

Plataforma de Serviços Inteligentes de Learning Analytics para grandes quantidades de dados

Thomas Rabelo¹, Manuel Lama², Ricardo Amorim^{1,3}, Juan C. Vidal²

¹Faculdade de Ciências Aplicadas e Sociais de Petrolina (FACAPE) -
Petrolina, PE – Brasil

²Centro de Investigación en Tecnoloxías da Información (CITIUS) -
Santiago de Compostela, Espanha

³Departamento de Educação
Universidade do Estado da Bahia (UNEB) – Senhor do Bonfim, BA– Brasil

thomas.rabelo@facape.br, manuel.lama@usc.es, amorim.ricardo@gmail.com

Abstract. *This article presents SmartLAK an architecture for large data analysis generated by the students during a course through Smart Services. Developed from Big Data technologies, ontologies and architectural patterns, which were combined to facilitate the development of a flexible and interoperable architecture and as a way to support intelligent services oriented to Learning Analytics. Currently, this platform is being used experimentally in a Department of Education with good results already signposted from 250,000 declarative analysis on a course in blended learning mode with 60 subscribers.*

Resumo. *Este artigo apresenta SmartLAK, uma arquitetura para a análise de grande quantidade de dados gerados pelos aprendizes durante um curso por intermédio de serviços inteligentes. Desenvolvida a partir de tecnologias de Big Data, ontologias e padrões arquiteturais, as quais foram combinadas para facilitar o desenvolvimento de uma arquitetura flexível e interoperável e como forma de suporte aos serviços inteligentes orientados à Learning Analytics. Atualmente, esta plataforma está sendo utilizada de forma experimental em um departamento de Educação com bons resultados já sinalizados a partir da análise de 250.000 declarativas em um curso na modalidade blended learning com 60 inscritos.*

1. Introdução

Os Ambientes Virtuais de Aprendizagem (AVA), plataformas software destinadas a promover a aprendizagem do aluno, online, ou "a qualquer hora e lugar", tem sido objeto de estudo nas últimas décadas. Neste sentido, surgiram inúmeras propostas relacionadas com o estudo ou desenvolvimento de plataformas AVA com o objetivo de facilitar a aprendizagem com base em diferentes estratégias tais como a aprendizagem colaborativa ou mesclada (*blended learning*), com o apoio de redes sociais e, em muitos casos, orientadas à adaptação e personalização ao perfil do estudante através de padrões comerciais, [De Souza Alencar, 2011], [Coelho, 2012], [Dotta, 2011]. No geral, tais

propostas incluem ferramentas diversas para facilitar a interação/comunicação de forma síncrona e/ou assíncrona entre os participantes de um curso; para a gestão de conteúdos de aprendizagem e para a avaliação da performance dos estudantes. No entanto, considerando que a aprendizagem é um processo extremamente complexo, a avaliação do mesmo em AVA, com qualidade, se constitui em uma tarefa bastante árdua. Neste sentido, surgiu, no início de 2010, o conceito de Learning Analytics (LA), entendido como a medição, coleta e análise de dados sobre os estudantes e seu contexto, com o objetivo de compreender e facilitar o ensino/aprendizagem em AVA [Siemens e Gasevic, 2011].

Para atingir este objetivo, é necessário ter uma arquitetura que permita, por um lado, obter, armazenar e recuperar de forma flexível e eficazmente a grande quantidade de dados gerados pelos aprendizes durante um curso e, por outro lado, fornecer serviços inteligentes que facilitem a análise do processo de ensino/aprendizagem em atividades tais como a previsão do desempenho dos estudantes e avaliação do comportamento destes ao longo de um curso (por exemplo, descoberta automática de formas de aprendizagem e/ou padrões de atividade seguidos por estudantes e rotulagem automática de texto inserido por alunos em atividades interativas ou nas memórias dos trabalhos práticos) [Vázquez-Barreiros, Mucientes e Lama, 2015], [Kinnebrew e Biswas, 2012], [Fernández-Delgado et al, 2014], [Romero et al, 2013], [Vidal et al, 2014]. Em Peña-Alya (2014) encontra-se uma visão geral do estado da arte das técnicas de análise inteligente de dados educacionais.

