
Systemes à base de trace pour l'apprentissage humain¹

Lotfi Sofiane Settouti* — **Yannick Prié*** — **Alain Mille*** —
Jean-Charles Marty**

* *LIRIS - Bât Nautibus - UFR Informatique*
Université Claude Bernard Lyon 1 / F-69622 Villeurbanne Cedex
{LotfiSofiane.Settouti,Yannick.Prié,Alain.Mille}@liris.cnrs.fr

** *SYSKOM - Université de Savoie*
Campus Scientifique / 73376 Le Bourget-du-Lac cedex
Jean-Charles.MARTY@univ-savoie.fr

RÉSUMÉ. L'observation à base de traces dans les EIAH est au coeur des expérimentations et des études. Dans cet article nous proposons une définition de la trace d'apprentissage humain médié. Nous présentons aussi le système à base de traces, un système informatique permettant l'exploitation des traces. Pour cela, nous présentons un modèle formel et une architecture supportant ce système.

ABSTRACT. Trace-Based observation is the core of the experiments and studies in context of EIAH. In this article, we propose a definition of trace training. We present also the Trace-Based system, an information processing system allowing the exploitation of the traces. For that, we present a formal model and an architecture supporting this system.

MOTS-CLÉS : trace d'apprentissage, système à base de trace, Observation, scénario d'apprentissage

KEYWORDS: training trace, Trace-Based system, Observation, learning scenarios

1. Le travail de recherche que nous exposons dans cet article est réalisé dans le cadre du cluster *Informatique, Signal, logiciels embarqués* financé par la région Rhône-Alpes.

1. Introduction

La personnalisation des environnements informatiques pour l'apprentissage humain (EIAH) est l'un des principaux obstacles de leur développement. L'enjeu dans ce domaine est de comprendre le comportement de l'apprenant, ou d'un groupe d'apprenants, qui utilise un EIAH. Mais c'est, aussi, de pouvoir fournir à l'enseignant ou au formateur une information précise et adéquate pour ses besoins propres sur l'évolution individuelle et collective des apprentissages. Pour cela, la personnalisation doit prendre en compte outre les caractéristiques des utilisateurs (apprenants ou formateurs), celles des tâches ou des problèmes, ainsi que celles du contexte de leur traitement (fonctionnalités particulières et caractéristiques techniques de l'EIAH, caractéristiques sociales, institutionnelles ou économiques de la formation).

Le travail que nous allons présenter s'insère dans le cadre du projet *personnalisation des EIAH*, plus précisément, sur les observations à base de traces dans les EIAH. La résolution du problème de la personnalisation des EIAH est essentiellement dépendante de la capacité à produire des traces pertinentes et exploitables de l'activité individuelle ou collective de l'apprenant, qui interagit avec un EIAH. L'exploitation des traces réalisée par les enseignants, formateurs, ou apprenants, permet non seulement de produire des éléments intéressants pour la modélisation comportementale ou conceptuelle, mais aussi une personnalisation pertinente des EIAH par production de feedbacks ou évolution des interfaces. Il est nécessaire dans ce contexte d'aider les utilisateurs de la trace à collecter, transformer et analyser les traces issues des observations d'apprentissage humain.

Dans cet article, nous traitons la question des traces d'apprentissage en tant qu'objet d'investigation. Nous souhaitons définir un cadre théorique qui permet de rechercher des données dans les traces, de poser des requêtes, de faire des transformations, des visualisations. Pour cela, nous visons la définition d'une théorie de la trace dans le contexte d'activités d'apprentissage médiées par un environnement informatique. Il s'agit tout d'abord d'étudier la notion de *trace informatique* telle qu'elle est actuellement utilisée à des degrés divers et pour des motifs variés dans les activités d'apprentissage et de mettre en place des bases formelles supportant cette théorie. A cet effet, nous proposons un modèle formel lié à la trace informatique et une architecture des *Systèmes à Base de Traces (SBT)*. Nous définissons les systèmes à base de traces comme étant des systèmes informatiques permettant et facilitant l'exploitation des traces.

Dans la suite de cet article, nous présentons les différentes utilisations des traces dans les EIAH. Nous proposons ensuite une formalisation de ces systèmes à base de traces en définissant la notion de trace et en détaillant les différentes composantes de leur architecture. Un modèle formel supportant ces systèmes sera proposé. Finalement, les perspectives de ce travail sont données en conclusion.

