

Arquitetura de Integração e Ontologia de Suporte para Recomendação de Objetos Educacionais em Ambientes Virtuais de Aprendizagem

Jorão Gomes Jr.^{1,2}, Jairo Francisco de Souza^{1,2},
Fabrício Martins Mendonça^{1,2}, Regina Braga²

¹ Laboratório de Aplicações e Inovação em Computação (LApIC)

² Programa de Pós-Graduação em Ciência da Computação
Universidade Federal de Juiz de Fora (UFJF)
36.360-900 – Juiz de Fora – MG – Brasil

{joraojunior, jairo.souza, fabricio.mendonca} @ice.ufjf.br

regina.braga@ufjf.edu.br

Abstract. *The growing current adherence to e-learning and the more recurrent use of Virtual Learning Environments (VLE) has generated a large volume of digital educational data and, consequently, problems in the representation, organization and recovery of information in this context. As a solution in this scenario, this research presents an integration architecture that uses an ontology, called Prov4LA, to standardize the generation of learning analytics data and provide support for educational recommendation systems. As partial results, a scenario of application of the proposed architecture is described through a course taught in VLE Moodle with students who participate in the same course and where recommendation events are triggered.*

Resumo. *A crescente adesão pela educação a distância e o uso mais recorrente dos Ambientes Virtuais de Aprendizagem (AVA) tem gerado um grande volume de dados educacionais digitais e, conseqüentemente, problemas na representação, organização e recuperação de informação nesse contexto. Como uma solução, a presente pesquisa apresenta uma arquitetura de integração que utiliza uma ontologia, denominada Prov4LA, para padronizar a geração de dados de learning analytics e prover suporte a sistemas de recomendação educacionais. Como resultados parciais, é descrito um cenário de aplicação da arquitetura proposta através de um curso ministrado no AVA Moodle com alunos que participam de um mesmo curso e onde são disparados eventos de recomendação.*

1. Introdução

A crescente adesão pela educação a distância fomenta o uso de Ambientes Virtuais de Aprendizagem (AVA). Conseqüentemente, o aumento da utilização de recursos educacionais virtuais leva a uma grande produção de dados, o que gera problemas e desafios para sua representação e organização. Como uma solução alternativa para tais problemas, o presente trabalho apoia-se na utilização de ontologias e em técnicas de *learning analytics* (LA) [Chatti et al. 2013] para representação e organização do conhecimento da área educacional.

O contexto educacional envolve a abrangência de diversos conteúdos informacionais. Em geral, a preparação, seleção e divulgação desses conteúdos ficam a cargo do professor [Costa et al. 2013]. Selecionar conteúdos e adaptar-se às necessidades de cada aluno em AVAs é uma tarefa bastante onerosa e complexa para ser feita de manualmente por professores. Nesse contexto, a utilização dos metadados gerados através de técnicas de *LA* apresentam-se como uma alternativa para resolver tais problemas.

Já as ontologias podem ser vistas como modelos de anotação [Uren et al. 2006, Burger et al. 2010, da Silva Lemos and Souza 2020] em uma visão de tratamento semântico destinado a dados e metadados envolvidos no processo de representação. Como cada AVA possui uma geração de dados diferente, ainda não existem soluções que padronizem, integrem e comuniquem esses diversos tipos de dados com aplicações inteligentes, tornando-se essencial o uso de ontologia para a manutenção e disponibilização de dados de proveniência relacionados aos metadados educacionais [Rosa et al. 2019].

Nesse sentido, a utilização conjunta de ontologias e técnicas de *learning analytics* possibilita transformar o conteúdo coletado dentro dos AVAs em informações relevantes para o aperfeiçoamento do ensino [Chatti et al. 2013]. Tais conteúdos podem ser usados por aplicações inteligentes, como por exemplo, sistemas de recomendação educacionais [Dwivedi and Roshni 2017], que consomem os metadados e aplicam-nos dentro de um cenário de *e-learning*. Nesse contexto, o uso de ontologias torna-se uma ótima alternativa [Oliveira et al. 2017]. Sendo assim, esse trabalho propõe uma arquitetura de integração que utiliza uma ontologia desenvolvida para esse cenário, denominada Prov4LA, para padronizar a geração de dados de *LA* e prover suporte a sistemas de recomendação de conteúdos educacionais.

