Ben De Meester, Sven Lieber, Anastasia Dimou, and Ruben Verborgh

Ghent University – imec – IDLab, Department of Electronics and Information Systems, Ghent, Belgium

{firstname.lastname}@ugent.be

Interoperable User Tracking Logs using Linked Data for improved Learning Analytics

Bio data

Ben De Meester joined the IDlab research group of Ghent University – imec, Belgium as a full-time researcher immediately after he graduated from Ghent University in 2013 as a Master of Computer Science. His research interests comprise digital publishing and education, and declarative solutions for generation and transformation of Linked Data.

Sven Lieber received his B.Eng. in Communication and Software Engineering from Albstadt-Sigmaringen University of Applied Sciences, Germany (2013) and his M.Sc. degree in Computer Science from the University of Freiburg, Germany (2016). After graduation he joined IDLab of Ghent University – imec, Belgium, where he currently pursues his PhD. His research interests cover Provenance, Privacy, and Semantic Web.

Anastasia Dimou is a post-doctoral researcher at the IDLab research group of imec/UGent, Ghent, Belgium since February 2013. Research interests include Linked Data generation and publication, data quality and integration, knowledge representation and management. Her research activities led to the development of the RML tool chain, and she currently conducts research on automated Linked Data generation, validation and assessment, and Knowledge integration from Big stream data.

Ruben Verborgh is a professor of Semantic Web technology at IDLab, a group of Ghent University/imec, and a research affiliate at the Decentralized Information Group of CSAIL at MIT. His research interests comprise the Web, decentralization, hypermedia clients, Web APIs, Linked Data, and much more.

Abstract

Learning analytics can provide adaptive learning and performance support by analyzing user tracking logs. However, data-driven learning is usually confined to a specific context (e.g., learning English within one application), and thus not interoperable across systems or domains. In this paper, we investigate ways to improve integration of data across applications and educational domains, by means of Linked Data, using existing standards such as the Experience API. Using JSON-LD, existing Learning Record Store tools can be used to store the tracking logs, which are then interpreted and aligned as Linked Data. We have applied the solution in an initial data capture resulting in more than two million statements span-

ning two different applications. This way, we aim to enrich adaptive learning and performance support across contexts.

Paper

Computer Assisted Language Learning (CALL) and other computer assisted learning systems enable automated tracking of *Information about the Learning Process* (i.e., learning analytics). Integration of learning analytics with *Information about the Learner* enriches data-driven learning. Learning analytics using user behavior and user's background variables can improve adaptive computer assisted learning systems. The initial skill level of a new user can be predicted based on its background variables (e.g., age, location), and this level can be adjusted based on the user's behavior (e.g., human-computer interactions) throughout the learning trajectory. This improves learner engagement and lowers dropouts.

However, user tracking data usually resides in a closed context, i.e., within a specific application and confined to a single educational domain (e.g., learning English). By opening this data, i.e., integrating it across contexts, more learning analytics are available, and data-driven learning can be enriched.

Within the LEAPS project (LEarning analytics for AdaPtive Support, 2016-2018)[1], we tackle the challenge of combining learning analytics across contexts and domains, and investigate ways to improve interoperability between computer assisted learning systems. This way, we aim to enrich adaptive learning systems. For example, within LEAPS, a first application provides auditory discrimination assessments, and a second provides relevant suggestions in writing exercises. The applications adapt the ability level based on the tracked behavior and the background variables of the user, based on the Elo-rating algorithm (Elo, 2008). By combining learning analytics across these applications, both adaptive algorithms can be improved even further. For example, when the first application tracks that a user needs improvement for a specific auditory discrimination activity (e.g., distinguishing between "i" and "a"), related suggestions can be enabled within the second educational application. Therefore, interoperable data across applications is needed.

By means of Linked Data, we capture learning analytics in a common data model – using the Resource Description Framework (RDF) (Cyganiak, Wood, & Lanthaler, 2014) – and thus enable interoperable data across applications. The main principles of Linked Data are: (i) use URIs as names for things; (ii) use HTTP URIs so that people can look up those names; (iii) when someone looks up a URI, provide useful information, using the standards; and (iv) include links to other URIs so that they can discover more things (Berners-Lee, Hendler, Lassila, & others, 2001). However, instead of requiring existing applications to adhere to these principles, our approach makes use of existing industry standards, and Linked Data is generated automatically. This eases the integration with existing industry workflows.

