



Méthodologie d'enrichissement semi-automatique d'un profil-apprenant basée sur une ontologie

Laurie Acensio

► To cite this version:

Laurie Acensio. Méthodologie d'enrichissement semi-automatique d'un profil-apprenant basée sur une ontologie. Rencontre Jeunes Chercheurs-EIAH (Environnement Informatique pour l'Apprentissage Humain), Jul 2016, Montpellier, France. hal-01375949

HAL Id: hal-01375949

<https://hal.archives-ouvertes.fr/hal-01375949>

Submitted on 6 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives | 4.0 International License

Méthodologie d'enrichissement semi-automatique d'un profil apprenant basée sur une ontologie

Laurie Acensio

2^{ème} année de doctorat

Laboratoire Lip6, 4 place Jussieu 75005, Paris, France

laurie.acensio@lexiane.fr

Résumé

Le contexte de la formation continue amène des adultes-apprenants de profils hétérogènes à s'inscrire à une même formation. Il est alors difficile de concilier leurs attentes au sein d'un groupe de formation en présentiel.

Notre proposition consiste à représenter et à enrichir le profil initial de l'adulte-apprenant en utilisant une formalisation ontologique.

Dans la continuité de cette analyse, il s'agit de composer par la suite des groupes de formation homogènes et raisonnés mais réunissant un nombre minimal de participants pour des contraintes budgétaires et logistiques.

Cet article de recherche présente une méthodologie pour clarifier et analyser les besoins de formation des adultes-apprenants instrumentés à partir des traces recueillies par un EPA (Environnement Personnel d'Apprentissage) communautaire.

Présentation

• Contexte

Ce travail s'inscrit dans le cadre du projet de recherche de développement TICLex¹ au sein d'un organisme de formation juridique.

Le contexte sociétal de la FLTV² et du CPF³ incite l'adulte-apprenant à s'engager dans une action de formation durant toute sa carrière professionnelle. Dans le domaine juridique, la formation professionnelle est souvent considérée comme indispensable du fait de l'évolution constante de la législation et de la réglementation mais reste une activité secondaire à l'activité professionnelle. Compte tenu de cette contrainte du temps de travail accordé à la formation, l'adéquation entre le choix de la formation et les besoins de l'adulte-apprenant s'impose davantage qu'en formation initiale. De plus, nos observations sur le terrain démontrent que les besoins de formation des

adultes-apprenants sont fortement individualisés et contextualisés selon leurs propres problématiques. L'adulte-apprenant est confronté à la difficulté de définir et d'adapter ses propres besoins liés à son contexte organisationnel à une offre de formation tierce générique et limitée. Sur le contenu de la formation, la personnalisation de la thématique est nécessaire pour espérer atteindre les objectifs visés par l'adulte-apprenant. Dans le contexte de formation hybride que nous proposons, le formateur aborde la session en groupe composé d'adultes-apprenants ayant des socles de compétences et d'expériences très disparates. Malgré cette contrainte, il apparaît que la session de formation en présentiel est une situation de rapprochement entre adultes-apprenants propices pour instaurer une dynamique de groupe.

Il nous apparaît pertinent de personnaliser tout autant le contenu éditorial de la formation que le groupe dans lequel l'adulte-apprenant va s'impliquer.

Dans cet article, après avoir rappelé quelques éléments de l'état de l'art, nous présentons notre problématique et la méthodologie pour enrichir le profil-apprenant à partir d'un ensemble d'informations déduites de son activité de formation.

• Etat de l'art

La majorité des travaux scientifiques utilisent les techniques d'algorithmes de classification pour la composition automatique de groupes de formation (Anaya et Boticario, 2009; Hogo, 2010; Pollalis et Mavrommatis, 2009). Quelques travaux utilisent une approche basée sur une ontologie (Isotani et al., 2009 ; Ounnas 2007) préconisant une approche hybride pour une meilleure qualité des résultats. Bien que les méthodes diffèrent, nous pouvons souligner le point commun à l'ensemble des travaux dont le calcul de la similarité se base principalement sur les caractéristiques explicites du profil-apprenant (exemple : style d'apprentissage, préférences, centres d'intérêt).