2. Fundamentação Teórica

2.1. Learning Analytics

Learning Analytics (*LA*) é um estudo voltado para análise de informações extraídas de dados educacionais, a fim de converter o conteúdo coletado em metadados válidos para melhoria dos processos de ensino e aprendizagem [Chatti et al. 2013]. O fluxo de desenvolvimento do *LA* tem como fonte primária a coleta dos dados educacionais, o pré-processamento, a aplicação de técnicas de mineração de dados para entendimentos dos mesmo e, por último, a aplicação de métodos computacionais para visualização do conhecimento extraído. O objetivo desse processo é, através do conhecimento gerado, discutir tomadas de decisão estratégicas pelos usuários desses sistemas [Dyckhoff et al. 2012]. Dentro de *LA*, diversos ambientes educacionais podem ser fomentados como é o caso de sistemas de recomendação educacionais [Dwivedi and Roshni 2017].

2.2. Uso de Ontologias para Representação de Dados na Web

Sendo a web um ambiente volátil, onde a produção de conteúdo é livre e dependente apenas do domínio de cada repositório, as publicações ficam inerentes à forma com que cada um dos repositórios de dados produzem e divulgam seus dados. Considerando ainda que a parte da geração de informação é realizada por meios manuais através de anotações realizadas por humanos, torna-se um processo de classificação de dados exaustivo e de alto custo, devido as demandas de tempo para produção das *tags* e as dependências vinculadas ao conhecimento de cada especialista sobre a área trabalhada em questão

[Baeza-Yates et al. 2011]. Nesse cenário, o uso de ontologias torna-se essencial para padronizar e interoperar a forma com que esses metadados são gerados [Horrocks 2008].

Com o passar dos anos, diversos padrões para publicação de dados foram propostos para o melhoramento da qualidade das informações produzidas, fundamentando o que é chamado de Web Semântica [Berners-Lee et al. 2001]. Dentro da Web Semântica, o termo Dados Ligados (do inglês, *Linked Data*) é utilizado para descrever um conjunto de práticas para publicar, conectar, compartilhar e divulgar dados estruturados na Web sobre diversos domínios através do uso de ontologias [Rautenberg et al. 2019]. Seguindo os padrões de dados ligados no cenário educacional, a Onto4LA [Carchedi et al. 2018] é uma ontologia que utiliza conceitos de Dados ligados abertos. Sua utilização é recomendada para facilitar e padronizar a integração de dados entre diversos tipos de informações gerados por *frameworks* voltados para LA. Por meio da Onto4LA é possível modelar a troca de informação através de uma estrutura baseada em eventos. Cada evento apresenta como classes mínimas: um agente, um local, uma duração, uma produção e uma fonte.

2.3. Ontologias para Proveniência de Dados

Outra questão importante a ser discutida sobre publicação de dados, tanto na Web quanto em outros contextos, é a manutenção dos dados. Como cada nova informação pode ser gerada de diversas formas em uma mesma aplicação, é importante rastrear a maneira com que cada novo dado é produzido a fim de entender porque, por quem e onde esses dados foram gerados [Cheney et al. 2009]. Com esse objetivo, surgiu a proveniência de dados.

Proveniência pode ser compreendida como a informação que pode ser coletada de diversas formas e em diversos níveis de detalhamento [Simmhan et al. 2005]. Proveniência de dados também pode ser compreendida como dados que falam sobre outros dados trazendo, como informações adicionais, a origem e o histórico de modificações do dado original. Para formalizar e padronizar as formas de captura de proveniência de dados, a W3C criou um modelo semântico PROV [Missier et al. 2013, Moreau and Groth 2013]. O PROV é composto por um conjunto de documentos construído para que tanto usuários quanto desenvolvedores possam aplicar os conceitos de proveniência em suas aplicações. Dentro da estrutura do PROV, a PROV-O é o modelo utilizado para padronizar a captura de proveniência de dados através de uma ontologia.

