Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

Preliminary Requirements Analysis towards an Integrated Learning Analytics System

Byung-gi CHOI[a], Yong-sang CHO[b], Jaeho LEE[c*]
[a]*School of Electrical and Computer Engineering, University of Seoul, Republic of Korea*
[b]*Korea Education & Research Information Service*
[c]*School of Electrical and Computer Engineering, University of Seoul, Republic of Korea*
**jaeho@uos.ac.kr*

Abstract: An integration of various information and processes for learning analytics into a united framework is the key to the development of an open and extensible learning analytics system. Recently, we have taken a step towards developing such a framework by starting to build a reference software architecture which in turn will allow us to identify the structure and the workflow of learning analytics systems. Our final goal is to develop an explicit specification of the learning analytics architecture as the international standard so that open and extensible learning analytics systems can be built for worldwide interoperability. In this paper, we present the result of our preliminary requirement analysis towards such an open and interoperable learning analytics system. The analysis focuses mainly on the system aspects of existing well-known frameworks such as IMS Global learning analytics platform.

Keywords: Learning Analytics, Requirement Analysis, Reference Architecture, Standards

1. Overview

Learning analytics (LA) systems require integration of the processes of measurement, collection, analysis and reporting of data about learners and their contexts, and thus involve multi-disciplinary areas including artificial intelligence, information science, statistics, visualization, and so on.

An effective integration of various information and processes for learning analytics into a united framework is the key to the development of an open and extensible learning analytics system because the system should be *open* to several related areas of research such as academic analytics, action research, educational data mining, personalized adaptive learning, and more. The system also should be *extensible* as new methodologies and technologies emerge rapidly. Especially, the big data technology has been evolved markedly to make it possible to collect data massively, analysis instantly, and visualize appropriately for learning analytics field.

Recently, we have taken a step towards developing such a framework by starting to build a reference software architecture which in turn will allow us to identify the structure and the workflow of learning analytics systems.

Our final goal is to develop an explicit specification of the learning analytics architecture as the international standard so that open and extensible learning analytics systems can be built for worldwide interoperability. In this paper, we present the result of our preliminary requirements analysis towards such an open and interoperable learning analytics system. The analysis focuses mainly on the system aspects of existing well-known frameworks such as IMS Global learning analytics platform.

The remainder of this paper is organized as follows. Firstly, we briefly review the purpose of learning analytics and survey related works and standardization activities in learning analytics field. Based on this survey, we then discuss the basic requirements of the reference software architecture. Finally we gives a summary of the main results of this paper and highlight directions for future work.

2. Backgrounds

In this section, we introduce the basic backgrounds that are needed for requirement analysis for learning analytics including the analysis levels of learning analytics and standardization activities.

2.1 Learning Analytics and Big Data

Learning analytics, as stated in the "Call for Papers of the 1st International Conference on Learning Analytics & Knowledge (LAK 2011)", is "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs." The tasks of measurement, collection, analysis, and reporting in this definition correspond closely to the major activities in big data, that is, collection, processing, analysis, and visualization of data, as shown in Figure 1.

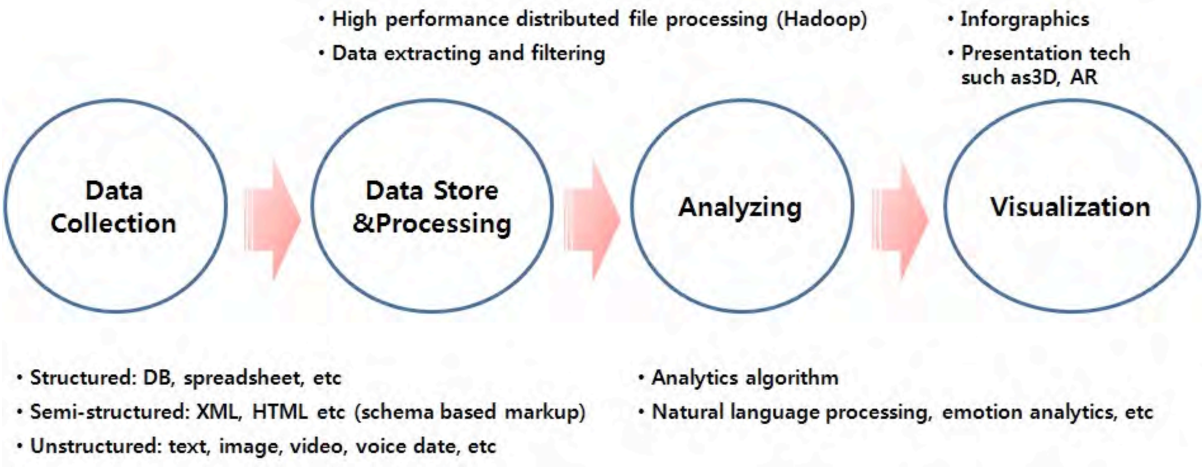

Figure 1. Big data workflow

Such correspondence is not coincidental but suggests that learning analytics can take advantage of the technological advancement of big data in building a learning analytics framework. Our requirement analysis for learning analytics is also largely borrowed from that of the big data framework.