2. Les traces et les EIAH

L'étude à base de traces¹, dans le cadre des EIAH, s'est imposée dans un mouvement de complexification des technologies supportant les EIAH et leurs usages. La question de la conceptualisation des traces d'activité à l'interface des EIAH, de leur représentation et de leur traitement, est loin d'être nouvelle. Le problème n'est plus seulement comment analyser les traces mais aussi comment vraiment compléter et exploiter les traces issues des observations faites pour améliorer l'apprentissage médié. Ce problème a été largement identifié et quelques travaux ont été proposés pour soutenir des observations dans le contexte des EIAH.

Une des nouvelles évolutions des EIAH est la possibilité de scénariser l'activité pédagogique de l'apprenant (Ferraris *et al.*, 2005). Ce progrès doit permettre à l'enseignant de définir des parcours destinés aux élèves, plus ou moins complexes et adaptés à ses objectifs pédagogiques. L'approche par les scénarios permet de bâtir des situations d'apprentissage personnalisées par les formateurs. Le scénario est l'expression de la tâche prescrite par l'enseignant alors que la trace d'exécution du scénario est l'expression de la tâche effective lors de l'utilisation par les acteurs. Une analyse comparant la tâche prescrite par le scénario et la tâche effective lors de l'utilisation par les apprenants participe à la régulation du processus d'apprentissage.

Dans le contexte des scénarios pédagogiques, les traces peuvent être utilisées pour atteindre des objectifs différents touchant aussi bien la régulation de l'activité d'apprentissage, la qualité du scénario pédagogique et la réingénierie des EIAH.

L'analyse des traces permet le contrôle et la régulation de l'activité d'apprentissage (Heraud *et al.*, 2005). Les traces permettent aux tuteurs et apprenants d'être aussi conscients que possible de l'activité d'apprentissage. L'observation apporte de nombreuses informations utiles à l'analyse d'une séance d'apprentissage, comme par exemple la compréhension de la réussite ou de l'échec des apprenants, pour que le tuteur et l'apprenant comprennent leurs activités. L'observation en se basant sur des traces est aussi un important facteur pour la qualité du scénario d'apprentissage (Marty *et al.*, 2004). Les traces sont utilisées pour améliorer le scénario pédagogique. Les traces d'apprentissage permettent l'évaluation du scénario pédagogique prédictif par rapport au scénario effectif extrait ou décrit par les traces d'usages. Donc, une correction ou amélioration du scénario sera possible après visualisation et analyse de ces traces.

Les expérimentations et les travaux sont nombreux à utiliser les traces pour la réingénierie des EIAH. L'étude des traces permet l'amélioration des scénarios pédagogiques du point de vue conceptuel (Barré *et al.*, 2005). Choquet et son équipe mettent en oeuvre la réingénierie du scénario pédagogique, en permettant de préconiser des observables à partir de la structure du scénario construit par l'équipe des concepteurs. Ensuite, ils proposent une façon de retranscrire ces observables dans un langage inter-

1. Le terme Trace est employé comme un concept et non dans son sens commun.

prétable en conception, dans le but de pouvoir comparer le scénario conçu a priori à celui résultant des usages observés *a posteriori*.

Beaucoup d'environnements supportant l'activité d'apprentissage utilisent les traces. Nous pouvons citer quelques systèmes d'apprentissage médié en montrant l'utilisation des traces qui en est faite. Ces systèmes permettent d'analyser nativement ou en externe les traces produites. Par exemple, le système Drew (Dialogical Reasoning Educational Web tool)(Séjourné *et al.*, 2004) qui est un EIAH collaboratif dont l'objectif est de favoriser un l'apprentissage collaboratif par l'argumentation ; le système Synergo (Avouris *et al.*, 2005) qui permet de mettre en oeuvre des situations de résolution de problèmes par un groupe d'étudiants distant ; ColAT (Collaboration Analysis Tool) (Avouris *et al.*, 2005) qui est un outil d'analyse de trace d'apprentissage indépendamment de tout système d'apprentissage ; APLUSIX (Nicaud, 1987) qui est un environnement d'aide à l'apprentissage de l'algèbre qui utilise la trace d'apprentissage pour aider les élèves à résoudre des exercices de calcul numérique.

