

# Capitaliser les processus d'analyse de traces d'e-learning

Alexis Lebis<sup>1,2</sup>, Marie Lefevre<sup>2</sup>, Vanda Luengo<sup>1</sup>, Nathalie Guin<sup>2</sup>

<sup>1</sup> Sorbonne Universités, UPMC, Univ Paris 6, LIP6 UMR 7606

{alexis.lebis, vanda.luengo}@lip6.fr

<sup>2</sup> Université de Lyon, CNRS, Université Lyon 1, LIRIS, UMR5205

{alexis.lebis, marie.lefevre, nathalie.guin}@univ-lyon1.fr

**Keywords:** processus d'analyses, traces, e-learning, learning analytics, capitalisation, interopérabilité, educational data mining.

## 1 Contexte de la recherche

Dans nos travaux, nous nous intéressons à la capitalisation des processus d'analyse de traces dans le contexte du e-learning. Ce type d'apprentissage est défini par l'utilisation d'outils informatiques par des apprenants, en présence ou non de tuteurs. Au sein de ces outils, divers types de ressources pédagogiques sont disponibles (vidéos, quiz...). Ces dispositifs génèrent des données qui relatent l'interaction des utilisateurs entre eux (messages privés, forum...) ainsi que les activités faites sur les ressources des dispositifs. Nous parlons dès lors de *traces d'interactions*, qui sont considérées comme réservoirs de connaissances ; connaissances pouvant être découvertes par l'analyse de ces traces.

Une analyse est représentée par un processus d'analyse, ici de traces. Il s'agit de l'application de techniques, d'opérations et de méthodes, appelées *opérateurs* [5], pour produire des connaissances comme des modèles ou des indicateurs. Les acteurs de cette analyse peuvent être des statisticiens, des analystes, des chercheurs... Concevoir un processus d'analyse de traces est une tâche bien souvent complexe et fastidieuse : le besoin doit être correctement cerné et les données, ainsi que les connaissances du domaine, correctement communiquées. La publication des résultats est également cruciale afin que ceux-ci puissent être utilisés convenablement. Actuellement, les processus d'analyse sont dépendants des outils utilisés, des opérateurs qui y sont disponibles, ainsi que de la représentation des données contenues dans les traces. Cela fait disparaître l'objectif des opérateurs utilisés et élude le sens des données manipulées et la démarche intellectuelle. De fait, en pratique, pour un même besoin, un processus d'analyse peut être décliné sur différents outils d'analyse, en fonction du format des données concernées : la capitalisation est inexistante [3].

Lorsque nous parlons de capitalisation, nous entendons la faculté de disposer de l'existant pour en tirer profit. Pour un processus d'analyse cela concerne notamment sa réutilisation, son adaptabilité, sa modification, son enrichissement, pour le spectre des outils d'analyse disponibles. Dans la littérature, nous pouvons distinguer deux styles d'approches allant dans le sens de la capitalisation.

L'une est axée sur le partage des processus, où tout du moins d'une certaine partie des opérateurs le constituant. Actuellement, la tendance est de proposer un unique outil où les processus et opérateurs sont partagés entre différents utilisateurs, projets et besoins. Mais il existe des travaux proposant de partager certains éléments vers d'autres outils, bien qu'ils se heurtent à des problèmes complexes, comme l'interopérabilité technique, la perte de sens des opérateurs utilisés ou encore l'évolution des outils d'analyses [1]. L'autre axe se concentre sur les données, où l'objectif est de palier au problème de leur représentation. Des travaux proposent d'effectuer un "mapping" vers des formalismes plus génériques, proposant ainsi des cadres de travail contrôlés pour ré-exploiter facilement les données. Cependant, augmenter la généricité des données n'exempte pas les opérateurs utilisés d'être contraints par les outils d'analyse : cela ne résout donc pas le problème de capitalisation des processus d'analyse de traces [2].