Embora alguns trabalhos tenham proposto estruturas conceituais nesta linha [IMS Global, 2013], [Greller e Drachsler, 2014], não encontramos, em uma revisão exaustiva da literatura atual, uma arquitetura com capacidade de efetivamente coletar o grande volume de informações gerado pelos alunos durante cursos de longa duração (ex.. um ano) e de oferecer serviços inteligentes que facilitam a compreensão por parte do professor do que está acontecendo no AVA. A construção de tal arquitetura é uma atividade complexa pois a mesma envolve uma série de (sub) sistemas independentes, autônomos, heterogêneos e com capacidade de interagir entre eles.

Assim, com base na revisão de literatura feita no contexto deste artigo e o estado da arte de Learning Analytics descrito em Peña-Alya (2014), propomos a plataforma SmartLAK, uma plataforma desenvolvida a partir da combinação de tecnologias para o gerenciamento e análise de grandes volumes de dados (*Big Data*) com ontologias e padrões para o sensoriamento/detecção de atividades estudantis utilizadas para facilitar o desenvolvimento de uma arquitetura flexível e interoperável que irá apoiar os serviços inteligentes voltados para Learning Analytics. Com SmartLAK, acreditamos que temos uma arquitetura que incorpora componentes que são, de fato, os subsistemas necessários ao suporte de um conjunto de serviços e aplicações inteligentes requeridas na criação de uma nova visão da aprendizagem. Cada subsistema tem requisitos, processos e saídas específicas. Coletam os dados por intermédio de sensores que são armazenados e processados com o intuito de obter informação útil capaz de ajudar na tomada de decisões, melhoria da qualidade de vida e estímulo a práticas sustentáveis.

Na seção 2, a arquitetura SmartLAK e seus principais componentes são descritos; a seção 3 apresenta um cenário de uso. Finalmente, a seção 4 faz uma discussão e propõe trabalhos futuros.

2. SmartLAK: estrutura e principais componentes

A figura 1 mostra os principais componentes da arquitetura SmartLAK, a qual suporta a extração e análise dos dados gerados pelos estudantes em sua interação com o AVA. A arquitetura SmartLAK foi estruturada com base em três fatores chave: uma camada de serviço padrão, para a captura e consumo de dados produzido no AVA (xAPI¹ services); uma ontologia que representa semanticamente os dados tendo em vista a interoperabilidade entre os componentes da arquitetura (xAPI ontology) e um banco de dados orientado a grafos de alta performance de forma que permita o consumo de grandes quantidades de dados (*Big Data*). Na próxima seção, descrevemos cada componente da arquitetura em detalhe.

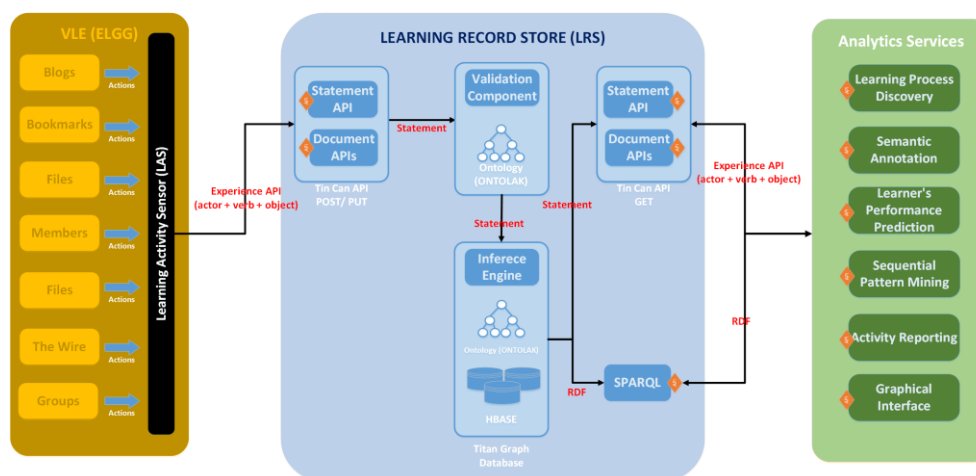


Figura 1. Arquitetura SmartLAK.