Une remarque sur tous ces projets est que les traces obtenues passent par différentes étapes avant d'être analysées et exploitées. Il y a certains systèmes qui facilitent les traitements et les transformations pour aboutir à des traces plus pertinentes, qui aident à la découverte d'indicateurs et de motifs (*patterns*) d'utilisation et de conception. La tendance est de collecter des traces les plus complètes possibles. Les traces utilisées sont de plus en plus représentatives de l'activité de l'utilisateur. Cette tendance peut s'expliquer par (1) la généralisation de l'utilisation des TIC dans EIAH et de leur instrumentation et donc la facilité d'obtenir des traces plus ou moins pertinentes voulues par les concepteurs de l'étude et (2) l'utilisation de moyens d'observation très performants générant des traces plus complexes (comme la vidéo), relatant l'aspect contextuel qui peut être un apport important aux observations humaine et qui permet donc de compléter les premières traces. Le besoin de gérer et de traiter des traces de plus en plus complexes n'a cesse de s'accroître dans le cadre des EIAH. Les traces se présentent sous des formes de plus en plus variées (texte structuré, vidéo, etc.), incluant des relations riches (composition, liens hypertexte, dépendances temporelles, etc.) et des possibilités de navigation poussées. Cette évolution met clairement en évidence la nécessité d'aider et d'assister les utilisateurs de ces traces dans leurs activités de production, d'exploitation et d'analyse de ces traces.

3. Systèmes à base de Traces (SBT)

La plupart des travaux emploient une certaine notion de trace, et la trace a un rôle et une utilisation particulière dans chaque approche. Dans le contexte des EIAH, les traces sont des données issues d'observation directe ou indirecte permettant la régulation, le contrôle, l'analyse et la compréhension de l'activité d'apprentissage. De manière abstraite, nous définissons *la trace comme une séquence temporelle d'observés*. Le terme *séquence temporelle* dénote l'existence d'une relation d'ordre organisant les données de la trace par rapport à un repère de temps et le terme *observés* dénote que les données de la trace sont issues d'une observation. Cette définition permet de dire

que la trace informatique est un document numérique qui révèle des données temporellement situées résultant d'une activité d'observation.

Le système à base de traces s'inspire des travaux au laboratoire LIRIS sur l'approche MUSETTE (Champin *et al.*, 2004). Dans cette approche, la trace d'utilisation d'un système informatique est structurée en état-transition et est exploitée afin d'être réutilisée selon le principe du Raisonnement à Partir de Cas. L'approche MUSETTE permet de définir différents niveaux : la collecte des informations en format natif au travers de ce que est appelé des *sources de traçage*, la mise en forme des informations natives dans un langage balisé autorisant des exploitations variées sous la forme de traces qualifiées de *brutes*, la définition de *modèles d'utilisation* permettant de fournir un premier niveau de traces *lisibles* qualifiées de *primitives*, et enfin la description de *signatures de tâches* pour obtenir des épisodes réutilisables dans la trace. Les systèmes à base de trace découlent d'une approche plus générale que MUSETTE permettant de représenter un cadre pour l'exploitation des traces. Un système à base de traces (SBT) est un système informatique permettant et facilitant la manipulation des traces. La figure 1 montre l'architecture d'un tel système. Nous allons en définir les différents composants.

