

Un environnement pour gérer des traces comme inscriptions de connaissances

Rapport de recherche

J. Laflaquière^{2, 1}, L.S. Settouti¹, Y. Prié¹, A. Mille¹

¹LIRIS - Bât Nautibus - UFR Informatique
Université Claude Bernard Lyon, F-69622 Villeurbanne,
{LotfiSofiane.Settouti, yprie, amille}@liris.cnrs.fr

²ICD - Tech-cico, UTT
10 rue Marie Curie, F-10000 Troyes
julien.laflaquiere@utt.fr

Résumé

Nous nous intéressons dans cet article aux traces numériques d'utilisation de systèmes informatiques, que nous proposons de penser comme des inscriptions des connaissances dans le cadre du « travail intellectuel ». Considérées en tant qu'inscriptions de connaissances, explicitement modélisées, les traces jouent d'une part un rôle similaire à celui qu'elles jouent dans le monde physique, d'autre part il est possible de proposer une relecture des travaux de la littérature sur les outils utilisant des traces. La deuxième partie de l'article présente d'abord le cadre conceptuel des systèmes à base de traces modélisées (SBTm) offrant des services de collecte, de transformation et de visualisation de traces puis notre implémentation de SBT (ATER) qui permet de gérer des traces modélisées en RDF/OWL ; enfin nous discutons l'ingénierie de trace comme ingénierie des connaissances.

Mots Clef

Traces, Inscription de connaissances, Modélisation de Traces, Systèmes à Base de Traces, Ingénierie de la trace

Abstract

In this article we consider the traces that can be collected from the use of a digital environment during "knowledge work" as knowledge inscriptions issued from this activity. Considered as such and explicitly modelled, traces can play a similar role as the one they play in the physical world. Literature can also be reviewed within that framework. The second part of the article firstly presents our conceptual framework called "modelled Trace-Based Systems" (mTBS), that offer collect, transforming and visualisation services. Secondly, our implementation of mTBS, that uses RDF/OWL modelled traces is presented. Lastly, we discuss the topic of trace engineering as knowledge engineering.

Keywords

Traces, Knowledge inscription, trace modelling, Trace-Based Systemes, Trace engineering

L'utilisation du terme de « trace » dans le langage courant peut donner l'illusion d'une simplicité de la notion qu'elle

ne possède pourtant pas. Comparable dans cette fausse simplicité à d'autres notions, telle que celle de « forme » par exemple, la notion de trace n'a que rarement été prise pour objet de recherche [1].

Signifiant à l'origine « empreinte ou la suite d'empreintes laissées par le passage d'un homme ou d'un animal », le terme possède aujourd'hui différentes facettes, dont nous ne retiendrons pour le travail que nous présentons ici, que celle d'« empreinte, de marque laissée par une action » [1,2]. L'action qui nous intéresse ici, en l'occurrence les actions, sont celles qu'implique un type particulier d'activité que nous appellerons « travail intellectuel ». Nous définissons ce dernier comme une « activité cognitive instrumentée d'interprétation et de manipulation de documents, de ressources, ou plus généralement d'inscriptions de connaissances » [3].

Les conditions d'interprétation des documents, des ressources, sont très différentes suivant que l'on se place dans un espace matériel ou dans un espace virtuel numérique. Si certains aspects liés aux traces numériques ont été mis en évidence et traités depuis le début de la construction de systèmes informatiques, il nous semble que la question des traces d'utilisation dans un environnement numérique n'a trouvé jusqu'à présent que des réponses partielles, tant au plan conceptuel que technique. Ce que nous semble montrer la grande hétérogénéité des travaux sur les traces numériques.

Cet article aura donc pour premier but, de reformuler la question des traces numériques, pour tenter d'y répondre de manière générique. Une comparaison entre les traces d'utilisation dans un environnement matériel et dans un environnement virtuel nous conduira à présenter les traces elles-mêmes d'une part comme nécessairement modélisées, d'autre part comme des inscriptions de connaissances. Ce point de vue nous offrira une grille de lecture originale pour relire les principales approches exploitant des traces numériques dans la littérature. Nous défendrons ensuite, dans la deuxième grande partie de l'article les principes de notre propre approche des Systèmes à Base de Traces modélisées, dont nous présenterons les principaux éléments, la plateforme en cours de développement, ainsi que les utilisations applicatives en cours.