Our approach comprises two steps. The first step tracks user behavior using generic object models, interfaces, and standards. We make use of the Experience API (xAPI)[2]. xAPI is a technical specification that allows applications to dynamically track, store and share data about learners in their context building on a standardized tracking vocabulary and APIs for learning applications and reporting tools to communicate and exchange data. Thus, adopters can make use of existing specifications and implementations to already track user data within their own context. More detailed context information is stored using the Semantic ExeRcise Interchange Format (SERIF) (De Meester, et al., 2015). This is a format for describing the content of educational assessments, based on the IMS QTI model from the IMS Global Learning Consortium[3]. These xAPI statements can be logged using any compli-

[1] https://www.imec-int.com/en/what-we-offer/research-portfolio/leaps_2
[2] https://xapi.com/
[3] http://www.imsglobal.org/question/

ant Learning Record Store. We make use of LearningLocker[4] given its popularity and open license. The second step generates Linked Data out of the user tracking logs. As xAPI statements as logged in the JSON format, we can make use of JSON-LD to interpret the existing structure as RDF (Sporny, Kellogg, & Lanthaler, 2014). To define the common semantic meaning of these statements, we make use of an xAPI ontology and a SERIF ontology, building upon previous research, namely, TinCan2PROV (De Nies, Salliau, Verborgh, Mannens, & Van de Walle, 2015)[5]. These ontologies define a common semantic meaning of these statements.

As this approach provides a common model across systems, the learning analytics can easily be integrated, on the one hand, with *Information about the Learner*, and on the other hand, with learning analytics across contexts, thus paving the way for transdisciplinary learning analytics. For example, we can then differentiate between learners that have difficulties learning foreign languages in general, or have troubles with one specific language. To integrate learning analytics across contexts, on the one hand, users need to be uniquely identified across the applications. When a user has different accounts for different applications, a unification of those accounts is needed to combine the learning analytics per user. This can be done using, e.g., another data source that unifies the different user accounts. As we make use of Linked Data, this process is made easier by design, as Linked Data improves integration of multiple data sources. On the other hand, the abilities of the users also need to be aligned across contexts. In this case, an exact match between abilities is not always possible. For example, distinguishing between the "i" and "a" sounds can be related, but not equal to being able to correctly write words containing "i" and "a". For this, semantic models exist – such as the Simple Knowledge Organization System (Miles & Bechhofer, 2009) – that can relate concepts without unifying them. Based on their relatedness, learning analytics can be more or less combined.

Our approach is implemented in two educational applications and tested in multiple schools, resulting in more than two million statements that track user behavior. This initial data gathering is currently used to improve the existing adaptive learning algorithms, and allow an initial evaluation of combining learning analytics across contexts. Future work includes tracking a third educational application and combining transdisciplinary learning analytics. We foresee that this integration can improve personalization (e.g., adaptive learning) and performance support.

References

Berners-Lee, T., Hendler, J., Lassila, O., & others. (2001, 5). The Semantic Web. *Scientific American, 284*, 28-37.

Cyganiak, R., Wood, D., & Lanthaler, M. (2014, 2 25). *RDF 1.1: Concepts and Abstract Syntax.* Tech. rep., World Wide Web Consortium (W3C). Retrieved from http://www.w3.org/TR/rdf11-concepts/

De Meester, B., Ghaem Sigarchian, H., De Nies, T., Verborgh, R., Salliau, F., Mannens, E., & Van de Walle, R. (2015, 10). SERIF: A Semantic ExeRcise Interchange Format. In S. Dietze, H. Sack, E. Simperl, M.-E. Vidal, & S. Peroni (Ed.), *LINKED2015 at ISWC2015.*

De Nies, T., Salliau, F., Verborgh, R., Mannens, E., & Van de Walle, R. (2015, 5). TinCan2PROV: Exposing Interoperable Provenance of Learning Processes through Experience API Logs. *Proceedings of the 24th International Conference on World Wide Web - WWW' 15 Companion* (pp. 689-694). ACM Press.

Elo, A. E. (2008, 5). *The Rating of Chess Players, Past and Present.* Ishi Press.

[4] https://learninglocker.net/

[5] xAPI used to be named the TinCan API.

Miles, A., & Bechhofer, S. (2009, 8). *SKOS Simple Knowledge Organization System Reference.* W3C Recommendation, W3C. Retrieved from https://www.w3.org/TR/skos-reference/

Sporny, M., Kellogg, G., & Lanthaler, M. (2014, 1). *JSON-LD 1.0.* Tech. rep., World Wide Web Consortium (W3C). Retrieved from http://www.w3.org/TR/json-ld/