Dans le cadre de notre travail, nous portons une attention particulière à la thématique de formation, qui constitue la raison principale de l'adulte-apprenant à s'engager dans une formation. Le « besoin de

¹ www.ticlex.fr

² Formation Tout au Long de la Vie

³ Compte Personnel de Formation

formation » est défini par l'AFNOR⁴ comme étant « l'identification d'un écart susceptible d'être réduit par la formation entre les compétences d'un individu ou d'un groupe à un moment donné et celles attendues ». Le besoin est préalablement défini par une analyse effectuée selon la politique des ressources humaine de l'organisation-commanditaire (entretien annuel, OPCA, évaluation...). La formation vise ainsi à combler un manque opérationnel du salarié au sein de son poste professionnel. Cependant, cette notion se révèle être en pratique subjective et ambiguë. Dans le domaine de la recherche en psychologie cognitive, la conduite et l'analyse des besoins dans un contexte professionnel ont été théorisées dans les années 1990 notamment avec les travaux de (Lapointe 1993) et de (Bourgeois 1991), mais restent peu expérimentées à ce jour. Selon (Bourgeois 1991), l'analyse des besoins nécessite un effort de réflexion, d'observation et de collecte de données (faits, opinions, attitudes) provenant de sources multiples visant à identifier et documenter les écarts (besoins) entre le résultat actuel et un résultat désiré. Dans cette optique, l'analyse des besoins est à la fois un processus d'élucidation et de négociation sociale. Dans le domaine informatique, de nombreux travaux scientifiques ont exploité les traces d'interactions pour optimiser le processus d'apprentissage. Ce champ d'application nous semble le plus proche de notre problématique d'analyse de besoin de formation. Les techniques de diagnostic citées par (Lallé 2013) comme le Knowledge Tracing, Constraint-based modeling, sont principalement axées sur l'évaluation des connaissances de l'apprenant. Ces techniques ont été largement exploitées et sont par conséquent matures mais restent liées à un domaine particulier et n'abordent pas la question du contexte. Cette nécessité de contextualiser les traces d'interactions a été soulignée lors du projet MULCE⁵ (Reffay et Betbeder, 2009). La méthode de structuration des traces proposées facilite ainsi le partage et la compréhension des résultats scientifiques publiés entre communautés de chercheurs tout en respectant les problématiques liées à la protection de la vie privée. D'un point de vue comportemental, les travaux de (Michelet et Luengo, 2012) ont révélé des contradictions intra-apprenant lors de ses activités pédagogiques. Ces conclusions rejoignent les travaux de l'auteur (Zekri et al., 2015) qui mettent en avant la difficulté de modéliser l'imprécision des résultats des tests effectués par le patient dans une finalité de diagnostic médical. Actuellement, les travaux scientifiques démontrent que l'enrichissement du profil de l'utilisateur via les traces d'interactions permet de rendre observable son évolution. Cependant, il apparaît nécessaire de prendre en considération l'imprécision et l'incertitude consubstantielles à ces traces.

⁴ Association française de normalisation-NF X 50-750

⁵ MULTimodal Corpus Exchange

• Problématiques

Le contexte de formation hybride favorise des actions de formation dispersées. Nous avons choisi de développer un prototype d'EPA (Environnement Personnel d'Apprentissage) pour unifier l'ensemble des activités de formation formelle et non-formelle effectué par l'adulte-apprenant. Bien que la conception de l'EPA a été orientée principalement vers un système contrôlé par l'adulte-apprenant, nous avons privilégié des fonctionnalités collaboratives (visualisation des groupes de formation, forum privé et public, réseautage social) pour favoriser la prise de conscience de groupe et éviter ainsi un isolement de l'adulte-apprenant. Cependant, les contraintes contractuelles (devoir de réserve, clause de confidentialité...) du contexte d'activité peuvent limiter les interactions en ligne comme l'entraide et la co-construction des connaissances. Il apparaît essentiel de ponctuer le parcours de formation de sessions en présentiel pour rapprocher l'adulte-apprenant de ses pairs. Cependant, le caractère spontané des sessions en présentiel crée des groupes de formation qui sont par nature éphémères, ce qui peut altérer la cohésion des relations interpersonnelles entre les adultes-apprenants. Côté utilisateur, notre positionnement est de considérer l'EPA comme un support de médiation informatisé pour maintenir un lien de l'adulte-apprenant avec les différents intervenants de la formation.