Nous serons conduits dans cette deuxième partie à utiliser plus spécialement une application particulière pour illustrer nos travaux. Nous avons ainsi choisi une situation mettant en jeu le travail intellectuel tel que défini plus haut : il s'agit d'un environnement de « classe virtuelle », utilisé par la société *eLycée campus*^{TM1} pour proposer des cours de français à des élèves anglophones scolarisés à l'étranger. Nous reviendrons lorsque cela sera nécessaire sur les détails de la plateforme proposée, mais il est important de souligner dès maintenant que toutes les activités (pédagogiques en l'occurrence) proposées aux utilisateurs passent par la manipulation complexe de ressources documentaires.

La première partie de l'article débutera donc par un retour sur la notion de trace, dans le cadre d'un travail intellectuel, d'une activité cognitive instrumentée.

2 Traces d'activités cognitives instrumentées

Nous présenterons progressivement dans cette section les notions de *trace matérielle* d'une activité cognitive instrumentée, puis de *trace numérique* pour enfin aboutir à la notion de *trace modélisée*.

2.1 Traces matérielles

La question qui nous intéresse tout d'abord est celle des traces d'un travail intellectuel se déroulant dans un environnement matériel. Dans un tel environnement, l'existence d'un document repose sur la nature matérielle de son support. La *matérialité* des supports d'inscriptions conditionne d'ailleurs fortement les moyens d'interagir avec les documents. Cette matérialité autorise toutefois, ce qui nous intéresse plus particulièrement ici, *l'existence de traces d'utilisation*.

Le travail intellectuel tel que nous l'avons défini dans l'introduction *s'appuie* sur l'utilisation de l'environnement physique, cette *utilisation* étant entendue comme l'ensemble des manipulations impliquées par l'exploitation des documents². Ces manipulations sont susceptibles de laisser des traces dans l'environnement et particulièrement sur les documents. Pour être plus précis ces traces peuvent être propres à un document (une page cornée, annotée ou abîmée), ou bien se rapportant à la relation de celui-ci au reste de l'environnement (un livre posé sur le bureau pour être étudié dans la journée par exemple). La matérialité de son support permet à un document de se donner à voir, à mémoriser, visuellement. Physiquement, la trace mémorielle [1] résultant de sa manipulation est peut-être la première des traces d'utilisation.

Les documents se trouvent donc « marqués » par leur utilisation, et ce de façon *continue* et *permanente*. Continue et permanente, mais non « pérenne », parce

qu'ils sont toujours susceptibles d'être marqués de nouveau et ne sont jamais totalement vierges de traces d'utilisation. Tout travail intellectuel, d'exploitation de document, et notamment toute interprétation de ces documents, est donc immergé dans une matérialité dont il ne peut se détacher : tout document, toute inscription, est interprétée *en tenant compte de ces traces*. Autrement dit, les traces d'utilisations jouent un rôle dans la construction du sens et dans l'appropriation du document [4].

Ces remarques peuvent faire penser, bien entendu, au marquage des documents par l'acteur lui-même, i.e. à l'annotation, à la lecture active [5], *etc.* Elles ne s'y limitent cependant pas. La condition matérielle dont nous venons de parler implique également que toute manipulation engendre et s'appuie sur une *organisation spatiale* et *temporaire* des éléments de l'environnement. La répartition des documents sur un bureau est nécessairement le résultat d'une manipulation, et elle participe pleinement à l'activité cognitive sous-tendant leur interprétation [6]. Cette organisation, cette *mise en relation* spatiale, qui peut se voir comme un marquage « passif »³ et temporaire des différents artefacts mobilisés par l'activité, relève également pour nous d'une trace d'utilisation.

Transposons maintenant cette situation dans un environnement numérique, dans lequel la matérialité des supports a justement disparu, pour nous demander ce que deviennent les traces que nous venons d'évoquer.