2.2 Range of Learning Analytics

According to Buckingham Shum (2012), the range of learning analytics can be defined as macro-, meso- and micro-levels.

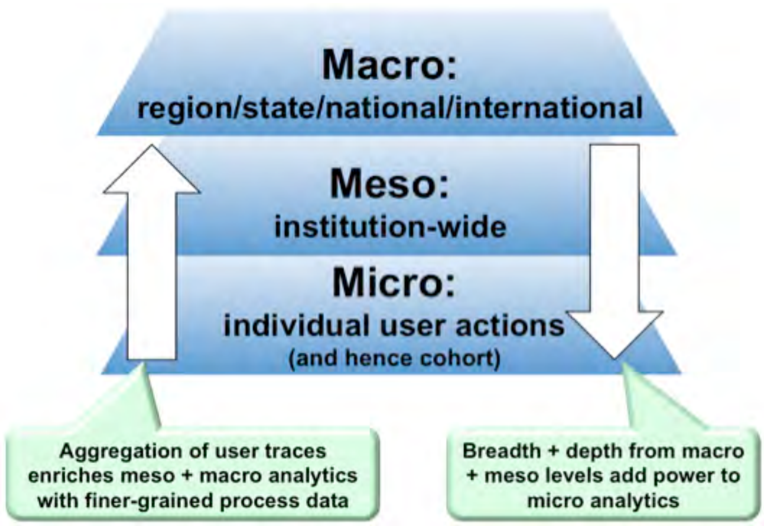

Figure 2. Levels of Learning Analytics (Buckingham Shum, 2012)

Buckingham Shum (2012) describes the levels of learning analytics as follows.

- Macro-level analytics is the cross-institutional analytics over region, state, national, or international institutions for students' lifetimes. Macro-analytics becomes increasingly real-time, incorporating more data from lower meso- or micro-levels, utilizing data integration methodologies that are developed in non-educational sections.
- Meso-level analytics operates at institution level, benefiting from the common business processes for business intelligence, utilizing the tools to integrate data silos in enterprise warehouse, optimize workflows, generate dashboards, mine unstructured data, predict future trends, and so forth.
- Micro-level analytics supports the tracking and interpretation of process-level data for individual learners or groups. This data is of primary interest to learners themselves and correspondingly the most personal, since it can disclose online activities as well as physical activities such as geolocation, library loans, purchases, and interpersonal data such as social networks.

As shown in Figure 2, while the aggregation of user date from the micro-levels enriches meso- or macro-level analytics, the breadth and depth at the macro- or meso-levels add power to micro-level analytics in building predictive models or providing feedback to learners. We thus believe that an effective integration of data and activities among these layers is essential requirement for mutual enrichment.

2.3 Reference Model for Learning Analytics

Chatti, Dyckhoff, Schroeder, and Thüs (2012) describe a reference model for learning analytics based on four dimensions: data and environments (what?), stakeholders (who?), objectives (why?), and methods (how?) as depicted in Figure 3.