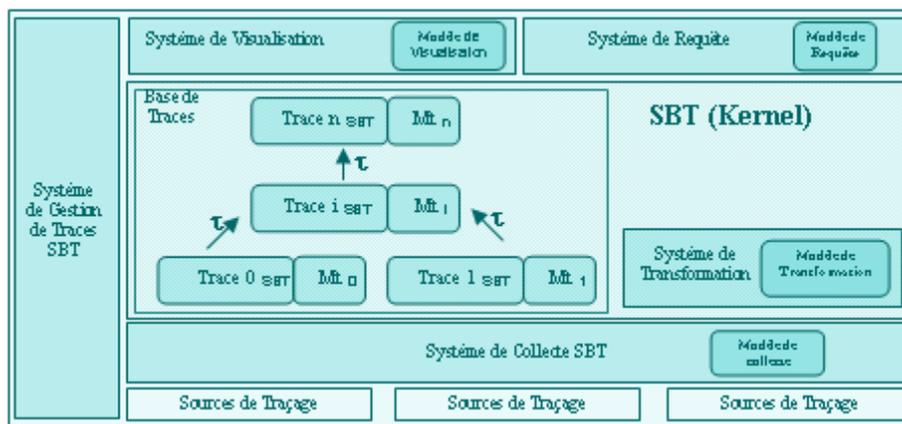


Figure 1. Architecture du Système à Base de Trace

Les *traces SBT* dans un SBT sont considérés comme une séquence temporelle d'observés et elles sont toujours accompagnées de leurs modèles de traces. Les traces les plus utilisées sont les fichiers Log mais de nombreux projets utilisent le format XML pour représenter les traces. La vidéo est aussi considérée comme une trace même si elle est utilisée la plupart du temps comme complément à d'autres traces étudiées. Le *modèle de trace* est un ensemble d'objets étiquetés représentant le vocabulaire de la trace. Ce modèle doit être explicite et permet d'expliquer la trace. Les modèles de trace font référence à d'autres modèles de traces. Par exemple dans le cadre des fichiers Log du serveur Apache, le modèle de trace est *Common Logfile Format* explicitant la sémantique de la trace. Dans Drew, le modèle de trace XML est une DTD. Dans le

cadre de l'approche MUsETTE (Olivier Georgeon, 2005), le modèle d'utilisation de la trace primitive est une ontologie exprimée en RDF structurant le vocabulaire de la trace.

Le *SBT Kernel* est le noyau du système. Il est constitué d'un système de transformation et d'une base de traces. La base de trace est l'ensemble des traces du SBT et de leurs modèles. Elle permet le stockage permanent et l'accès à tout moment aux traces. La base de traces peut être par exemple une collection de fichiers XML ou une base de données temporelle. L'exploitation de la trace consiste en partie en sa transformation. Le système de transformation s'occupe des différentes transformations τ requises lors de la modification et manipulation de la trace. Ce système permet de modifier la trace en enrichissant ou filtrant ses données, de modifier le modèle de trace, de mettre à jour la base de traces ou bien de faire des transformations automatiques en utilisant un modèle de transformation. Le modèle de transformation est un ensemble de règles formelles explicitées dans un certain langage.

Les traces peuvent être des sources d'information volumineuses d'où la nécessité de les interroger pour mieux les exploiter. Le système d'interrogation permet de poser des requêtes sur la base de traces. Ces requêtes peuvent concerner une ou plusieurs traces. Le système permet de partager des modèles de requêtes entre différents utilisateurs. Une *source de traçage* est un fichier ou un flux de données dans un format explicite quelconque. C'est une ressource disponible à tout moment dans le SBT. Dans notre approche, les sources de traçage deviennent des traces lorsqu'on leur associe un modèle de trace.

Le système de collecte est un ensemble structuré de processus permettant de convertir les données en traces en utilisant les outils adéquats. Le système de collecte permet de sélectionner des sources de traçage. Les systèmes de collecte peuvent recourir à des techniques de synchronisation des bases de temps des différentes sources. Ils peuvent recourir aussi à méthodes d'intégration de données pour agréger les différentes sources. Une propriété intrinsèque du processus de collecte est qu'il itératif. En effet, la trace collectée est souvent obtenue après plusieurs retours aux sources de traçage afin de l'améliorer.

Le système de visualisation permet de visualiser les traces et facilite donc leur analyse et interprétation. Le système de visualisation doit permettre de jouer certaines traces en accédant aux sources de traçage et aux données relatives à ces traces. Ce système doit gérer et accéder aux différentes applications associées aux sources de traçage comme les vidéos. Il donne l'accès direct depuis la trace à la source de traçage pour interroger celle-ci plus directement. Par exemple, exploiter la vidéo comme trace temporalisée liée par le temps avec une trace du SBT. Le système de visualisation peut être intégré dans un système de transformation à un niveau quelconque. Le système de visualisation doit être capable de visualiser les traces issues de transformations.

Enfin, le système de gestion des traces concerne la gestion des modèles (modèles de trace, de transformation, de requête). Il permet l'ajout, la suppression des traces. Il regroupe les processus de conservation des traces et des droits d'administration de ces

traces. Il permet de gérer le graphe de transformation des traces permettant de relier l'évolution des traces et leur cycle de vie. Il peut être considéré comme un système assurant la traçabilité des transformations effectuées sur les traces.

4. Un modèle formel simplifié pour les Systèmes à Base de Traces

Dans cette section, nous allons définir un cadre formel (Framework) pour des systèmes à base de traces. Ce cadre fournit une base simplifiée d'un SBT où les composantes sont la Trace, Modèle de trace. Des situations plus complexes peuvent être modélisées en exprimant des extensions de ce cadre afin de représenter, par exemple, le système de transformation et les transformations, le langage de requêtes. Il est important pour notre système de mettre en place les bonnes bases formelles pour définir formellement par la suite les autres composantes du SBT en ce qui concerne les transformations, les requêtes, les visualisations et les analyses.