2.2 Traces numériques

Dans un environnement numérique, la dématérialisation des supports d'inscription annule certaines des contraintes physiques que nous venons d'évoquer. La disparition de ces contraintes fait également disparaître les conditions d'émergence et d'existence des traces d'utilisation telles que nous venons de les décrire.

La conséquence de ce changement est double. Premièrement, s'agissant de l'utilisation de l'organisation spatiale des documents : il est toujours possible de défendre l'idée du maintien d'une spatialité dans un environnement numérique. Cependant cette spatialité est limitée aux possibilités offertes par une interface Homme-Machine, c'est-à-dire celle de l'écran et du clavier⁴. La métaphore du « bureau », commune aux grands systèmes d'exploitation, reste d'ailleurs limitée pour les mêmes raisons. Nous verrons dans la partie 2.4 que si la dimension *spatiale* de l'organisation des documents est fortement amoindrie, la dimension *temporelle* de celle-ci gagnera le premier plan.

Deuxièmement, s'agissant cette fois du « marquage » par l'activité dans le cadre de leur manipulation. On peut

¹ <http://elycee.net/>

² Exploitation que l'on peut décrire à l'instar de [4] comme une *objectivation* et une *appropriation*.

³ Au sens où il n'est pas nécessairement intentionnel.

⁴ Nous ne prenons pas en compte ici les environnements de réalité virtuelle.

remarquer que certains documents numériques⁵ possèdent des (méta)données, dates de création ou de modification par exemple, qui pourraient être considérées comme des traces numériques d'utilisation. Ces données cependant sont loin de représenter la dynamique d'utilisation des inscriptions dans une activité donnée, et sont surtout liées au fonctionnement même des systèmes informatiques⁶, et surtout n'ont pas été conçues *en tant que traces d'utilisation*, i.e. en tant que « marquage » par l'activité.

Ce qui nous conduit à faire le constat suivant : pour obtenir des traces numériques d'utilisation, nous sommes condamnés à *créer de toutes pièces* leurs conditions d'existence au sein de l'environnement numérique qui soutient l'activité. Cela signifie que l'environnement doit être *instrumenté* pour capter dans l'utilisation de l'environnement, des traces de la manipulation et de l'exploitation des documents.

Le terme de *trace d'utilisation* faisant pour nous référence à l'idée d'un « marquage » par l'activité.», nous proposons de définir les traces numériques, de manière générale, comme tout enregistrement automatique temporalisé et exploitable, dont les constituants sont liés à la manipulation de l'environnement numérique impliqué. Cette définition globale étant posée, nous pouvons chercher à déterminer comment ces traces numériques peuvent être obtenues.

2.3 Traces numériques modélisées

Il existe deux approches générale permettant de récupérer des traces numériques :

La première approche consiste à considérer que le fonctionnement logiciel de l'environnement informatique fournit « naturellement » (par construction) un enregistrement des interactions, qui peut être considéré comme trace d'utilisation d'un environnement dans le cadre d'une activité donnée. Dans ce cas, la démarche qui consiste à considérer les traces numériques d'utilisation de documents est une démarche d'utilisation des traces *a posteriori*, qui s'appuie sur une instrumentation native des systèmes, qui n'ont pas été conçus pour cela.

La seconde approche consiste à instrumenter spécifiquement l'observation des interactions pour récupérer une trace dont on aura défini *a priori* les constituants. La démarche est alors moins opportuniste et nécessite un travail logiciel non négligeable.

Dans une démarche d'ingénierie des connaissances, nous nous intéressons plus particulièrement à la mise en place *explicite* d'une instrumentation de l'observation. En effet,

⁵ Nous adopterons un point de vue général sur la notion de document numérique sans rentrer dans les débats menés au sein du RTP Document [6],

⁶ La date de modification permettant par exemple de régler automatiquement un conflit de version entre deux fichiers du même nom.

le fait d'instrumenter un environnement dans le dessein d'en tirer des traces d'utilisation implique que soit mise en place une observation de l'interaction, de l'activité, dont certains éléments, considérés comme pertinents seront tracés. Ces éléments (appelés *observés*) sont les constituants de la trace numérique générée par l'activité réalisée dans l'environnement par l'utilisateur. Le point important ici est que, du fait de l'instrumentation explicite de l'environnement, ces observés sont obligatoirement modélisés en tant que tels. C'est pourquoi on parlera de *traces numériques modélisées*.