Malheureusement, il n'est pas non plus possible de trivialement de combiner ces deux axes, principalement dû aux dépendances structurelles des opérateurs et des données. Et quand bien même, de telles approches n'apportent aucune assistance aux utilisateurs dans la compréhension et la réutilisation directe des analyses. Malgré cela, ces travaux constituent une preuve en soit que le besoin d'échange et de partage est bien réel.

## 2 Abstraire pour partager

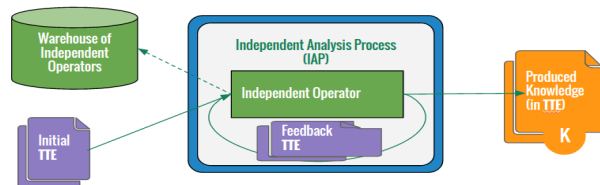
Notre objectif est d'apporter une solution au problème de la capitalisation des processus d'analyses de traces d'interactions. L'ambition est, à terme, de disposer d'un entrepôt de processus d'analyse éprouvés qui offre un panorama de l'existant pour les partager, les réutiliser et les enrichir, tout en indiquant les outils d'analyses capables de les réaliser. Pour cela, nous proposons d'abstraire et de documenter les processus d'analyse pour représenter la méthodologie nécessaire à l'obtention des connaissances, de manière indépendante des outils d'analyses, en décrivant les différents éléments constituant de l'analyse. À la différence des approches décrites auparavant, l'objectif n'est pas de rendre le processus d'analyse exécutable sur des valeurs pour produire une connaissance concrète. Notre objectif est de décrire le processus ; cette description servira de manifeste pour les analystes qui implémenteront le processus dans les outils à disposition de la communauté. Cette approche a l'avantage d'exploiter les outils d'analyses déjà existants. Cela nous permet de nous inscrire dans une démarche d'aide à la création, à l'expression et à la décision, nous évitant de proposer un nouvel outil de calcul, contraignant lui aussi par ses caractéristiques intrinsèques.

L'étude de la littérature nous montre que la capitalisation peut se faire à deux moments importants du cycle de vie d'une analyse : (1) au moment de la démarche cognitive qu'a un acteur lors de la création d'un processus, pour répondre à un besoin. Le processus n'ayant pas été encore implémenté, il n'est pas altéré par des spécificités techniques. Ce moment se prête alors bien à la réification, par la description, de la démarche -abstraite- dans un formalisme indépendant des outils d'analyse. Et (2) lorsque qu'un acteur a terminé, et éventuellement

validé, une analyse décrite dans un ou plusieurs outils, et désire la partager. Il s'agit alors de représenter et d'abstraire le processus implémenté grâce à un formalisme indépendant. Cela équivaut à une démarche de rétro-conception, au sens où l'acteur extrait du processus d'analyse les éléments saillants impliqués et les appaillie à leur concept générique.

Pour qu'un processus d'analyse puisse être considéré comme indépendant des outils, il est nécessaire que chacun des éléments le constituant le soit aussi (paramètres, données, opérateurs...). Nous proposons donc un formalisme permettant de représenter un opérateur de manière indépendante des outils d'analyse. Un opérateur indépendant est l'expression d'un même concept véhiculé par un ou plusieurs opérateurs, qu'importe leur spécificités techniques au sein des outils d'analyse. Par exemple, réaliser un intervalle de temps peut être fait de différentes manières en fonction des outils, mais un opérateur indépendant *Time Range* représentera le concept de créer un intervalle, sans dépendre d'outils. Nous l'avons vu précédemment, les processus d'analyse sont basés sur l'exploitation des données et sont intrinsèquement liés à elles. Il n'est donc pas possible pour obtenir une description efficace du processus de ne pas les considérer. Cependant, ces données sont elles aussi soumises à des contraintes, comme leur formalisme (e.g. CSV ou RDF), posant des problématique d'interopérabilité et de capitalisation. Dès lors, il convient de trouver un point d'équilibre entre l'information contenue, son utilité et les contraintes techniques qu'elle génère. Nous suggérons, comme point d'équilibre pour une capitalisation efficace, de travailler avec les concepts que dégagent les données -appelés types d'éléments tracés (TTE, e.g. heure de début), à l'instar des opérateurs indépendants, et non pas directement avec leurs valeurs (e.g. 12h45) : cela permet de s'affranchir du problème de la représentation des données. Le descripteur de l'analyse déclare alors les TTEs requis initialement pour effectuer l'analyse. Ces TTEs seront exploités par les opérateurs indépendants pour générer des informations tout au long de la description du processus.

