

Recherche intelligente de processus d'analyse de traces d'e-learning *via* des inférences sémantiques

Alexis Lebis^{1,2}, Marie Lefevre², Vanda Luengo¹, Nathalie Guin²

¹ SORBONNE UNIVERSITÉS, UPMC, CNRS, LIP6 UMR 7606, F-75005 Paris, France
alexis.lebis@lip6.fr, vanda.luengo@lip6.fr

² UNIVERSITÉ DE LYON, CNRS, LIRIS UMR 5205, F-69222 Villeurbanne, France
marie.lefevre@univ-lyon1.fr, nathalie.guin@univ-lyon1.fr

Résumé : Le partage et la réutilisation de processus d'analyse de traces d'apprentissage, ainsi que leur adaptation à des contextes différents, sont devenus des enjeux importants dans le domaine des Learning Analytics. L'objectif est double : impliquer et soutenir la communauté dans le cycle du processus d'analyse, de son élaboration à son utilisation, et lui donner les outils nécessaires pour promouvoir une co-construction de ces analyses. Dès lors, il apparaît important de fournir aux différents acteurs des outils pour les assister dans leur tâche respective, au sein de l'analyse. Dans cet article, nous présentons une assistance à la réutilisation des processus d'analyse de traces d'apprentissage *via* une recherche exploitant la représentation sémantique de ces processus d'analyse.

Mots-clés : Processus d'analyse de traces, Environnements Informatiques pour l'Apprentissage Humain (EIAH), narration, sémantique, ontologie, inférence, raisonnement, assistance.

1 Introduction

Les processus d'analyse de traces d'apprentissage, mis en oeuvre pour répondre à des besoins d'analyse spécifiques dans le domaine des EIAH, sont des artefacts informatiques complexes. Ils sont concrétisés par une succession ordonnée d'opérations, implémentées dans un outil d'analyse, et appliquées sur des traces d'apprentissages (Mandran *et al.*, 2015). Le partage, l'adaptation et la réutilisation de ces *processus* sont complexes voire peu pertinents (Clow, 2013), en cela qu'ils sont intrinsèquement contraints à la fois par les contextes d'apprentissage et les spécificités techniques des données et des outils d'analyse.

Pour renforcer ces propriétés nécessaires à la capitalisation des *processus*, nous avons, dans nos précédents travaux, changer le paradigme dans lequel ils s'ancrent en nous appuyant sur une approche narrative (Lebis *et al.*, 2018). Il s'agit d'utiliser une ontologie¹ pour les structurer sémantiquement ainsi que pour y intégrer les informations en lien avec l'analyse (*e.g.* contexte, hypothèse). Nos résultats expérimentaux confirment qu'aborder la description des *processus* par la narration est une piste solide pour les capitaliser, adopter une ontologie offre des possibilités d'assistances renforcées pour les différents acteurs de l'analyse.

En ce sens, nous présentons dans cet article une recherche intelligente des *processus*. Ce mécanisme, pionnier pour les outils d'analyse de l'état de l'art, a pour nature de permettre au système d'exploiter les propriétés sémantiques existantes ou induites par notre ontologie pour, d'une part, interpréter le besoin d'analyse de l'utilisateur et, d'autre part, comprendre foncièrement les *processus*, leurs étapes et les choix d'analyse adoptés. C'est ainsi qu'il est possible de répondre au besoin d'analyse utilisateur en lui proposant des *processus* pertinents.

2 Présentation de la recherche basée sur l'inférence sémantique

La recherche intelligente repose tout d'abord sur la possibilité pour l'utilisateur de décrire son besoin d'analyse dans le système lors de recherche, afin que ce dernier puisse l'exploiter

1. <https://perso.liris.cnrs.fr/alexis.lebis/research/CAPTEN/ontology.html>

et utiliser les bonnes relations sémantiques. Pour ce faire, nous définissons un besoin d'analyse \mathcal{B} comme un ensemble fini de **dimensions** \mathcal{D} . Chacune représente une propriété atomique du besoin (*e.g.* contexte, objectif). Elles sont décrites par l'utilisateur *via* des **tokens** sémantiques : ce sont les concepts et les propriétés qui existent dans l'ontologie. Par exemple, dans notre approche, un enseignant pourrait définir le contexte de son besoin d'analyse comme étant des apprenants impliqués au sein d'un jeu sérieux. Cela revient à définir une dimension $\mathcal{D}_{\text{contexte}} = \text{involved}(\text{Apprenant}, \text{Jeu_Sérieux}), \mathcal{D}_{\text{contexte}} \in \mathcal{B}$. Ainsi, un processus faisant intervenir une personne inscrite dans une unité d'enseignement où un jeu sérieux est dispensé peut être un candidat potentiel après raisonnement sémantique.