2.4 Traces et inscriptions de connaissances

Le fait d'introduire la notion de *modélisation* de la trace nous permet de reconsidérer ce que des traces numériques peuvent apporter au travail intellectuel dans un environnement numérique. Nous allons voir, dans le prolongement de ce qui a été amorcé dans la partie 2.2, que les possibilités nouvelles offertes par le numérique nous laissent envisager les traces numériques comme de véritables inscriptions de connaissances.

Considérons, de manière générique, une trace numérique fruit de l'enregistrement automatique d'éléments d'interaction d'un utilisateur avec son environnement, permettant de donner une information sur l'utilisation de ce dernier dans le cadre d'un travail intellectuel. Il est tout à fait possible de considérer cette trace également comme une *organisation* des éléments qui la composent. Ces éléments d'interaction permettant de traduire la manipulation effective des documents, on peut alors voir la trace comme une organisation *par leur manipulation elle-même* des documents manipulés.

Bien sûr cette organisation diffère franchement de celle que nous avons précédemment évoquée (partie 2.2). Elle constitue bien une *mise en relation* des éléments manipulés dans leur environnement relativement à leur manipulation dans le cours de l'activité. Cette fois-ci cependant cette mise en relation est avant tout *temporelle*, plus que spatiale (cf. partie 2.2), mais elle peut aller bien au-delà.

D'une part, du fait de la séquentialité/linéarité des interactions avec un environnement numérique, chaque manipulation, chaque action de l'utilisateur est située dans le temps. Un enregistrement automatique de ces interactions aboutit donc nécessairement à une trace numérique, organisation des observés sur une base *temporelle*.

D'autre part la particularité de l'environnement numérique permet d'envisager d'autres mises en relations dans cette même trace. Prenons l'exemple d'un navigateur *Web*. La majorité des navigateurs *Web* possèdent un historique, permettant (au minimum) de revenir en arrière. Deux sites visités successivement se suivront dans la trace (ou bien des informations enregistrées permettront d'établir que deux affichages d'url se sont succédés).

Ces deux sites seront donc liés dans la trace d'un point de vue temporel et séquentiel. Mais ces mêmes pages peuvent également être mises en relation dans la trace en suivant d'autres critères : celui de la structure de la navigation, celui de l'auteur de la première visite, celui d'être des résultats issus de la même requête, etc.

Finalement, une trace numérique comme *organisation* de ses éléments constitutifs, et par conséquent des documents manipulés, a toutes les raisons de jouer un rôle similaire à celui que nous avons attribué aux traces matérielles. En effet, et à partir du moment où ces traces numériques sont exploitables par un utilisateur, celles-ci vont conditionner, en l'enrichissant, l'interprétation des documents.

En quelque sorte, le numérique amoindrit fortement la possibilité d'un marquage « actif » des documents, tout en démultipliant les possibilités d'un marquage « passif » de ces derniers. Ces possibilités s'articulent autour de la structuration qu'offre l'activité elle-même, ici le travail intellectuel, de sorte que la trace se trouve porteuse de sens et capable d'enrichir les conditions d'interprétation d'un document. C'est pourquoi nous revendiquons aujourd'hui pour ces traces le statut d'*inscriptions de connaissances* à part entière [7].

En effet, bien qu'une trace numérique modélisée soit une inscription *automatique*, elle possède le sens que lui confère l'utilisation de l'environnement au regard de l'activité menée par l'utilisateur. L'interprétation d'une trace, ou d'un observé relativement à la trace est une connaissance : elle a un impact sur l'action possible sur un document, ne serait-ce qu'une action d'interprétation. Cette proposition est intéressante parce qu'elle nous permet de considérer les traces numériques de manière générique, et nous guide vers dans le développement nécessaire d'une instrumentation des traces, notamment parce qu'elle fait dépendre la sémantique de la trace, du caractère plus ou moins explicite de la modélisation de celle-ci.