3. Arquitetura de Integração e Ontologia Prov4LA

3.1. Prov4LA

Embora a Onto4LA, apresentada na Seção 2.2, abranja a maior parte dos *frameworks* para LA, peca quando o assunto é recomendação de conteúdos educacionais. Por exemplo, em [Barrère et al. 2018] é apresentado um processo de recomendação de vídeos educacionais no AVA Moodle através de indexação e recomendação semântica de videoaulas. O processo de recomendação utiliza de diversas etapas, desde a utilização de ferramentas de reconhecimento automático de fala, até recomendação baseada em metadados semânticos. Através da Onto4LA não é possível identificar quais são as principais ferramentas de recomendação e quais obtiveram os melhores resultados durante a recomendação. Além disso, não é possível avaliar alunos que participam de cursos com assuntos similares para saber se eles são candidatos a receber recomendações similares a de outros alunos com

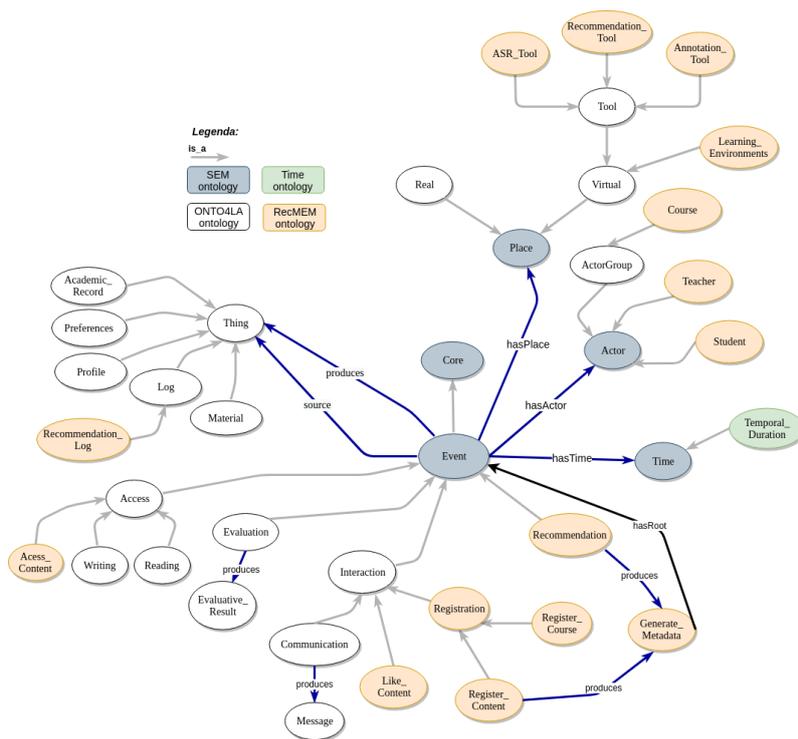


Figura 1. Onto4LA + Recomendação(RecMEM)

perfis análogos receberam. Com a necessidade de analisar o cenário de recomendação de mídias educacionais (RecMEM) em AVAs, a Onto4LA foi expandida (Figura 1).

A expansão da Onto4LA teve como objetivo inserir novas ferramentas e resultados aplicados ligados diretamente com o ambiente de recomendação. Na Figura 1 apresentam-se novas classes, o evento de recomendação e outros eventos ligados ao monitoramento das atividades dos alunos dentro do AVA. A ideia desse monitoramento é além de entender o comportamento de cada aluno durante a utilização de AVAs, gerar metadados que auxiliem também outras ferramentas para o cenário de *learning analytics*, como recomendação de materiais, exercícios, caminhos de aprendizagem, parceiros para grupos de estudo, algoritmos de aprendizado curricular adaptativo, entre outros. [Machado et al. 2019, Machado et al. 2020].

