Weaving a Semantic Web Across OSS Repositories: Unleashing a New Potential for Academia and Practice

Olivier Berger, Institut Telecom, France
Valentin Vlasceanu, Institut Telecom, France
Christian Bac, Institut Telecom, France
Quang Vu Dang, Institut Telecom, France
Stéphane Lauriere, Mandriva, France

ABSTRACT

Several public repositories and archives of “facts” about libre software projects, maintained either by open source communities or by research communities, have been flourishing over the Web in recent years. These have enabled new analysis and support for new quality assurance tasks. This paper presents some complementary existing tools, projects and models proposed both by OSS actors or research initiatives that are likely to lead to useful future developments in terms of study of the FLOSS phenomenon, and also to the very practitioners in the FLOSS development projects. A goal of the research conducted within the HELIOS project is to address bugs traceability issues. In this regard, the authors investigate the potential of using Semantic Web technologies in navigating between many different bugtracker systems scattered all over the open source ecosystem. By using Semantic Web techniques, it is possible to interconnect the databases containing data about open-source software projects development, which enables OSS partakers to identify resources, annotate them, and further interlink those using dedicated properties and collectively designing a distributed semantic graph.

Keywords: Bugtracker, Interoperability, Ontology, OSLC-CM, RDF, Repository of Repositories, Semantic Web

INTRODUCTION

The HELIOS project\(^1\) is a joint project between french academics and industrials in the frame of the Paris area System@tic cluster (under the “Libre software” thematics group), to build an open source Application Lifecycle Management (ALM) platform.

Among other goals, HELIOS aims at addressing bugtracker synchronization issues,

DOI: 10.4018/jossp.2010040103
and bug traceability. To that purpose, the potential of using Semantic Web technologies for navigating between the many similar bugs filed in the different bugtracker systems has been experimented, together with implementing REST protocols for manipulation of bug reports by different client tools.

The first section of this article introduces several open source tools illustrating the need for semantically interconnected databases and interoperability standards: bts-link, and UDD (Ultimate Debian Database) developed by the Debian community, and the Eclipse-Mylyn bugtracker client. The second section discusses new use-cases we foresee for researchers and open source practitioners, with the advent of more semantics in open source related software engineering facts repositories. It will be illustrated by recent initiatives like the LD2SD methodology from DERI, doc4 developed by Mandriva, the OSLC-CM specifications and finally our FetchBugs4.me project.

This will not constitute a detailed analysis nor the presentation of all results achieved during the HELIOS project. The objective is mainly to attract attention to novel interesting projects, and present ideas that may trigger the interest of the research community, and maybe receive useful comments on the way the work done in the frame of HELIOS can be further shaped, and on how the use of such tools by the open source communities can be maximized.

Finally, a word of caution: this paper does not enter into the details of the Semantic Web approach, that some qualify as “the next revolution of the Internet” (see instead Berners-Lee, 2001); it just focuses on the progressive adoption of Semantic Web concepts in various services and tools adopting interoperable representations of data through the use of standards such as RDF, RDFa, OWL, microformats and others. The reader unfamiliar with Semantic Web concepts and techniques is advised to read the gentle introduction presented by Howison in 2008.

LARGE SCALE INTEROPERABILITY CHALLENGES IN FLOSS DEVELOPMENT

We’ll start by introducing the reader to some key projects that have been developed by the FLOSS communities recently, that will illustrate the need for more interconnected databases, and more interoperability between tools, that may be addressed with the help of Semantic Web techniques.

NAVIGATING THE NET OF BUGS: BTS-LINK, DEBIAN’S BUG LINKS WATCHER

Open-source GNU/Linux distributions such as Debian or Mandriva are composed of thousands of assembled packages (downstream) providing software which has been developed within hundreds of independent projects (upstream).

Each GNU/Linux distribution usually maintains a central bugtracker (for instance, Debian’s debbugs which is available at http://bugs.debian.org/, or Mandriva’s bugzilla running at https://qa.mandriva.com/) that is open to reports from its users in case of problems or requests for changes. Such bugtrackers are key to the quality assurance process of the distributions. Unlike forums or mailing-lists, they often are the only place where thousands of users and maintainers can coordinate on the technical problems in a semi-structured way (thanks to the workflows of bug reports imposed by the bugtrackers). Apart from these popular bugtrackers for the distributions (running bugzilla, debbugs or launchpad), in turn, each independent FLOSS project generally maintains its own dedicated bugtracker which is mainly used by its developers and a few “power users”. They may then be running their own instance of Bugzilla, Mantis, Trac, Jira, Roundup or other such tools, or also using co-hosted track-
Factors Affecting the Development of Absorptive Capacity in the Adoption of Open Source Software
www.igi-global.com/article/factors-affecting-development-absorptive-capacity/54244?camid=4v1a