Conforme a figura, à esquerda (em Marrom) representa o ambiente virtual de aprendizagem e suas respectivas atividades, que para esse cenário consideramos o ELGG². O ELGG é um motor de rede social premiado, que possui um conjunto rico de funcionalidades que permite universidades e escolas criarem suas próprias redes sociais com: Blogs, Fóruns, Páginas, Arquivos e Favoritos que emitem ações (eventos) que são capturadas pelo sensor de atividades de aprendizagem (LAS). Esse componente é responsável por coletar os dados e enviá-los em um formato consistente (Experience API, xAPI), por meio de serviços REST (Statement API/ Document APIs); por sua vez os dados são submetidos a validação de conformidade, através de uma ontologia axiomática que representa toda a especificação Experience API, garantindo a integridade antes de salva-los no repositório de atividades de aprendizagem (LRS). O LRS é responsável por gravar os dados validados, como instancias da ontologia (OntoLAK). Sua arquitetura é composta por um Motor de Inferência, e um banco de dados orientado a grafos distribuído em um cluster Hadoop³, que o possibilita manipular e gravar grandes volumes de dados. Uma vez armazenados esses dados, é possível

¹ <http://experienceapi.com/>

² <https://elgg.org/>

³ <http://hadoop.apache.org/index.html>

acessa-los por meio de um serviço REST (xAPI), ou pelo serviço de consultas semânticas SPARQL.

À direita (Em Verde), os serviços analíticos que são componentes independentes que consomem os dados do LRS, e aplicam algoritmos de mineração de dados e técnicas de inteligência artificial para extração de informação. Por se tratar de um modelo arquitetural extensível, é possível criar e implantar serviços de maneira independente, permitindo a colaboração de comunidades interessadas no tema.

2.1. Sensores de Atividades de Aprendizagem

O propósito dos Sensores de Atividades de Aprendizagem, em inglês, Learning Activity Sensors (LAS), é capturar os eventos gerados pelos aprendizes e/ou professores em suas interações com o AVA. Neste contexto, os eventos são peças atômicas de informação descrevendo algo que os usuários do AVA fazem tais como inserir um post em um fórum, leitura de um Blog, responder a questões em um Quiz, deletar um comentário, etc. Muitos dos AVA implementam sistemas de gestão de eventos desenhados a partir de um padrão de publicação que, no geral, representam os mesmos de forma declarativa indicando apenas o que foi feito e o contexto no qual ocorrem. Com a arquitetura SmartLAK, ampliamos a forma de apresentação de eventos com a especificação Experience API, formalmente conhecida como TinCan API, a qual define (i) um modelo de dados com as principais características de um evento e (ii) uma interface que poderá ser invocada⁴ para armazenar as declarativas de eventos em um Repositório de Registro de Aprendizagem, em inglês, Learning Record Store (LRS).

Este tipo de arquitetura é necessária considerando que o rastreamento de eventos de forma contínua pode ser considerado como uma atividade do tipo stream que requer uma infraestrutura de alta performance para armazenar de forma efetiva os eventos quando o número de estudantes for muito grande (por exemplo em MOOCs). Levando isto em conta, os componentes LAS foram definidos como um barramento de eventos que centraliza as notificações de eventos em um ponto e que permite o roteamento destas como declarativas para o LRS onde o stream de atividades é armazenado.

2.2. Ontologia OntoLAK

As principais classes e relações de OntoLAK estão descritas na Figura 2. A classe Statement da ontologia proposta se refere às declarativas capturadas nesta entidade, cujo objetivo é a formulação de sentenças na forma: “Eu falei com Mike” ou “John escreveu um ensaio sobre surfe”. Statement tem três propriedades em sua estrutura: (i) Actor, propriedade que representa o sujeito da declarativa ou a pessoa/grupo que disparou o evento (“Eu” e “John” nas sentenças); (ii) Verb, para representar o predicado da declarativa (“falei com” e “escreveu”); e, (iii) Object, que indica o tipo de objeto (Activity, Agent, Group ou outro tipo de declarativa).

Conforme a Figura 2, uma declarativa também pode detalhar o resultado do evento ou as condições sobre as quais esta foi criada, por intermédio das propriedades Result e Context, respectivamente. Uma declarativa pode incluir anexos (Attachment)

⁴ Posta em execução, no contexto de chamada de eventos, em inglês, invoked.

que fazem parte do registro de aprendizagem. As quatro propriedades restantes de uma declarativa, nomeadamente uuid, timestamp, stored, and version, são propriedades de dados que descrevem a identificador UUID atribuída à declarativa, no momento que o evento ocorreu e no momento que foi armazenado no LRS e a versão da xAPI, respectivamente. Finalmente, a propriedade Authority é usada para identificar o agente ou grupo que afirma que a declarativa é verdadeira.