Au niveau de l'organisme de formation, nous utilisons cet environnement informatique pour stocker et exploiter les profils initiaux des adultes-apprenants et les affiner en fonction des traces générées. Nous portons une attention particulière aux spécificités du besoin de formation de l'adulte-apprenant, car il constitue l'objet principal du déclenchement d'une action de formation.

En amont du projet de formation, l'une des difficultés de l'organisation-commanditaire est la formulation et l'interprétation des besoins de formation. L'analyse des besoins est peu approfondie résultant d'une formulation du besoin souvent incomplète et/ou erronée. Cependant, cette étape s'avère fondamentale pour garantir d'atteindre les objectifs visés par l'adulte-apprenant.

Nous pouvons résumer notre problématique de la manière suivante :

-Comment identifier les besoins de formation des adultes-apprenants à partir de son expérience tracée sur l'EPA ?

D'un point de vue technique, cette problématique soulève des questions de recherches notamment liées à la méthodologie pour modéliser des traces d'interactions contextuelles.

Notre choix de modélisation doit également résoudre les difficultés liées à l'interopérabilité et l'hétérogénéité des données.

D'un point de vue cognitif, la formulation du besoin recouvre à la fois un mécanisme intellectuel et psychologique de l'adulte-apprenant, qui est par nature

imprévisible et incertain, ce qui rend difficile une automatisation. Nous sommes contraints de limiter notre approche cognitive en analysant les énoncés émis par l'adulte-apprenant. Pour cela, nous avons mis à disposition des fonctionnalités de communication pour intensifier les interactions en expression libre.

• **Hypothèse de recherche**

Notre hypothèse principale s'oriente vers un besoin de formation qui évolue, s'affine en interaction et en rétroaction entre les différents intervenants de la formation, soit le formateur, l'adulte-apprenant et l'organisme de formation. Toutefois, notre intuition nous oriente davantage vers une évolution limitée des besoins de formation que vers des changements profonds. En effet, l'adulte-apprenant souhaite clarifier, revoir ou approfondir certains aspects d'une thématique déjà connue. La formation permet de résoudre des incertitudes et de contextualiser ses besoins de formation en fonction de ses problématiques métier.

Notre approche d'utilisation des traces d'interactions est donc plus globale et superficielle : nous les considérons comme étant des indicateurs thématiques pour connaître les besoins de formation réels de l'adulte-apprenant tout en prenant en compte de leurs évolutions et leurs imprécisions.

Comme indiqué dans l'état de l'art, il peut s'avérer que cette information soit également biaisée par une perception erronée de l'adulte-apprenant.

Proposition

Notre proposition consiste à s'appuyer sur les profils d'apprenants pour personnaliser la composition de groupes de formation.

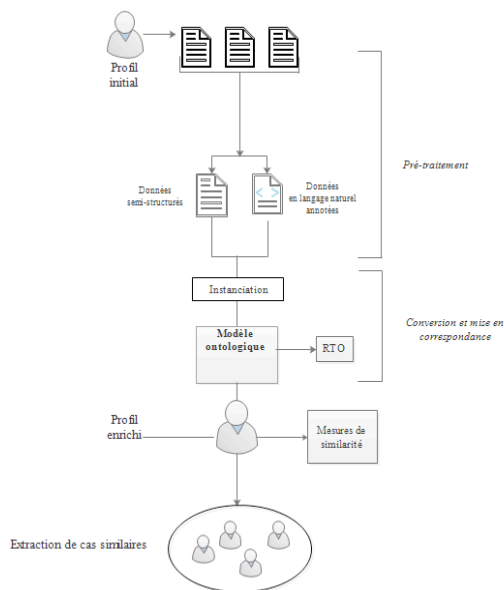


Figure 1-Schéma conceptuel simplifié

La construction d'un modèle va nous permettre d'exploiter l'ensemble de traces hétérogènes d'une manière cohérente. D'un point de vue technique, nous avons choisi une formalisation ontologique qui constitue le pivot central de notre chaîne de traitement présenté selon la figure 1.

• **Structure du profil apprenant**

L'évolution des profils s'appuiera sur l'ensemble des traces d'interactions et d'activités hétérogènes pour enrichir la structure initiale du profil de l'adulte-apprenant.

