Accepted Manuscript

Trade-off between interoperability and data collection performance when designing an architecture for learning analytics

Juan Manuel Dodero, Enrique Juan González-Conejero, Guillermo Gutiérrez-Herrera, Sonia Peinado, José Tomás Tocino, Iván Ruiz-Rube

PII: S0167-739X(16)30281-3
DOI: http://dx.doi.org/10.1016/j.future.2016.06.040
Reference: FUTURE 3144

To appear in: *Future Generation Computer Systems*

Received date: 1 December 2015
Revised date: 20 May 2016
Accepted date: 27 June 2016

Trade-off between interoperability and data collection performance when designing an architecture for learning analytics

Juan Manuel Dodero, Enrique Juan González-Conejero, Guillermo Gutiérrez-Herrera, Sonia Peinado, José Tomás Tocino, Iván Ruiz-Rube

Superior School of Engineering, University of Cadiz, Spain

Abstract

The heterogeneity of external systems that can be connected in an e-learning environment can impose interoperability and performance requirements for recording and storing the learning data. Web-based protocols have been developed to improve e-learning systems' interoperability and capability to perform meaningful analytics. The present paper describes a web-based learning environment aimed at training how to command and control unmanned autonomous vehicles, provided with analytic capabilities. It integrates an external web content management system and a simulation engine that present different performance requirements for recording all significant events that occur during the learning process. Its record store construction, based on standard interoperability protocols, is explored here from the performance viewpoint. The tests that were conducted to assess regular data stores used for learning analytics show that performance should not be overlooked when constructing and deploying learning analytics systems.

Keywords:
Technology-Enhanced Learning (TEL), Virtual Learning Environment (VLE), Learning Analytics (LA), Experience API (xAPI), Learning Record Store (LRS),

1. Introduction

Modern e-learning software systems enable instructors to design and host the learning resources that must be completed by the learners, as well as reusing them as learning assignments to teach others. In addition, the analysis of interactions with learning resources enables to assess whether learners fulfill the learning objectives of a course in order to redesign it —for example, to find out whether estimated lesson durations are appropriate.

Learning Analytics (LA) has emerged in the e-learning arena as a discipline that aims at processing and analyzing data streams generated from online learning experiences, for the purposes of understanding and optimizing the learning and the environments in which it occurs [1]. User participation in a Virtual Learning Environment (VLE) or Learning Management System (LMS) leaves a trail of data that can be analyzed for insights. Collecting such data is of particular interest for measuring, personalizing, and assessing the learning experiences. Data-driven approaches are paramount for LA information systems [2]. Two types of non-functional requirements have to be satisfied, i.e., interoperability and performance, which are especially critical for all kinds of real-time and embedded systems. Non-functional requirements are often interconnected and their satisfaction cannot be considered in isolation, as they can have direct impacts on each other [3]. The ability to balance a robust design with good performance is of strategic importance. Software systems have to be secure and usable, as well as extensible, interoperable and scalable. Security and usability are types of *execution* features observable at run time, while interoperability and scalability are *evolution* features embodied in the architecture of the software system and depend on a robust design. A careful trade-off analysis of non-functional requirements is necessary for many types of real time systems, including embedded systems [4] and big data analytics [5] among others. Such information systems that do real-time analytics have to examine large data sets to uncover patterns, correlations and other useful information to understand and optimize data-driven processes with a diversity of purposes for business, research and education. LA is a clear field for observing the trade-off between the robustness of a system design (i.e. its evolution features) and the performance of its implementation (i.e. its execution features).

On the one hand, the real-time implementation of an LA system requires having sufficient technical resources to pay the computational cost of associating the relevant analysis techniques with the software services and data storage in either remote or local servers [6]. Learning data collection involves a wide range of computing issues that are common to big data research [7] and can affect all phases of the LA process cycle, including capture, report, predict, act and refine [8]. On the other hand, managing diverse data sources and formats to collect and organize LA data is a challenge for robust designs. First, data

Email addresses: juanma.dodero@uca.es (Juan Manuel Dodero), juanenrique.gonzalez@uca.es (Enrique Juan González-Conejero), guillermo.gutierrez@uca.es (Guillermo Gutiérrez-Herrera), sonia.peinado@uca.es (Sonia Peinado), josetomas.tocino@uca.es (José Tomás Tocino), ivan.ruiz@uca.es (Iván Ruiz-Rube)