Nous allons présenter un début de formalisation de la notion de trace et modèle de trace et graphe des transformations. Afin de présenter le temps dans ce cadre formel, nous devons faire au moins quelques choix de base sur les modèles de temps. Les deux solutions traditionnelles sont les modèles de temps basés sur les instants et les modèles de temps basés sur les intervalles. Dans les modèles basés sur les instants, les entités temporelles primitives sont des instants. Ainsi, l'écoulement du temps est représenté comme ensemble d'instant avec une relation d'ordre binaire de précédence ($T, <$), l'égalité est implicitement assumée. Dans ce cadre formel, nous travaillerons dans le cas le plus traditionnel et le plus simple des modèles. Les modèles de temps que nous utiliserons sont basés sur les instants, mais il est tout à fait possible de considérer le modèle basé sur les intervalles dans un travail ultérieur. Nous présentons le Modèles de trace, Trace, Graphe des transformations.

– **Définition 1 :** Un *modèle de Trace* est une structure $\Theta = (\Theta_c, \Theta_r)$ où Θ_c est un ensemble fini de concepts θ_{ci} , $\Theta_c = \{\theta_{c0}, \theta_{c1}, \dots, \theta_{cn}\}$ et Θ_r est un ensemble fini de concepts relations θ_{ri} , $\Theta_r = \{\theta_{r0}, \theta_{r1}, \dots, \theta_{rn}\}$. Le modèle de la trace est une structure à deux ensembles de concepts θ_{ci} et de concepts relations θ_{ri} .

– **Définition 2 :** Un *domaine temporel* est une structure $D_p = (T, \leq)$ ordonnée linéairement où T est l'ensemble fini des instants de temps et \leq est une relation d'ordre linéaire dans T .

– **Définition 3 :** Une *Trace* est un quintuplé $Trace = (D_p, O_{tr}, \Theta_t, R_t, R_s)$ où D_p est un domaine temporel O_{tr} est un ensemble fini des éléments O de la Trace, $O_{tr} = \{o_0, o_1, \dots, o_n\}$ Tel que $\forall O_i \in O_{tr}, f(O_i) \in \theta_{ci}$, f est une fonction d'étiquetage $f : O_{tr} \rightarrow \theta_{ci}$. Θ_t est un modèle de trace associé à la trace R_t est une relation représentant les liens temporels $D_t \times O_{tr}$ tel que $R_t \subseteq D_t \times O_{tr}$. et $\forall R_{ti} \in R_t, h(R_{ti}) \in \theta_{ri}$, h est une fonction d'étiquetage $h : R_t \rightarrow \theta_{ri}$. R_s est une relation représentant les liens structurels $O_{tr} \times O_{tr}$ tel que $R_s \subseteq O_{tr} \times O_{tr}$. et $\forall R_{si} \in R_s, g(R_{si}) \in \theta_{ri}$, g est une fonction d'étiquetage $g : R_s \rightarrow \theta_{ri}$.

La trace est une structure où les objets de la trace O_{tr} sont liés au domaine temporel D_t par R_t , les objets de la trace sont liés entre eux par une relation de structure R_s et les objets de la trace, les relations temporelles et structurelles sont liées aux modèles de trace par des fonctions d'étiquetages.

– **Définition 4 :** *Graphe de transformation* $G_f = (Traces, Transformations)$ où Traces est un ensemble de noeud de type Trace et Transformations est un ensemble de relations de Traces Traces.

Le modèle formel qu'on présente se veut un modèle simple et générale. Par exemple, nous ne définissons pas les liens dans le modèle de trace Θ , et les liens de objets de la traces O_{tr} avec ce modèle. Cependant, une spécialisation précisant cette structure permettra une meilleure modélisation et une meilleure extension de ce modèle.

5. Conclusion

Nous avons défini dans cet article le concept de trace et de systèmes à base de trace. Ce système facilite et prend en charge complètement l'exploitation des traces. Un début de proposition de modèle formel à été présenté pour supporter les SBT. Dans le cadre des EIAH, les applications possibles sont nombreuses concernant l'amélioration de l'apprentissage et du scénario pédagogique et les EIAH.

Les perspectives de ce travail concernent la formalisation complète des différents composants et services du SBT. Notamment, la formalisation d'un langage de requêtes en se basant sur les modèles de trace. Ensuite, une implémentation de ce système permettra une application concrète dans le contexte des EIAH.