2.5 Tour d'horizon des systèmes utilisant des traces numériques

Le point de vue sur les traces d'utilisation que nous venons d'exprimer nous offre une grille de lecture originale pour aborder dans cette section les différents (et nombreux) travaux utilisant des traces numériques, présents dans la littérature. Dans cette courte revue, nous insisterons en particulier sur les différences marquées entre les approches concernant la modélisation plus ou moins explicite des traces utilisées. Une analyse succincte nous permettra ensuite d'introduire notre propre approche.

Remarquons tout d'abord qu'il est difficile d'établir une typologie claire des nombreuses approches utilisant des traces numériques, notamment parce que le nombre de critères discriminants est très (voire trop) élevé. Ceci est à mettre en rapport avec les nombreuses finalités possibles pour des systèmes utilisant des traces : de l'assistance à

l'utilisateur, à la création de scénarii, en passant par des applications destinées à la gestion des connaissances [8].

Pour pouvoir présenter rapidement un tour d'horizon global, nous prendrons le risque de caricaturer une classification en proposant de regrouper les différents travaux suivant de deux « tendances » distinctes.

La première tendance regroupe des approches plutôt quantitatives, s'appuyant sur des méthodes statistiques, généralement tournées vers *l'analyse*. Si nous devons choisir un exemple représentatif de ce type d'approche, nous pourrions faire référence aux travaux de *Web-mining* ou *Web usage mining*. Globalement ces travaux adoptent une démarche qui consiste à se baser sur des *logs* nécessaires au fonctionnement de la navigation Web. L'ensemble des *logs* ainsi collectés est ensuite traité statistiquement pour tenter d'en dégager des informations remarquables. Généralement ces approches sont plutôt tournées vers l'analyse de l'activité. Sont ainsi construits des assistants à la navigation [9] ou des systèmes de typage des parcours sur le *Web* [10]. Il s'agit également de soutenir des personnalisations d'interface ou des modélisations de l'utilisateur [11], ou encore des fournir des moyens d'évaluation de l'utilité et de l'utilisabilité d'un système [12, 13]. Dans ces approches, la modélisation des traces reste *implicite* dans le sens où celles-ci ne dépendent que du modèle de conception des systèmes et non de l'utilisation effective qui en est faite [14].

La seconde tendance regroupe des approches dont les traces sont modélisées et définies de manière plus explicite. Si nous devons prendre un exemple représentatif nous pourrions proposer cette fois les approches consistant à instrumenter des EIAH dans l'objectif d'en rendre l'utilisation plus réflexive. Contrairement aux approches précédentes, la trace n'est alors pas destinée à une analyse externe à l'activité elle-même, mais à l'utilisateur dans le cadre de son activité [15, 16, 17].

Le fait que des traces soient destinées à être exploitées par l'utilisateur lui-même nécessite que celles-ci soient appréhendables facilement, et explique le fait qu'elles aient fait l'objet d'une modélisation, ou tout au moins d'un choix motivé. En effet, le modèle de la trace, s'il apparaît plus explicite, parce que plus proche de l'activité menée en termes de niveau d'abstraction, n'en est pas pour autant clairement *explicité formellement* par les auteurs. Ainsi, les outils que proposent ce genre d'approches sont généralement d'abord conçus comme des solutions pratiques à un besoin utilisateur identifié. La conséquence est que ces solutions sont généralement développées de façon *ad hoc*, sans réflexion globale sur le statut des traces numériques.

Après ce rapide tour d'horizon des approches employant des traces numériques, nous pouvons remarquer que celles-ci sont hétérogènes, tant sur le plan théorique et conceptuel que technique et qu'aucune d'entre elles ne prend le recul nécessaire à la prise en compte globale des traces. Cela est lié au fait que la modélisation des traces reste souvent un aspect secondaire, alors même qu'elle représente souvent la pierre d'angle de ces travaux. C'est

pourquoi nous avons développé et présentons dans la deuxième partie de cet article un *cadre conceptuel générique* pour penser les traces, que nous nommons *Système à Base de Traces modélisées* [19].