Analisando a expansão da Onto4LA para o cenário apresentado no parágrafo anterior, percebe-se a característica de proveniência de atividades agregada a ela. Afim de formalizar a captura de informação através dos padrões propostos pela W3C criou-se a Prov4LA. A Prov4LA é uma ontologia de proveniência que estende os padrões básicos da PROV-O através da Onto4LA + RecMEM. A Figura 2 apresenta a visão geral da Prov4LA demonstrando através das linhas tracejadas as classes e propriedades pertencentes a Onto4LA. Com essa integração é possível utilizar todas as propriedades da PROV-O para capturar o rastreamento dos dados produzidos além de gerar as inferências próprias da Onto4LA.

3.2. Arquitetura de Integração Proposta

Para que seja possível capturar a proveniência dos dados em AVA, é necessário que a Prov4LA seja utilizada de forma correta. Para isso, é fundamental a monitoração da ge-

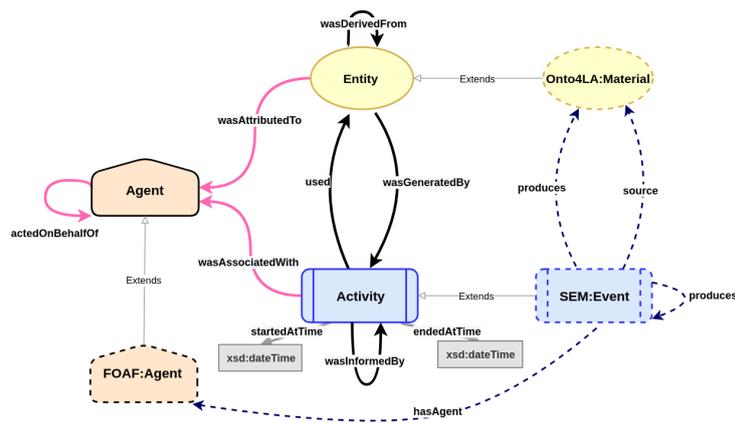


Figura 2. Prov4LA

ração de dados dentro do AVA e instanciação dos mesmos dentro da Prov4LA. Pensando nessa integração, é proposto a criação de um *plugin* que captura os dados gerados no AVAs, processa-os na Prov4LA e, por fim, submete os resultados a aplicações para recomendação de conteúdo ou *dashboards* para monitoramento e avaliação dos dados gerados para LA. No contexto de ambientes virtuais de aprendizagem, a iniciativa de se trabalhar com um *plugin* torna-se uma alternativa viável, já que é uma prática comum nesse cenário. Por exemplo, para o AVA Moodle já existem diversos *plugins* utilizados pelo usuários. A Figura 3 apresenta a visão geral do funcionamento da arquitetura.

Inicialmente, os alunos serão cadastrados nos cursos de um AVA, no qual irão realizar suas atividades normalmente. O AVA irá disponibilizar o acesso ao material requisitado por cada aluno, salvando seus logs de acesso. Nessa etapa, o *plugin* se integra ao AVA funcionando como um *listener* entre as requisições realizadas pelo AVA e o acesso ao banco de dados do AVA. Os dados capturados pelo *plugin* são instanciados (populados) na Prov4LA via OWLAPI ¹, o *reasoner* é acionado e as inferências são capturadas. Ao fim, o *plugin* irá trabalhar com um SGBD NoSQL pelo qual os *dashboards* poderão consultar os resultados processados pela Prov4LA, coletando os históricos de atividade e outras informações para utilização em LA, e os sistemas de recomendação poderão consultar quais são os melhores alunos para recomendar determinados assuntos ou então, dado que algum aluno recebeu uma recomendação anteriormente, quais outros possuem perfis parecidos e são potenciais para receber a mesma recomendação.

A utilização dessa arquitetura torna-se uma alternativa viável, pois utilizando a Prov4LA, além de gerar a proveniência dos dados e facilitar a integração destes entre AVAs, há uma padronização dos dados gerados para aplicações, tanto para o cenário de *Learning analytics* quanto para sistemas de recomendação de conteúdos educacionais.

