part of the runtime environment

such as the JDK or application servers. We also excluded non-Java based components such as the JavaScript library JQuery. A small tool has been developed in order to retrieve the reuse data by extracting, transforming and analyzing deployment artifacts in an automatic fashion. In short, it looks for characteristics such as the filenames and the checksum of files to identify the name and the version of a reused third party component.

During the identification of third party components in 36 Java based web applications 3311 different artifacts were considered leading to 651 unique third party components. 70 third party components are reused on average ranging from a minimum of 16 (Agorum) to a maximum of 161 components (Alfresco) per web application. The results are comparable to the existing studies, underlining that software reuse happens in this specific way of third party component reuse, independent from the kind of reuse (calling, extending, inheriting, instantiating).

As part of this exploratory study these questions were raised from looking at the data which are worthwhile to be asked as part of our future research: Do integrators keep track of the huge amount of reused third party components, especially with respect to bugfix and security updates and what are the risks of not doing so? Specific findings in the data might be of interest to integrators during their selection process: Some third party components are reused by almost all of the analyzed web applications. 50 percent of the 40 most reused third party components are maintained by the Apache Foundation. Certain reuse characteristics were sepecific to particular web application domains. How can this knowledge be used to support the selection process of third party components? Having laid the foundation with this work, we further look into how valuable findings in the data set can be made available to integrators. We already started to create component recommendations to support the component identification and selection phase [SE12].

## References

- [Ga08] The Evolving Open-source Software Model. Predicts from December 2008. Gartner.
- [He11] Heinemann, L. et al.: On the Extent and Nature of Software Reuse in Open Source Java Projects. In (Schmid, K. Ed.): Top Productivity through Software Reuse. ICSR 2011, Pohang, South Korea, June 13-17, 2011. Proceedings. Springer Berlin Heidelberg, 2011; pp. 207–222.
- [Li09] Li, J. et al.: Development with Off-the-Shelf Components: 10 Facts. In IEEE Software, 2009, 26; pp. 80–87.
- [Ro08] Rossi, G.: Web engineering. Modelling and implementing web applications. Springer, London, 2008.
- [RvV12] Raemaekers, S.; van Deursen, A.; Visser, J.: An Analysis of Dependence on Third-party Libraries in Open Source and Proprietary Systems: Proceedings of the Sixth International Workshop on Software Quality and Maintainability (SQM 2012), 2012.
- [SE12] Schwittek, W.; Eicker, S.: Decision Support for Off-the-Shelf Software Selection in Web Development Projects. In (Grossniklaus, M. Ed.): Current trends in web engineering. Springer, Berlin, 2012; pp. 238–243.
- [SE13] Schwittek, W.; Eicker, S.: A study on third party component reuse in Java enterprise open source software. In (Kruchten, P.; Giannakopoulou, D.; Tivoli, M. Eds.): Proceedings of the 16th International ACM Sigsoft symposium on Component-based software engineering. ACM, New York, NY, 2013; pp. 75–80.