3. Systèmes à Base de Traces modélisées

L'approche globale des systèmes à base de traces modélisées (SBTm en abrégé) repose fondamentalement sur le concept de trace modélisée, ou M-trace. Le principe général d'une trace modélisée ayant été rapidement abordé dans la section précédente, et nous allons cette fois en proposer un développement rigoureux.

3.1 Les M-traces

Nous pouvons commencer par poser les définitions suivantes [19] :

On appelle *observé* tout objet informatique décrivant un élément issu de l'observation de l'activité d'utilisation d'un environnement informatique. En tant que résultat de l'observation de cette activité / processus se déroulant dans le temps, une trace sera constituée d'observés temporellement situés. On décrira par exemple les actions effectuées et le moment de leur effectuation (ouvrir un document, modifier un paragraphe, utiliser un lien hypertexte, etc.), auquel cas, on considérera des observés liés à des périodes de temps, par exemple qu'un mail a été rédigé pendant cinq minutes, ou que l'utilisation d'une application a duré plusieurs heures.

On appelle relation entre observés toute relation permettant d'exprimer des liens entre observés autres que temporels. Par exemple, que telle action a conduit à la modification de tel document, que tel document manipulé tel jour est bien le même que tel autre, etc.

On appelle modèle de trace le vocabulaire de la trace. Le modèle de trace permet la compréhension et la manipulation de la trace en décrivant abstraitement les éléments, c'est-à-dire en typant les observés et les relations.

On appelle trace modélisée, ou M-trace, l'association d'une collection d'observés temporellement situés structurée par leurs relations et d'un modèle explicite de cette collection.

De ce point de vue, une trace numérique obéit à un modèle de trace, qui décrit les objets qui en font partie. Quel que soit le niveau d'abstraction des éléments de la trace (i.e. leur proximité aux événements informatiques), le modèle de trace vient préciser comment il est possible de les comprendre et de les utiliser. Un modèle de trace peut se limiter à une description des classes d'objets observés, mais définira aussi des classes de relations entre ceux-ci. Une M-trace est toujours associée à un modèle de trace définissant les éléments qui la composent : observés et relations. Un modèle de trace peut être « implicite »,

c'est-à-dire uniquement présent dans le code de l'outil l'utilisant comme c'est le cas par exemple pour le *Common Log Format* qui est le modèle des traces du serveur Apache. Il peut au contraire être « explicite », c'est-à-dire formalisé suffisamment pour permettre l'échange et la réutilisation de traces (un schéma XML décrivant et contraignant le contenu d'une trace XML par exemple).

Considérer les traces numériques comme inscriptions de connaissance implique que le but de la modélisation d'une trace est que cette dernière doit faire sens tant pour l'utilisateur que pour la machine. En ce sens le modèle de la trace doit être non seulement explicite, mais *explicitement formellement* (pour l'utilisateur).

Il s'agit alors de considérer les traces numériques comme une inscription de connaissances modélisées explicitement d'une part, et d'autre part de leur adjoindre les moyens de prendre connaissance de la signification du modèle ainsi que la possibilité de manipuler les traces elles-mêmes. On en vient alors à conférer aux traces d'utilisation le statut de *représentation de connaissance*, au sens que lui accorde habituellement le domaine de l'ingénierie des connaissances. Comme nous allons le voir, ce statut s'acquiert par une *formalisation* de l'approche qui conduit à l'utilisation d'outils spécifiques à ce type de représentation.

3.2 Principe général d'un SBTm

Les concepts de trace et de modèle de traces étant posés, ils permettent de définir un SBTm comme tout système informatique implémentant ces deux notions et qui offrira divers services de manipulation de traces. Autrement dit, on appellera système à base de traces modélisées tout système informatique dont le fonctionnement implique à des degrés divers, la gestion, la transformation et la visualisation de traces modélisées.