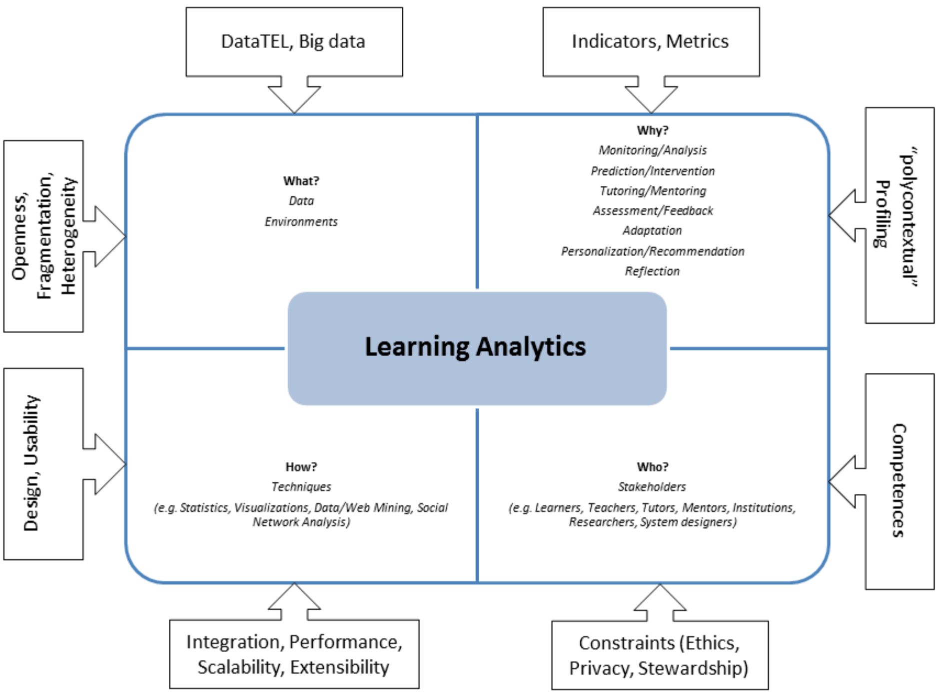

Figure 3. Learning Analytics Reference Model (Chatti et al., 2012)

This reference model is going to be utilized in our requirement analysis as it provides a classification schema on software components in our ongoing development of reference software architecture.

2.4 Standardization Activities

As the importance of the interoperability of learning analytics is recognized in Korea, a three-year-long research project funded by Telecommunications Technology Association (TTA) of Korea is recently launched to build *a reference model of learning analytics based on educational content and unstructured data*. In this project, Korea Education and Research Information Service (KERIS) runs the project in close coordination with ISO/IEC JTC 1/SC 36 and IMS Global.

ISO/IEC JTC 1/SC 36, Information Technology for Learning, Education and Training is a standardization subcommittee (SC), which is part of the Joint Technical Committee ISO/IEC JTC 1, that develops and facilitates standards within the field of information technology (IT) for learning, education and training (LET) to support individuals, groups, or organizations, and to enable interoperability and reusability of resources and tools. Recently, SC 36 established an *Ad Hoc Group on Learning Analytics Interoperability* to ascertain the necessity of standards to facilitate interoperability among diverse learning analytics components such as data collection, analysis, privacy protection, qualification of data and accessibility. We expect to submit the result of requirement analysis to this Ad Hoc Group as contributions for assessment, and anticipate to participate in the future Working Group, if established, with the reference software architecture.

3. Related Work

3.1 IMS Global

IMS Global Learning Consortium (usually referred to as IMS GLC, IMS Global or simply IMS) is a global, nonprofit, member organization that strives to enable the growth and impact of learning technology in the education and corporate learning sectors worldwide. IMS GLC members provide leadership in shaping and growing the learning industry through community development of interoperability and adoption practice standards and recognition of the return on investment from learning and educational technology. Their main activity is to develop interoperability standards and adoption practice standards for distributed learning, some of which like Learning Tools Interoperability (LTI), Question & Test Interoperability/Accessible Portable Item Protocol (QTI/APIP), Common Cartridge, Learning Information Services and Content Packaging are very widely used (Wikipedia, 2014a).

Especially, IMS Caliper (IMS Global Learning Consortium, 2013) is a work in progress to define a learning measurement framework using existing IMS specifications to provide a standardized representation, capture, and marshaling of learning activity generated metrics targeted for consumption by any conforming sensor API endpoint enabled analytics store/ services. IMS Caliper is built around the following concepts:

- **IMS Learning Metric Profiles** that provide a Learning Activity centric focus to standardize on metrics (actions and related context) captured across consumer and producer learning tool's delivery activities and delivery platforms that consume and orchestrate activity based curriculum, while providing for custom extensions and future additions to the profiles;
- **IMS Learning Sensor API and Learning Events** drive standardized instrumentation and metric capture and marshal between tools and their delivery platforms and/or associated analytics service solution aggregating metrics;
- **IMS LTI™/LIS/QTI™ leverage and extensions** enhance and integrate granular, standardized learning measurement with tools interoperability and the underlying learning information models, inclusive of course, learner, outcomes and other critical associated context.

In a recent IMS publication, "Learning Measurement for Analytics Whitepaper" (IMS Global Learning Consortium, 2013), they claim standards for learning analytics are required so they can be combined across all of the educational sources by asserting that *"equipped with a standards based common foundation for learning measurement, the quality, efficacy and performance derived analytics for the online curriculum across the ecosystem can be achieved more effectively."*

3.2 UNESCO IITE

UNESCO Institute for Information Technologies in Education (IITE) identifies three kinds of predictors and indicators, and two kinds of interventions as follows.