Nous avons regroupé ces notions d'*opérateur indépendant* et de *types d'éléments tracés (TTE)* au sein d'un méta-modèle permettant de décrire des processus d'analyse indépendants. Les différents éléments de notre formalisme ont été obtenus par des études empiriques de différents outils d'analyse existants. Ils ont été implémenté au sein d'une application web qui a servie de support à six expérimentations impliquant différents acteurs [4].



**FIGURE 1.** Représentation des éléments impliqués dans la description d'un processus d'analyse indépendant permettant, sur cet exemple, d'obtenir une connaissance (K).

### 3 Problématiques actuelles et à venir

Bien que les résultats des expérimentations menées soit encourageants, il faut cependant noter plusieurs limites, notamment : (1) le manque de sémantique des différents éléments manipulables dû au caractère abstrait de l'approche, limitant de ce fait la compréhension globale de l'analyse et les facultés descriptives et (2) un manque avéré de feedback pour les utilisateurs, concernant entre autres, l'influence que peut avoir un opérateur indépendant sur des TTEs.

Nous travaillons actuellement sur la question d'une description plus riche des processus d'analyses indépendants, de leurs objectifs et de leurs contextes applicatifs ; cela afin qu'ils puissent se suffire à eux même sur un plan sémantique, et pour servir de support d'aide à la compréhension, à la décision, à la réutilisation et à la diffusion scientifique. Nous prévoyons d'exploiter le vocabulaire xAPI au sein d'une ontologie décrivant des processus d'analyses.

Cependant, cette description requiert aussi le besoin de s'adapter aux spécificités propres au domaine du e-learning. Il est intéressant de se demander s'il est possible de faire ressortir des concepts récurrents et des habitudes d'analyses propres aux données de e-learning, dans l'idée d'aiguiller l'effort scientifique. Pour ce faire, nous envisageons un écosystème évolutif et relationnel concernant la base de vocabulaire contrôlé offerte par xAPI, mais aussi des opérateurs et processus d'analyses indépendants.

Cette description sémantique soulève également la question de savoir quelles informations utiles peuvent être inférées par une description des processus d'analyse de haut niveau ? Nous nous posons aussi cette question afin d'assister l'utilisateur dans son élaboration *via* des mécanismes de recommandation et de lui apporter des feedbacks pertinents et contextuels. Il faudra donc voir comment assister efficacement les différents acteurs de l'analyse... Il serait intéressant, par exemple, de rechercher des processus connexes à des besoins, ou encore de s'assurer de la crédibilité scientifique d'une analyse et de la bonne lecture des résultats.

### Références

1. PMML 4.1 by DMG. <http://dmg.org/pmml/pmml-v4-1.html>, accessed : 2016-09-20
2. Choquet, C., Iksal, S. : Modeling tracks for the model driven re-engineering of a tel system. *Journal of Interactive Learning Research* 18(2), 161 (2007)
3. Dyckhoff, A.L., Zielke, D., Bültmann, M., Chatti, M.A., Schroeder, U. : Design and implementation of a learning analytics toolkit for teachers. vol. 15, pp. 58–76. *JSTOR* (2012)
4. Lebis, A. : Vers une capitalisation des processus d'analyse de traces. In : *Rencontres Jeunes Chercheurs en EIAH (RJC-EIAH 2016)*. Montpellier, France (Jun 2016)
5. Mandran, N., Ortega, M., Luengo, V., Bouhineau, D. : Dop8 : merging both data and analysis operators life cycles for technology enhanced learning. In : *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*. pp. 213–217. *ACM* (2015)