Cette approche est illustrée dans la figure 1. Un système à base de traces permet de *gérer* et *d'exploiter* des M-traces et leurs modèles. Il est alimenté par un processus de collecte de traces à partir de l'utilisation du système observé par l'utilisateur (interactions). Plus précisément, la collecte consiste à transformer un ensemble de *sources de traçage*, issues de l'observation des interactions (approches *a priori* ou *a posteriori*) en une *trace première* (Figure 1). Cette trace première est comme son nom l'indique la première dont on puisse considérer qu'elle est modélisée explicitement dans les termes du système, et donc pleinement manipulable et exploitable par celui-ci. Le système peut ensuite offrir différents services d'exploitation des traces, notamment leur *transformation* (fusion, sélection, réécriture) afin de changer de niveau d'abstraction et d'expression des observés, leur *visualisation* interactive (navigation dans la trace), ainsi que le *calcul d'indicateurs* (par exemple numériques), en *fonction du ou des modèles de traces établis*.

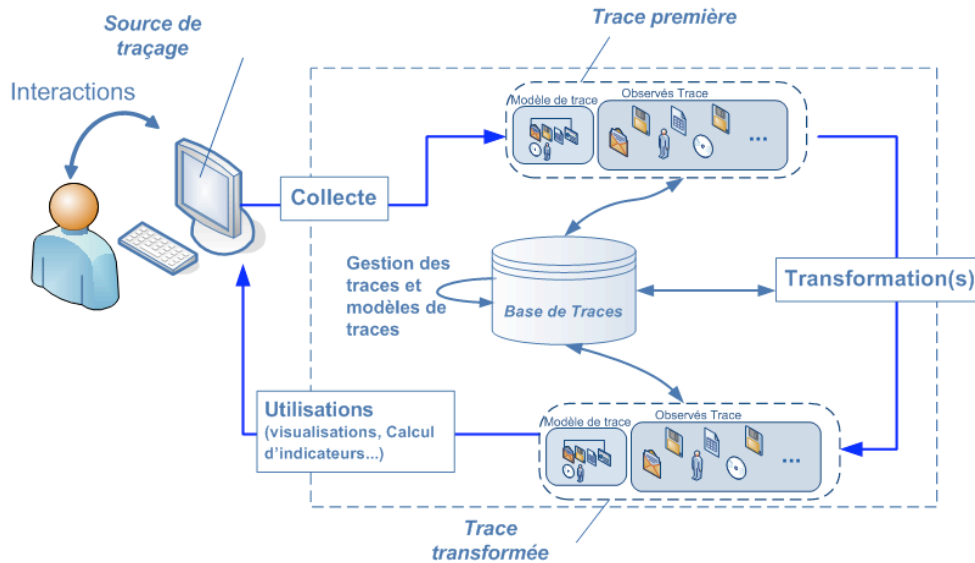


Figure 1: Principe général d'un Système à Base de Traces modélisée

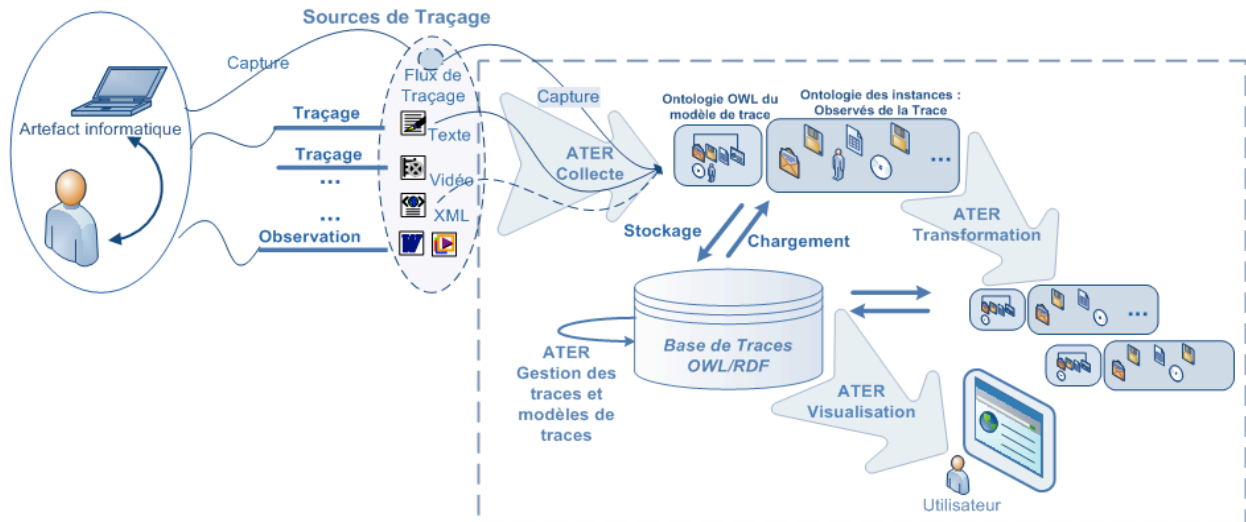


Figure 2 : Architecture générale du système ATER

Cette prégnance des modèles de traces permet d'envisager une ingénierie de la trace basée sur ces modèles, ainsi que la réalisation d'outils génériques réutilisables et partageables dans des contextes applicatifs différents.

3.3 Formalisation

Dans un SBTm, nous formalisons le modèle de trace comme une structure $\Theta=(\Theta_c, \Theta_r)$ où :

- Θ_c est un ensemble fini de classes d'objets θ_c décrivant et typant les observés de la trace
- Θ_r est un ensemble fini de types de relations θ_{ri} avec $\theta_{ri} \subset (\theta_{ci})^k$ et qui est noté $\theta_{ri}(\theta_{c1}, \theta_{c2}, \dots, \theta_{ck})$ ($k > 1$ étant l'arité de la relation θ_{ri}).

Le modèle de trace est finalement un ensemble d'objets étiquetés représentant le vocabulaire de la trace. Une manière d'implémenter un tel modèle est de décrire la trace par une ontologie⁷ comme c'est le cas par exemple dans l'approche Musette [20], où le modèle de trace est une ontologie (exprimée en RDF) décrivant des « objets d'intérêts », entités ou événements, ainsi que des états et des transitions observés lors d'une activité.