- **LMS/VLE Analytics Dashboards**: The first kinds of analytics that many institutions will encounter will be the analytics dashboards now appearing in most online learning platforms., Data logs are now rendered via a range of graphs, tables and other visualizations, and custom reports designed for consumption by learners, educators, administrators and data analysts.
- **Predictive Analytics**: From the pattern of learners' static data (e.g. demographics; past attainment) and dynamic data (e.g. pattern of online logins; quantity of discussion posts) one can classify the trajectory that they are on (e.g. "at risk"; "high achiever"; "social learner"), and hence make more timely interventions (e.g. offer extra social and academic support; present more challenging tasks).
- **Adaptive Learning Analytics**: Adaptive learning platforms build a model of a learner's understanding of a specific topic (e.g. algebra; photosynthesis; dental surgical procedures), sometimes in the context of standardized tests which dictate the curriculum and modes of testing. This enables fine-grained feedback (e.g. which concepts you have grasped and at what level), and adaptive presentation of content (e.g. not showing material that depends on having mastered concepts the learner has failed on).
- **Social Network Analytics**: Social network analysis (sometimes called Organizational Network Analysis in corporate settings) makes visible the structures and dynamics of interpersonal networks, to understand how people develop and maintain these relations. People may form 'ties' of different sorts, ranging from extended, direct interaction reflecting significant ties, to more indirect ties.
- **Discourse Analytics**: Analytics could go beyond simple quantitative logs, and provide feedback to educators and learners on the quality of the contributions. Researchers are beginning to draw on extensive prior work on how tutors mark essays and discussion posts, how spoken and written dialogue shape learning, and how computers can recognize good argumentation, in order to design analytics that can assess the quality of text, with the ultimate goal of scaffolding the higher order thinking and writing that we seek to instill in students.

4. Requirements of the Reference Software Architecture

In this section, we present preliminary results of requirement analysis of the reference software architecture based on the survey described in the previous sections, Section 2 and 3. Even though, the requirements are multifold over data, analysis, and application requirement, the results can be summarized as design requirements of reference software architecture as follows:

- **Open and extensible**: It should be open to incorporate new sensors or analytics functionality, desirably without interrupting the task being serviced. It also should ensure incorporation or modification of new workflows at the task level.
- **Distributed**: It should be able to handle multiple sources of data and functionalities distributed over multiple systems. It is also desired to be able to distribute data and to delegate functionality dynamically and transparently.
- **Interoperable:** It should provide compatibility for various learning platforms or VLE by providing interoperable interface to the data and operations.
- **Reusable and configurable**: The functional components and data interfaces should be modular and thus reused and configured for different tasks or more complex tasks as building blocks.
- **Real-time and predictable**: Learning analytics should be performed satisfying the real-time constraints and should be able to estimate the time to completion.
- **Usable:** It should acceptable user experience (UX) by providing appropriate data visualization and user interfaces for monitoring and tasking throughout the learning analytics process.
- **Secure and traceable:** It should protect personal user information to secure privacy and preserve confidential information. Some analytics functionality should be ensured not to be performed as

required. Furthermore, the history of execution of analytics functions and access to data should be recorded, if needed, to ensure traceability.

5. Reference Software Architecture

A reference software architecture is a software architecture where the structures and respective elements and relations provide templates for concrete architectures in a particular domain or in a family of software systems. A reference architecture often consists of a list of functions and some indication of their interfaces (or APIs) and interactions with each other and with functions located outside of the scope of the reference architecture (Wikipedia, 2014b).