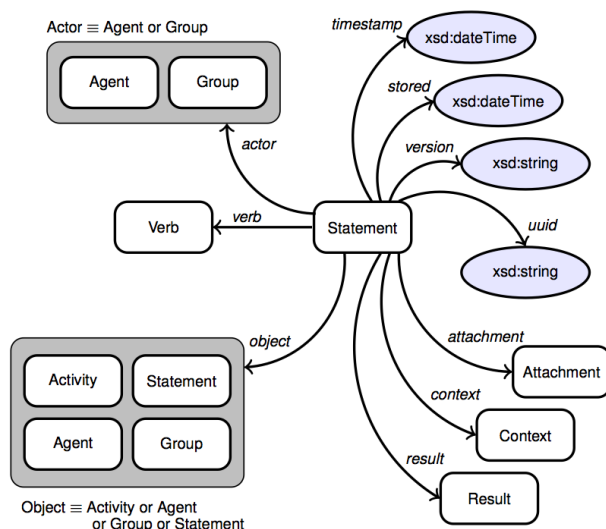


Figura 2. Modelo Statement.

Uma série de axiomas foram definidos para a OntoLAK. Estes axiomas são utilizados para verificar um dos requisitos da especificação xAPI que afirma que a declarativa deve usar cada propriedade não mais do que uma vez. No entanto, e ao contrário da especificação xAPI, os anexos são modelados como um conjunto de elementos, em vez de uma lista. Nós escolhemos esta opção porque (i) os anexos não são ordenados e podem aparecer em ordem diferente sem alterar a semântica da declaração, e (ii) ontologias proporcionar melhores mecanismos de raciocínio para conjuntos do que para listas.

Tabela 1. Axiomas do modelo Statement.

Class			
Domain	Range	Property	Axiom
Statement	Actor	actor	functional
Statement	Actor	actor	exactly 1
Statement	Verb	verb	functional
Statement	Verb	verb	exactly 1
Statement	Object	object	functional
Statement	Object	object	exactly 1
Statement	Result	result	functional
Statement	Context	context	functional
Statement	Actor	authority	functional
Statement	Attachment	attachments	functional
Statement	xsd:string	uuid	functional
Statement	xsd:dateTime	timestamp	functional
Statement	xsd:dateTime	stored	functional
Statement	xsd:string	version	functional

Os restantes três axiomas da Tabela 1 foram definidos para satisfazer a exigência de que uma declarativa deve ter um ator, verbo e objeto. Especificamente, esses axiomas representam uma restrição a cardinalidade das antigas propriedades, exigindo que não é exatamente uma instância da propriedade para cada instrução.

Tabela 2. Axiomas adicionais do modelo *Statement*.

#	Axiom
1	<p>Sub-statements must have a <i>uuid</i> but no <i>stored</i>, <i>version</i>, or <i>authority</i> properties:</p> <pre>Statement(?a), Statement(?b), object(?a, ?b) -> (authority exactly 0 foaf:Agent)(?b), (stored exactly 0 xsd:dateTime)(?b), (version exactly 0 xsd:string)(?b), (uuid exactly 1 xsd:string)(?b)</pre>
2	<p>Sub-statements cannot be nested:</p> <pre>Statement(?a), Statement(?b), object(?a, ?b), object(?b, ?c) -> (not (Statement))(?c)</pre>

Finalmente, a Tabela 2 mostra dois axiomas adicionais para verificar a semântica do conceito *Statement*. Especificamente, o axioma 1 lida com o caso de declarações que desempenham o papel de um sub-declarações. Contrariamente às declarações normais, quando uma instrução é o objeto de uma outra declaração, ele deve ter um *uuid* mas há propriedades armazenadas, versão ou de autoridade. O axioma 2 verifica se sub-declarativas não podem ser aninhadas, isto é, uma sub-declarativa não pode conter outra declarativa na parte objeto. Isto significa que as declarativas têm apenas um nível de composição.