3.3. Cenário de Aplicação da Arquitetura

Como prova de conceito da arquitetura de integração proposta, esta subseção descreve um cenário possível para sua aplicação da arquitetura. Especificamente, descreve-se um exemplo de curso ministrado no Moodle, o qual corresponde a um curso X que é ministrado no AVA e onde os alunos A e B participam do mesmo curso. O Aluno A acessa os conteúdos disponibilizados pelo professor e, através de um evento de recomendação

¹<http://owlcs.github.io/owlapi/>

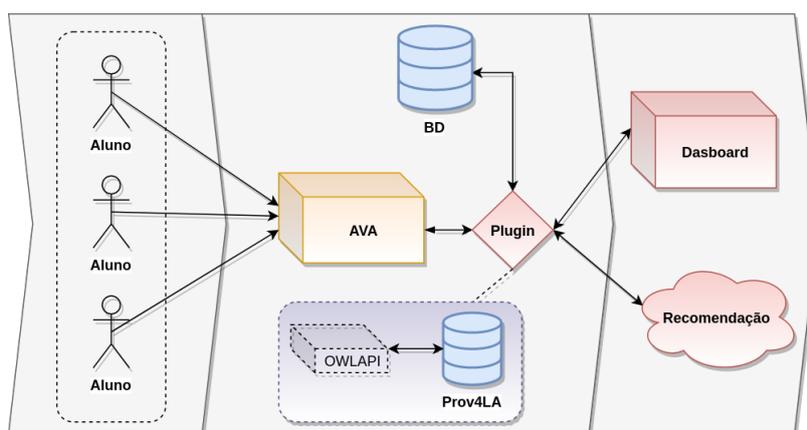


Figura 3. Arquitetura de integração

recebe uma recomendação de um conteúdo Y como resultado. Como o Aluno B, que faz o mesmo curso que o aluno A e recebeu, inicialmente, acesso ao mesmo conteúdo que A e A recebeu a recomendação Y, então é inferido que o Aluno B torna-se um potencial agente a receber a mesma recomendação de Y. A Figura 4 apresenta a exemplificação do caso descrito dentro da ontologia.

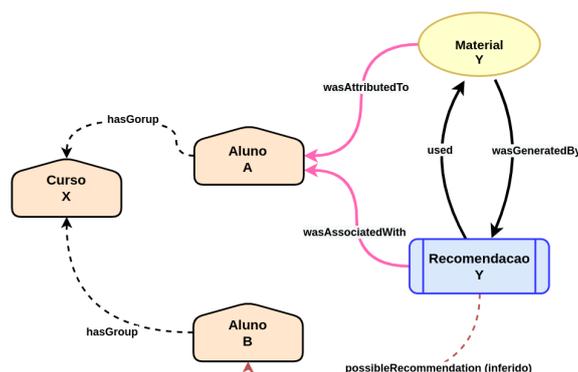


Figura 4. Exemplificação do cenário de aplicação

4. Considerações Finais e Trabalhos Futuros

Este trabalho apresentou uma estrutura para padronizar a geração de dados de LA e prover assistências a sistemas de recomendação. Foi apresentada a Prov4LA, uma ontologia para captura de proveniência de eventos dentro de ambientes virtuais de aprendizagem. Para utilização correta da Prov4LA foi proposta uma arquitetura de integração na qual um *plugin* funcionará como um *listener* que captura os acessos a banco de dados realizados por cada ator dentro de um AVA. Através do cenário de aplicação apresentado, foi possível identificar e analisar o rastreamento de atividades, além dos principais eventos ocorridos.

Como trabalhos futuros está prevista a etapa de validação da arquitetura proposta dentro de um curso ministrado no Moodle. Além disso, espera-se a utilização dos metadados gerados dentro de outras ferramentas de LA, por exemplo, com a proveniência coletada conseguir gerar caminhos de aprendizados automáticos para cada aluno, afim de melhorar a interação entre os alunos e os ambientes virtuais de aprendizagem.