6. Bibliographie

- Avouris N., Komis V., Margaritis G. F. M., Voyiatzaki E., « Logging of fingertip actions is not enough for analysis of learning activities », *Workshop AIED'05, Amsterdam*. 1-8, 2005.
- Barré V., El-Kechai H., Choquet C., « Re-engineering of collaborative e-learning systems : evaluation of system, collaboration and acquired knowledge qualities », *Workshop AIED'05, Amsterdam*. 9-16, July, 2005.
- Champin P.-A., Prié Y., Mille A., « MUSETTE : a Framework for Knowledge Capture from Experience », *EGC'04, Clermont Ferrand*, Jan, 2004.
- Ferraris C., ans L Vignollet A. L., David J., « Modélisation de scénarios d'apprentissage collaboratif pour la classe », *Environnements Informatiques pour l'Apprentissage Humain EIAH'05, Montpellier*, 26 et 27 Mai, 2005.
- Heraud J.-M., Marty J.-C., France L., Carron T., « Helping the Interpretation of Web Logs : Application to Learning Scenario Improvement », *Workshop AIED'05, Amsterdam*, 18-22 July, 2005.

- Marty J.-C., Héraud J.-M., Carron T., France L., « A quality approach for collaborative learning scenarios », *Learning Technology Newsletter of IEEE Computer Society*, Vol. 6, Issue 4p. 46-48, 14 -16 December, 2004.
- Nicaud J., Aplusix : un système expert en résolution pédagogique d'exercices d'algèbre, PhD thesis, Thèse de l'Université Paris XI-Orsay, 1987.
- Olivier Georgeon theiry Bellet A. M. D. L. R. M., « Driver behaviour modelling and cognitive engineering tools development in order to assess driver situation awareness », *Workshop on Modelling Driver Behaviour in Automotive Environments. 25-27 May 2005, Ispra*, 2005.
- Séjourné A., Baker M., Lund K., Molinari G., « Schématisation argumentative et co-élaboration de connaissances : le cas des interactions médiatisées par ordinateur », *In E.A. Théodile Lille 3 (Ed.), Actes du colloque international "Faut-il parler pour apprendre ?" (Arras)*, 2004.

ANNEXE POUR LE SERVICE FABRICATION
A FOURNIR PAR LES AUTEURS AVEC UN EXEMPLAIRE PAPIER
DE LEUR ARTICLE ET LE COPYRIGHT SIGNE PAR COURRIER
LE FICHER PDF CORRESPONDANT SERA ENVOYE PAR E-MAIL

1. ARTICLE POUR LA REVUE :

L'objet. Volume 8 – n°2/2005

2. AUTEURS :

Lotfi Sofiane Settouti — Yannick Prié* — Alain Mille* —
Jean-Charles Marty***

3. TITRE DE L'ARTICLE :

Systèmes à base de trace pour l'apprentissage humain²

4. TITRE ABRÉGÉ POUR LE HAUT DE PAGE MOINS DE 40 SIGNES :

Système à Base de Trace

5. DATE DE CETTE VERSION :

20 février 2006

6. COORDONNÉES DES AUTEURS :

– adresse postale :

* LIRIS - Bât Nautibus - UFR Informatique
Université Claude Bernard Lyon 1 / F-69622 Villeurbanne Cedex
{LotfiSofiane.Settouti,Yannick.Prié,Alain.Mille}@liris.cnrs.fr

** SYSCOM - Université de Savoie
Campus Scientifique / 73376 Le Bourget-du-Lac cedex
Jean-Charles.MARTY@univ-savoie.fr

– téléphone : 00 00 00 00 00

– télécopie : 00 00 00 00 00

– e-mail : Roger.Rousseau@unice.fr

7. LOGICIEL UTILISÉ POUR LA PRÉPARATION DE CET ARTICLE :

\LaTeX , avec le fichier de style `article-hermes.cls`,
version 1.2 du 03/03/2005.

8. FORMULAIRE DE COPYRIGHT :

Retourner le formulaire de copyright signé par les auteurs, téléchargé sur :
<http://www.revuesonline.com>

SERVICE ÉDITORIAL – HERMES-LAVOISIER
14 rue de Provigny, F-94236 Cachan cedex
Tél : 01-47-40-67-67
E-mail : revues@lavoisier.fr
Serveur web : <http://www.revuesonline.com>