2.3. Learning Record Store

O objetivo do LRS é duplo: por um lado, para armazenar de forma eficiente as informações referentes à aprendizagem (declarativas xAPI) sobre os eventos produzidos pelos aprendizes e professores em sua interação com o AVA; por outro lado, para permitir um alto rendimento no acesso e processamento à informação de aprendizagem a partir de técnicas de *Big Data* (uso de MapReduce, por exemplo). Para alcançar estes objetivos, foram definidos para o LRS, os seguintes componentes:

- Componente de Validação Semântica - é responsável por verificar se uma declarativa é compatível com a especificação xAPI ou não. Para isso, este componente utiliza o raciocinador Pellet [Sirin, Parsia e Grau, 2007] que, em conformidade com a ontologia, faz a validação das declarativas que vêm da camada de serviço xAPI, verificando ambos os axiomas e regras de OntoLAK. Apenas neste caso, a instrução será armazenada na base de dados LRS, garantindo que as restrições da especificação xAPI sejam cumpridas.
- O componente Titan [Aurelius, 2014] é um banco de dados orientado a grafos com o qual dados e relações podem ser representados explicitamente. O grafo no banco de dados é definido no formato Blueprints, mas como os dados armazenados nos LRS são declarações xAPI, este grafo deve ter a mesma estrutura do grafo RDF em OntoLAK: neste caso, as instâncias da ontologia são vértices do grafo e as relações de instância são as arestas deste. A estrutura de modelos de dados de Titan permite a apresentação do grafo no banco de dados na forma de tripla RDF através de uma interface Sail [Broekstra, Kampman e

Harmelen, 2002], o que significa que este grafo pode ser apresentado e consultado semanticamente por uma interface de comunicação (*Endpoint*) SPARQL. Do ponto de vista das características de *Big Data*, Titan permite distribuir grafos no banco de dados através de várias máquinas em clusters, efetuar o manuseio de complexos cruzamentos de grafos em tempo real e a análise de grandes volumes de dados através do framework Apache Hadoop. Portanto, Titan é um banco de dados de alto desempenho que facilita a integração de componentes de base ontológica para checagem de conformidade e consulta semântica.

O LRS interaciona com os componentes SmartLAK por intermédio dos serviços RESTful que implementam as interfaces em conformidade com especificação xAPI: para armazenar uma nova declarativa no banco de dados orientado a grafos, os métodos POST/PUT HTTP são invocados a partir do sensor de atividades de aprendizagem, enquanto acessa o banco de dados orientado a grafos, os métodos GET HTTP são invocados a partir dos serviços de Learning Analytics.

2.4. Serviços Inteligentes

Os Serviços Analíticos Inteligentes, em inglês, Intelligent Analytics Services (IAS), são componentes independentes que processam os dados do LRS para prover informação valiosa que permite aos professores entender melhor o que está acontecendo no curso ou para facilitar o processo de ensino e aprendizagem. Um IAS é composto de um algoritmo desenvolvido a partir de técnicas de inteligência artificial de acordo com os requisitos do serviço em si mesmo; um ponto final RESTful, que permite o serviço ser consumido por outro(s) serviço(s) IAS; um conjunto de widgets, que permitem visualizar os resultados da execução do serviço; e, um sensor de Learning Analytics que rastreia as atividades do usuário quando o IAS é integrado como parte de um painel de ferramentas de Learning Analytics.

A versão atual da arquitetura SmartLAK inclui um conjunto de serviços IAS que foram especialmente desenvolvidos para a avaliação de aprendizes:

- Serviço de descoberta em processos de Aprendizagem. Para a obtenção automática da trilha de aprendizagem seguida pelos aprendizes durante um curso. Este serviço foi implementado com o algoritmo ProdDiGen [Vázquez-Barreiros, Mucientes e Lama, 2015];
- Serviço de mineração de Padrão Sequencial. Em inglês, Sequential Pattern Mining (SPM) service, detecta padrões de atividades que os aprendizes seguem durante um curso. Estes padrões são entendidos como sequencias de atividades de aprendizagem que ocorrem com alta frequência durante o curso, em comparação com outras sequencias. Os SPM tem sido aplicados para detectar padrões de atividades em aprendizagem auto regulada;
- Serviço de Anotação Semântica. Classifica e anota o conteúdo textual (por exemplo, entradas em blogs, posts em fóruns, comentários, etc.) produzido pelos aprendizes durante um curso, facilitando o acesso de professores aos conteúdos através de tópicos ou anotações. Este serviço foi implementado com o algoritmo ADEGA [Vidal et al, 2014].