Referências

- Baeza-Yates, R., Ribeiro, B. d. A. N., et al. (2011). *Modern information retrieval*. New York: ACM Press; Harlow, England: Addison-Wesley,.
- Barrére, E., Souza, J., Vitor, M. A., and de Almeida, M. A. (2018). Recomendação automática de videoaulas no moodle. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 29, page 1613.
- Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The semantic web. *Scientific american*, 284(5):34–43.
- Burger, T., Zaihrayeu, I., Andrews, P., Babenko, D., Pane, J., and Popov, B. (2010). Report on the state-of-the-art and requirements for annotation representation models. Technical report, University of Trento.
- Carchedi, L., Souza, J., Barrére, E., and Mendonça, F. (2018). Onto4la: uma ontologia para integração de dados educacionais. *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, 7(1):439.
- Chatti, M. A., Dyckhoff, A. L., Schroeder, U., and Thüs, H. (2013). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6):318–331.
- Cheney, J., Chiticariu, L., and Tan, W.-C. (2009). Provenance in databases: Why, how, and where. *Foundations and Trends® in Databases*, 1(4):379–474.
- Costa, E., Aguiar, J., and Magalhães, J. (2013). Sistemas de recomendação de recursos educacionais: conceitos, técnicas e aplicações. *Jornada de Atualização em Informática na Educação*, 1(1).
- da Silva Lemos, D. L. and Souza, R. R. (2020). Representação de recursos multimídia na web: uso e reúso de padrões de anotação. *Perspectivas em Ciência da Informação*, pages 202–232.
- Dwivedi, S. and Roshni, V. K. (2017). Recommender system for big data in education. In *2017 5th National Conference on E-Learning & E-Learning Technologies (ELEL-TECH)*, pages 1–4. IEEE.
- Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., and Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society*, 15(3):58–76.
- Horrocks, I. (2008). Ontologies and the semantic web. *Communications of the ACM*, 51(12):58–67.
- Machado, M. d. O. C., Barrére, E., and Souza, J. (2019). Solving the adaptive curriculum sequencing problem with prey-predator algorithm. *International Journal of Distance Education Technologies (IJDET)*, 17(4):71–93.
- Machado, M. d. O. C., Bravo, N. F. S., Martins, A. F., Bernardino, H. S., Barrere, E., and de Souza, J. F. (2020). Metaheuristic-based adaptive curriculum sequencing approaches: a systematic review and mapping of the literature. *Artificial Intelligence Review*, pages 1–44.

- Missier, P., Belhajjame, K., and Cheney, J. (2013). The w3c prov family of specifications for modelling provenance metadata. In *Proceedings of the 16th International Conference on Extending Database Technology*, EDBT '13, pages 773–776, New York, NY, USA. ACM.
- Moreau, L. and Groth, P. (2013). Provenance: an introduction to prov. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 3(4):1–129.
- Oliveira, F. H., Salvador, L. N., and Novais, R. L. (2017). Um experimento com a ontologia ims ld na construção de modelos conceituais para e-learning. *X Seminar on Ontology Research in Brazil*.
- Rautenberg, S., de Souza, L., Dall’Agnol, J. M. H., and Michelon, G. A. (2019). *Guia Prático para Publicação de Dados Abertos Conectados na Web*. Appris Editora e Livraria Eireli-ME.
- Rosa, F. L., da Silva Machado, R., Primo, T. T., Yamin, A. C., and Pernas, A. M. (2019). Análise de desempenho de ferramentas para persistência de dados ontológicos em triplas: Experimentos e resultados. *XII Seminar on Ontology Research in Brazil*.
- Simmhan, Y. L., Plale, B., and Gannon, D. (2005). A survey of data provenance in e-science. *SIGMOD Rec.*, 34(3):31–36.
- Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., and Ciravegna, F. (2006). Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Journal of Web Semantics*, 4(1):14–28.