Le profil de l'adulte-apprenant comporte quatre sections. La figure ci-jointe propose une structure du modèle. Chacune des sections est subdivisée en sous-sections.

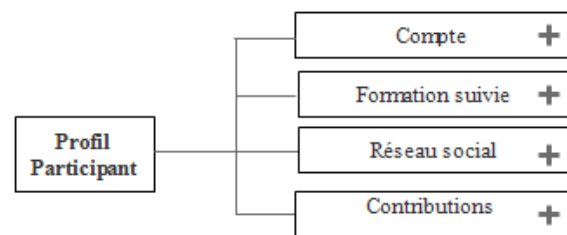


Figure 2-Structure globale du profil

La section « Mon compte » contient les données personnelles et professionnelles qui sont des sources d'informations stables. L'adulte-apprenant renseigne son métier, ses domaines d'expertises et ses centres d'intérêts de manière explicite. Ces champs sont obligatoires afin que nous puissions nous baser sur un profil initial complet.

Les sections « Formation suivie », « Réseau social » et « Contributions » du profil sont des sources d'informations évolutives. Elles seront enrichies par l'ensemble des traces d'activités, soit l'analyse de contenu des traces d'interactions et l'observation en situation de formation en présentiel.

L'évolution du profil de l'adulte-apprenant est gérée selon un processus incrémental basé sur l'addition de nouvelles informations dans la représentation structurel du profil.

• **Méthode de traitement**

La structure du profil de l'adulte-apprenant comporte des données structurées stockées dans une base de données relationnelle mais également non-structurées en format texte brut.

Les méthodes de traitement diffèrent selon la nature des données : structurées ou en langage naturel.

La particularité des traces d'interactions en langage naturel (mail avec le formateur, forum, commentaires, questions ouvertes des questionnaires) implique une approche à base de règles d'annotations. Nous supposons que les interactions écrites émises au

formateur en mode asynchrone seront les sources d'informations les plus pertinentes. Ces échanges avec le formateur s'effectuent dans un cadre confidentiel et limitent ainsi la menace du jugement professionnel.

Nous nous sommes orientés vers une méthodologie de catégorisation semi-supervisée de la nature des questions en utilisant des règles d'extraction. L'approche à base de règles nous semble plus appropriée que les méthodes à base d'apprentissage statistique, car la quantité de données disponibles dans notre corpus n'est pas suffisamment volumineuse pour que les algorithmes statistiques soient représentatifs.

Cette étape s'apparente à une analyse de contenu que nous avons choisie d'appuyer par une démarche linguistique computationnelle combinant une analyse sémantique et pragmatique. L'analyse pragmatique prend en compte le contexte énonciatif émis par l'adulte-apprenant et une analyse sémantique qui traite de l'hétérogénéité des vocabulaires utilisés par les différents intervenants. Les futures annotations générées seront les instances de peuplement de l'ontologie (voir le paragraphe suivant).

Ces transformations visent à représenter un profil avec les traces de manière unifiée pour simplifier la correspondance et respecter le formalisme ontologique cible.

- **Modèle ontologique**

La structuration des données est formalisée par un modèle ontologique qui constitue le pivot de notre chaîne de traitement. Le rôle principal de ce modèle est d'unifier et de représenter les informations collectées via l'import d'instances.

Il est guidé par une RTO⁶ (Ressource Terminologique) qui nous permet d'introduire le vocabulaire du domaine et un schéma d'annotation qui guide le processus d'instanciation des traces d'interactions.

- **Construction d'une RTO Métier**

Cette extension de l'ontologie a pour but de résoudre plusieurs difficultés d'ordre sémantique notamment liées à la divergence des expressions terminologiques des métiers des adultes-apprenants.

Nous nous sommes fondés sur le référentiel officiel de la CNFPT⁷, qui a fait l'objet d'une harmonisation des nomenclatures métiers de la fonction publique.