- Serviço de predição de rendimento do aprendiz. Obtêm o nível de rendimento relativo aos objetivos educacionais de um curso, usando como entrada acertos de avaliações de atividade de aprendizagem desenvolvidas pelos aprendizes. Este serviço foi implementado pelo algoritmo descrito em [Fernández-Delgado et al, 2014].
- Serviço de Relatório de Atividades, fornece relatórios gerados automaticamente em linguagem natural a partir dos dados de cada aprendiz. Este serviço é baseado em técnicas de descrição linguística adaptados obtidas a partir de ferramentas dos campos de lógica fuzzy e linguagem natural. Este serviço foi implementado com o algoritmo SLAR [Barreiros et al, 2015].
- Serviço de Interface gráfica permite ao professor entender o comportamento do aprendiz por intermédio da visualização dos dados disponíveis no LRS e os resultados obtidos na execução do IAS. Este componente não é independente de outros IASs, na medida que necessita da execução destes para provê informação aos professores. A Fig. 3 mostra uma captura de tela da interface gráfica de SmartLAK.

3. Cenário de uso

A versão atual da arquitetura SmartLAK tem sido usada para apoiar o processo de avaliação dos alunos do curso de graduação de Tecnologia Educacional da Licenciatura em Pedagogia na Faculdade de Educação da Universidade de Santiago de Compostela. Este curso tem 60 alunos inscritos. Durante o primeiro semestre de 2014, foi desenvolvido um curso na modalidade blended learning com atividades virtuais, onde os alunos realizam atividades de aprendizagem através de um e-portfolio social com blogs, ferramentas de micro-blogging, páginas de favoritos, etc. Neste cenário, a arquitetura SmartLAK capturou e armazenou mais de 250.000 declarativas, e seus serviços de Learning Analytics ajudou os professores a avaliar o desempenho dos alunos com base em seus percursos de aprendizagem, a classificação e anotação de conteúdo textual, e as estatísticas de participação suportados por um painel de análise de aprendizagem (Figura 3).

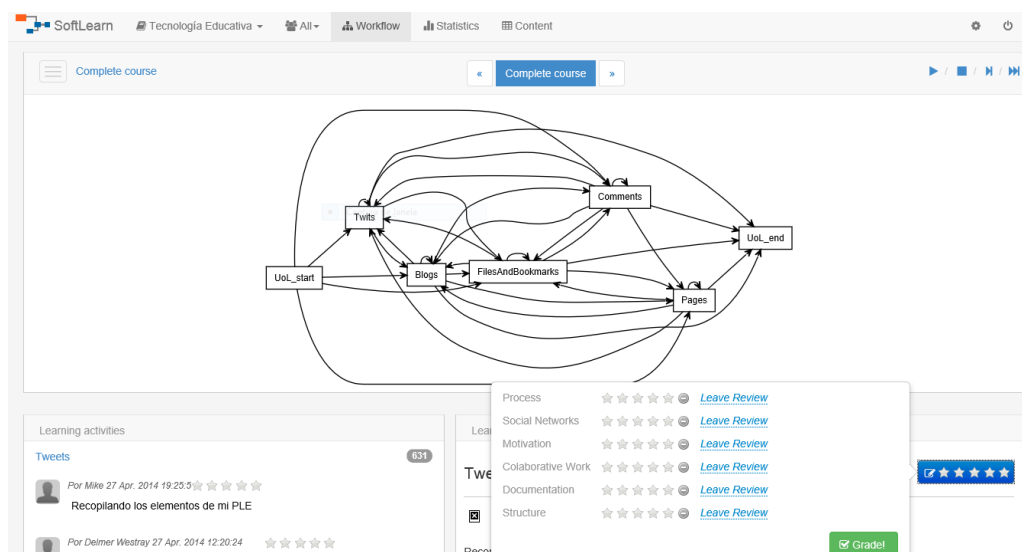


Figura 3. interface gráfica da plataforma SmartLAK.

4. Considerações finais e trabalhos futuros

Apresentamos SmartLAK, uma arquitetura baseada em *Big Data* que utiliza uma ontologia baseada na xAPI para representar o fluxo de atividades gerado pelos alunos durante o curso, favorecendo a interoperabilidade entre os componentes da arquitetura. A Ontologia é utilizada para checar a conformidade do fluxo de dados com a especificação xAPI, garantindo que os dados disponíveis no LRS serão compatíveis com especificação para armazenamento dos dados em um banco de dados de alto desempenho orientado a *Big Data*.

Como trabalhos futuros a nossa intenção é melhorar a arquitetura com um Enterprise Service Bus que permite integrar diferentes fontes de stream de dados, incluindo uma fila de mensagens orientada a *Big Data* para aumentar o desempenho do stream de atividades. Finalmente, pretendemos realizar vários experimentos com vários cursos e uma grande quantidade de alunos que iriam gerar milhões de dados.