Puisque la description des processus est riche, il est toutefois nécessaire de circonscrire les éléments de l'ontologie contribuant activement aux dimensions, pour en limiter notamment le bruit. Aussi, chaque dimension est projetée individuellement dans l'ontologie ; chaque élément y contribuant est identifié et un score d'importance vis-à-vis de la dimension lui est associé. Ces règles de projection sont appelées **patrons de recherche** et peuvent aussi exploiter les propriétés sémantiques des éléments (*e.g.* transitivité, subsomption).

Néanmoins, le fait de ne pas trouver un token ne signifie pas qu'un concept similaire n'existe pas dans les processus à disposition qui pourraient intéresser l'utilisateur. D'après les travaux d'approximation sémantique menés dans Corese (Corby *et al.*, 2006), nous définissons la notion de **termes similaires** . Deux termes sont dits similaires selon un certain score lorsqu'ils partagent tous deux un même arbre taxonomique (distance verticale ou horizontale), ou lorsqu'ils sont indiqués comme tel au sein de l'ontologie (score arbitrairement choisi). Outre le fait qu'utiliser des termes similaires nous permettent d'étendre l'espace de recherche en raisonnant avec, ce score de similarité influe directement sur la pertinence des processus résultats. De surcroît, toujours soucieux de l'imprécision latente résidant dans les descriptions, nous définissons un double mécanisme d'approximation à l'échelle de l'utilisateur. Le premier est de lui permettre de **relaxer des contraintes** lors de la recherche, le deuxième est de lui permettre d'exprimer l'importance qu'il donne aux dimensions et tokens. Pour ce dernier point, nous avons proposé des **modificateurs flous** (*e.g.* extrêmement similaire, très peu équivalent) modifiant l'espace d'appartenance des résultats par pondération.

Enfin, toujours dans l'objectif de laisser l'utilisateur au centre du processus de décision, il est nécessaire de doter une telle recherche d'un mécanisme d' **explicabilité des résultats** . Conséquemment, chaque résultat de la recherche présente les éléments de l'ontologie ayant contribué et l'influence qu'ils ont eu sur leur score de pertinence respectif.

3 Conclusion

Nous avons mis en œuvre cette recherche intelligente dans un prototype de recherche² et avons testé notre proposition avec des scénarios d'usage. Globalement, les résultats préliminaires sont bons et très encourageants : nous avons réussi à obtenir les candidats les plus pertinents pour chacun de nos scénarios d'usages, alors que certains nécessitaient l'exploitation de relations sémantiques complexes. Il convient désormais d'évaluer nos travaux avec les différents acteurs impliqués dans l'analyse.

Références

- CLOW D. (2013). An overview of learning analytics. *Teaching in Higher Education*, **18**(6), 683–695.
- CORBY O., DIENG-KUNTZ R., GANDON F. & FARON-ZUCKER C. (2006). Searching the semantic web : Approximate query processing based on ontologies. *IEEE Intelligent Systems*, **21**(1), 20–27.
- LEBIS A., LEFEVRE M., LUENGO V. & GUIN N. (2018). Capitalisation of Analysis Processes : Enabling Reproducibility, Openness and Adaptability thanks to Narration. In *Proceedings of 8th International Conference on Learning Analytics and Knowledge*, p. 245–254, Sydney, Australia.
- MANDRAN N., ORTEGA M., LUENGO V. & BOUHINEAU D. (2015). Dop8 : merging both data and analysis operators life cycles for technology enhanced learning. In *Proceedings of LAK'15*, p. 213–217 : ACM.

2. <https://github.com/alexislebis/CAPTEN> - cf. branche *astrolabe*