La méthodologie de construction d'une ontologie et plus particulièrement d'une RTO à partir de textes repose sur le postulat que les textes produits par des experts véhiculent les connaissances du domaine, et

⁶ Une RTO (ou TOR en anglais, pour Terminological or Ontological Resource) se définit comme une ontologie dans laquelle les termes sont rattachés aux concepts de façon systématique et exhaustive. Il y a plusieurs façons de rattacher des termes à des concepts selon ce qu'on veut pouvoir représenter (Reymonet, 2007)

⁷ Centre National de la Fonction Publique Territoriale

qu'il est alors possible d'en extraire suffisamment d'informations pertinentes pour construire une représentation conceptuelle. Celle-ci est plus ou moins isomorphe à la structure lexicale du domaine, en considérant qu'un terme est un concept lexicalisé et que la relation de subsomption se calque sur la relation d'hyponymie.⁸

L'instance principale est l'appellation la plus objective du métier. La synonymie des appellations métiers est principalement assurée par la fonction « same_individual_as » dans les tâches de raisonnement.

Pour être exhaustif dans le recouvrement terminologique des instances, nous y intégrons des instances supplémentaires non-officielles en faisant une extraction manuelle via les sites d'emplois généralistes.

Nous portons une attention particulière à la maintenance de cette portion de l'ontologie car nous considérons le métier comme la principale caractéristique discriminante pour les regroupements d'adultes-apprenants en présentiel. De plus, le métier induit le niveau hiérarchique de l'adulte-apprenant : le cadre statutaire de la fonction publique est fortement régi par l'appartenance du métier selon la catégorie (catégorie A, B, C). Cette catégorisation définit la délégation d'autorité et de responsabilité pour les adultes-apprenants titularisés.

Dans le cadre de notre ontologie, nous nous limiterons à l'appartenance d'un métier à une catégorie, ce concept étant suffisamment significatif pour ne pas se spécialiser aux concepts proposés par le référentiel officiel dont le grade et l'échelon. De plus, il est peu probable que l'adulte-apprenant utilise une description aussi finement détaillée dans son profil professionnel. Il s'agit également de ne pas négliger les adultes-apprenants ayant un statut contractuel de droit privé mais ayant un métier et un niveau hiérarchique équivalent à un adulte-apprenant titularisé.

Analyse et interprétation

A l'issue de la chaîne de traitement présenté dans la figure 1, nous obtenons en sortie un profil de l'adulte-apprenant enrichi.

Cette étape d'interprétation consiste à analyser les traces générées pour compléter la formulation du besoin initial. Celle-ci étant de nature incertaine, il s'agit d'analyser la progression thématique pour réajuster la formation accomplie. L'interprétation vise à réduire l'écart de perception entre le besoin ressenti de l'adulte-apprenant et la proposition de formation.

D'après les premiers tests effectués sur des profils fictifs, les résultats des requêtes donnent un ensemble de sous thématiques mais sans ordre de priorité. Pour optimiser ces résultats, il s'agirait d'attribuer un indice

⁸ Relation hiérarchisante d'une unité lexicale de sens plus général qui englobe une unité lexicale plus spécifique

de priorité selon les besoins de formation. Il est envisagé qu'il soit attribué de manière explicite par l'adulte-apprenant selon un degré de croyance calculé par le système grâce à des heuristiques.

Si l'on respecte les principes de construction de l'ontologie, les mesures se calculent en fonction de la longueur du chemin en considérant les concepts comme des sommets et les arêtes comme des relations sémantiques. L'inconvénient principal de cette mesure réside dans l'hypothèse forte selon laquelle toutes les relations de subsumption de l'ontologie interviennent dans la même proportion au cours du calcul de similarité. Cette contrainte peut être gérée en s'appuyant sur une mesure qui quantifie le degré de ressemblance essentielle entre deux concepts par une valeur numérique.

Un autre verrou apparaît sur la consistance de notre modèle ontologique. La nature des concepts des métiers de la RTO implique que nous pouvons conceptualiser l'ontologie de domaine de manière précise et certaine. Les concepts des besoins de formation sont par nature imprécis et incertains.

La prochaine étape de notre travail consistera à l'extraction de profils apprenants similaires pour le regroupement de personnes.

Nous considérons que la détection de centres d'intérêts de formation similaires est prioritaire pour l'assignation au sein d'un groupe de formation. Les caractéristiques identitaires professionnelles telles que le métier, le niveau hiérarchique et la zone géographique sont également des items déterminants.

Les caractéristiques identitaires personnelles (sexe, âge) sont considérées comme secondaires.