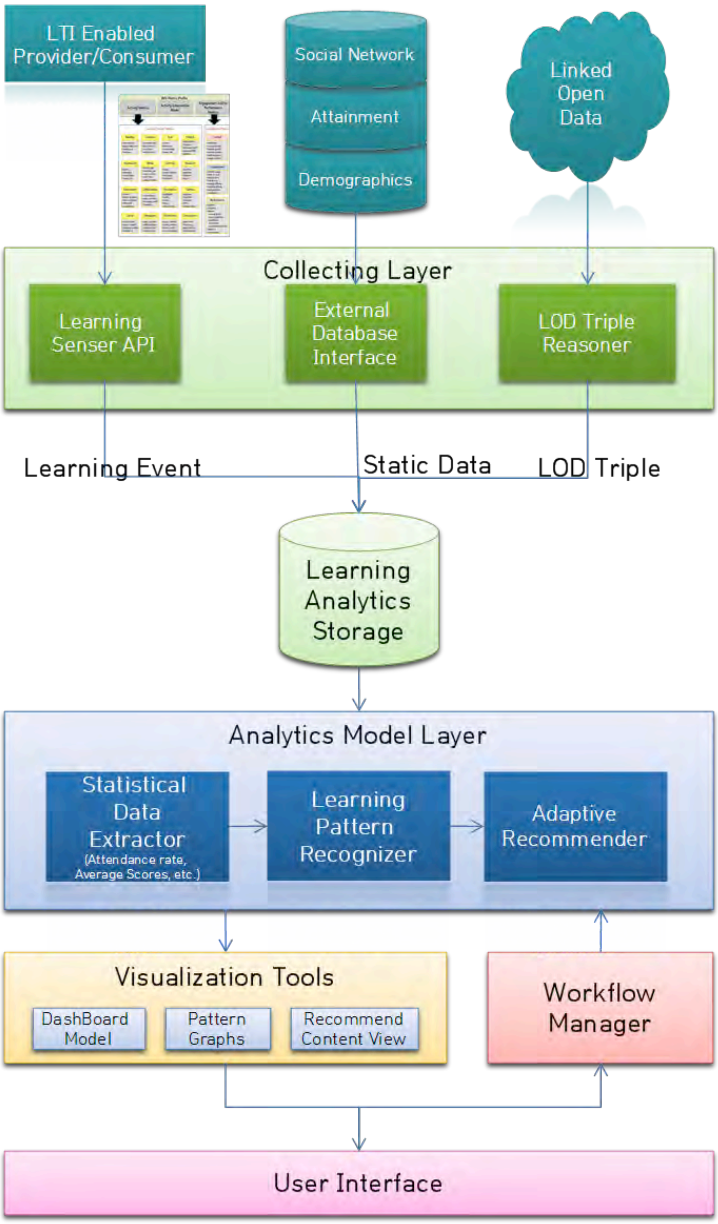

Figure 4. a preliminary reference software architecture

In this section, we present our preliminary reference software architecture as shown in Figure 4. The main purpose of this initial work is to identify the necessary components and data interactions of the components before the results of the requirement analysis are fully applied. The initial architecture is then instantiated with supporting software tools to validate the effectiveness. Table 1 summarizes the basic input and output interfaces for each component and applicable software to implement the functionality of the component.

Table 1: Specifications of the Reference Software Architecture

Component	Description	Input	Output	Applicable Software
Learning Sensor API	External API to Collecting Learning Activity Data	Metric Profile	Learning Event Data	Apache Storm
External Database Interface	Collect Structured Data from RDB or Web Services	Well-Formatted Data		Apache Sqoop
LOD Triple Reasoner	Collect LOD and extension Triple Data	LOD Triple	LOD Triple	Webpie
Statistical Data Extractor	Calculate Statistical Information	Literal data set	Analyzed Information	*R*
Learning Pattern Recognizer	Recognize Specific pattern by Data Mining	Literal Data Set, Analyzed Information		*R*
Adaptive Recommender	Recommend Associated Content	Analyzed Information	Literal Data	*R*
Visualization Tools	Visualize each Types of Analyzed Data	Analyzed Information	HTML	Google Charts

6. Discussion and Future Work

Realizing learning analytics systems indispensably requires an integrated and holistic approach. In this paper, we first identified the basic requirements of reference software architecture to capture the holistic and integrated view. The primary requirements includes the system to be open, distributed, interoperable, reusable, real-time, usable, and secure. So far these requirements were derived from the survey of the state of the art in learning analytics field, but they need to be refined and applied to the reference software architecture by the real use cases and application scenario. We are currently collecting such use cases and application scenarios of learning analytics. We are well aware that our goal to develop the explicit specifications of the learning analytics architecture is far away from this preliminary work, but we believe this is an essential step toward the goal.

Acknowledgements

This research was supported by the ICT Standardization program of MISP (The Ministry of Science, ICT & Future Planning)

References

Buckingham Shum, S. (2012). Learning analytics. *IITE Policy Brief*.
Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5), 318-331.
IMS Global Learning Consortium. (2013). Learning Measurement for Analytics Whitepaper.
Wikipedia. (2014a). IMS Global. 2014, from http://en.wikipedia.org/wiki/IMS_Global
Wikipedia. (2014b). Reference Architecture. 2014, from http://en.wikipedia.org/wiki/Reference_architecture