5. Referências

- Aurelius, “Titan: Distributed graph database.” <http://thinkaurelius.github.io/titan/>, 2014.
- Barreiros, B. V., Lama, M., Mucientes, M. e Vidal, J. C. “Softlearn: A process mining platform for the discovery of learning paths,” in *IEEE 14th International Conference on Advanced Learning Technologies (ICALT 2014)*. IEEE Press, 2014, pp. 373–375.
- Barreiros, B. V., Lama, M., Mucientes, M. e Vidal, J. C. “Towards textual reporting in learning analytics dashboards,” in *IEEE 15th International Conference on Advanced Learning Technologies (ICALT 2014)*. IEEE Press, 2015.
- Broekstra, J., Kampman, A., e Van Harmelen, F., “Sesame: A generic architecture for storing and querying rdf and rdf schema,” in *The Semantic WebISWC 2002*. Springer, 2002, pp. 54–68.
- Coelho, Willyans Garcia. Uso dos recursos de mídias sociais na Educação a Distância: impactos na percepção da presença social. In: **Anais do Simpósio Brasileiro de Informática na Educação**. 2012.
- De Souza Alencar, Wanderley et al. RHODES: Um Ambiente Virtual de Aprendizagem para Programação Linear. In: **Anais do Simpósio Brasileiro de Informática na Educação**. 2011.
- Dotta, Sílvia. Uso de uma mídia social como ambiente virtual de aprendizagem. **Anais do XXII Simpósio Brasileiro de Informática na Educação**. Citado na pág, v. 1, 2011.
- Fernández-Delgado, M., Mucientes, M., Lama, M., Vázquez-Barreiros, B. (2014): Learning Analytics for the Prediction of the Educational Objectives Achievement. 44th Annual Frontiers in Education Conference (FIE 2014), pp. 2500-2503, Madrid, España.
- Greller, W., Drachsler, H. (2014): Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society* 15(3):42-57.
- IMS Global Learning Consortium (2013): Caliper Analytics. <http://www.imsglobal.org/caliper>.

- Junior, Elias Vidal Bezerra; Gomes, Alex Sandro; Souza, Flávia Veloso. Plataforma Social Educacional Redu. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. 2012.
- Kinnebrew, J.S., Biswas, G. Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution. 5th International Conference on Educational Data Mining (EDM 2012), pp. 57-64, Chania, Grecia.
- Longaray, Ariane NC; Behar, Patricia A.; Longhi, Magalí T. Afetividade em um ambiente virtual de aprendizagem: um estudo sobre os indicadores pedagógico. **Revista de Informática Aplicada**, v. 9, n. 2, 2014.
- Peña-Alaya, A. (2014): Educational Data Mining: A survey and data mining-based analysis of recent works. *Expert Systems with Applications* 41:1432-1462.
- Rayo, A., Guenaga, M. e Nunnez, A. N. “Ensuring the integrity and interoperability of educational usage and social data through Caliper framework to support competency-assessment,” in *44th Annual Frontiers in Education Conference (FIE 2014)*. IEEE Press, 2014, pp. 2275–2783.
- Ribeiro, Ana Carolina Ribeiro; Grande, Tássia Priscila Fagundes; Behar, Patricia Alejandra. Letramento digital na Ciberinfância: uma leitura a partir do ambiente virtual Planeta ROODA 2.0. In: **Anais do Simpósio Brasileiro de Informática na Educação**. 2012.
- Romero, C., López, M.I., Luna, J.M., Ventura, S. (2013): Predicting students' final performance from participation in on-line discussion forums. *Computers & Education* 68:458-472.
- Siemens, G. e Gasevic, D. (2012): Guest Editorial: Learning and Knowledge Analytics. *Educational Technology & Society* 15(3):1-2.
- Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A. e Katz, Y. “Pellet: A practical owl-dl reasoner,” *Web Semantics: science, services and agents on the World Wide Web*, vol. 5, no. 2, pp. 51–53, 2007.
- Vázquez-Barreiros, B., Mucientes, M., Lama, M. (2015): ProDiGen: Mining complete, precise and minimal structure process models with a genetic algorithm 294:315-333.
- Vidal, J.C., Lama, M., Otero-García, E., Bugarín, A. (2014): Graph-based semantic annotation for enriching educational content with linked data. *Knowledge-Based Systems* 55:29-42.