Conclusion et perspectives

Dans cet article, nous avons présenté une méthodologie d'approche pour enrichir le profil de l'adulte-apprenant à partir de données hétérogènes.

L'adulte-apprenant peut avoir une incertitude concernant ses besoins de formation, ce qui complexifie l'extraction des caractéristiques pertinentes et nécessite un processus de découverte dynamique de ses objectifs de formation.

Nous avons choisi d'axer l'enrichissement du profil principalement sur une caractérisation structurelle de l'adulte-apprenant, soit son métier combiné avec une caractérisation sémantique, soit son besoin de formation.

L'inconvénient principal de cette démarche est qu'elle implique un adulte-apprenant actif sur un temps restreint.

De nombreuses améliorations sont envisagées pour affiner le profil de l'adulte-apprenant.

En matière de perspectives, nous planifions d'effectuer des raisonnements qui doivent être capables d'exploiter l'ensemble des liens présents dans notre modèle. L'expression et l'exécution de règles sur les modèles OWL se fait principalement de deux manières : soit l'expression d'axiomes qui restreint un concept

automatiquement et ses instances associées, soit par l'écriture de règles SWRL⁹. Des premiers tests ont été effectués et semblent être plus souples pour agencer le raisonnement. Il s'agirait également d'utiliser une mesure sémantique pour pondérer les besoins de formation en fonction de son contexte temporel. En fin de parcours, le besoin de formation est davantage affiné et par conséquent la thématique est limitée dans une perspective de personnalisation de la formation.

Bibliographie

Actes de Conférences

Anaya, A. and Boticario, J.G. 2009. Clustering learners according to their collaboration', in Proceedings of the 13th International Conference on Computer Supported Cooperative Work in Design, IEEE Press, Los Alamitos, pp.540–545.

Harispe S, Sánchez D, Ranwez S., S. Janaqi and J. 2014. A framework for unifying ontology-based semantic similarity measures : A study in the biomedical domain. Journal of Biomedical Informatics, Volume 48, Elsevier, pp. Montmain.

Hogo, M.A. 2010. Evaluation of e-learners behaviour using different fuzzy clustering models: a comparative study', Int. Journal of Computer Science and Information Security, Vol. 7, No. 2, pp.131–140.

Isotani, S., Inaba, A., Ikeda, M. and Mizoguchi, R. 2009. An ontology engineering approach to the realization of theory-driven group formation', International Journal of Computer-Supported Collaborative Learning, Vol. 4, No. 4, pp.445–478.

Lallé S. ; Mostow J. ; Luengo V. ; Guin N. 2013. Comparaison de l'impact de techniques de diagnostic des connaissances sur l'apprentissage d'une stratégie d'aide. In Journée EIAH&IA 2013.

Pollalis, Y. and Mavrommatis, G. 2009. 'Using similarity measures for collaborating groups formation: a model for distance learning environments', European Journal of Operational Research, 1 March, Vol. 193, No. 2, pp.626–636.

Reffay C. ; Betbeder ML. 2009. Construction of a Mulce corpus from the VMT dataset Pre-conference Workshop on Common Objects for Productive Multivocality in Analysis. Computer Supported Collaborative Learning Rhodos, Greece.

⁹ Semantic Web Rule Language

Reymonet A. 2007. Modélisation de ressources termino-ontologiques en OWL. In F. TRICHET, Ed., Actes des 18es Journées Ingénierie des Connaissances, p. 169–180, Grenoble, France.

Zekri F. ; Turki E. ; Bouaziz R. 2015. AlzFuzzyOnto: Une ontologie floue pour l'aide à la décision dans le domaine de la maladie d'Alzheimer. In Inforsid, Biarritz, France.

Article de Revue

Bourgeois, E. 1991. L'analyse des besoins de formation dans les organisations : un modèle théorique et méthodologique, *Mesure et évaluation en éducation*, 14 (1)17-60.

Lapointe, J. 1993. L'analyse des besoins d'apprentissage-*Revue des sciences de l'éducation*, 9 (2), 251-266.

Michelet S. ; Luengo V. 2012. Prise en compte des contradictions intra-apprenant dans le diagnostic-Étude de cas : DiagElec un diagnostic informatique. *STICEF* (19), 1-29.