La trace, qui est donc toujours associée à un modèle, est formellement définie comme un 4-uplet $\text{Trace} = (D_p, O_{tr}, R_t, R_s)$ associée toujours à un modèle Θ_{trace} où :

⁷ Une ontologie exprimant des concepts et des relations représentant les observés de la traces.

- D_p est un domaine temporel ($D_p = (I_t, \leq)$) où I_t est un ensemble fini d'instantanés et \leq une relation d'ordre dans I_t .
- O_{tr} est un ensemble fini des éléments (donnés) de la trace représentant l'ensemble des observés (instances des θ_{ci} de Θ_{trace}).
- R_t est une relation représentant les liens temporels $D_p \times O_{tr}$ ($R_t \subseteq D_p \times O_{tr}$)
- R_s est une relation représentant les liens structurels dans $(O_{tr})^k$ représentant l'ensemble des instances de θ_{ri} de relations type de Θ_{trace} .

La formalisation minimale que nous proposons donne les bases générales permettant d'envisager l'implémentation de systèmes à bases de traces modélisées, et d'en considérer les modèles, transformations et visualisations au niveau général. La partie qui suit présente l'implémentation du cadre SBTm que nous développons au LIRIS : le système ATER.

3.4 Le Système ATER : un environnement pour exploiter / réutiliser les traces

ATER est une application implémentant la notion de systèmes à base de traces. ATER permet de collecter, transformer, requêter, visualiser et de partager des traces modélisées en tant que représentations de connaissances. ATER supporte l'expression de modèles de trace comme ontologies OWL décrivant des concepts et relations structurant les éléments de la trace. Les traces décrites par un modèle de trace ontologique sont constituées d'instances (individus) représentant leurs observés.

La Figure 2 montre l'architecture du système ATER. Le niveau « collecte » fournit une trace première décrite par un modèle. Plus précisément, la collecte permet de passer, en temps réel ou différé, d'une ou plusieurs sources de traçage quelconques à une trace première stockée dans une base de traces OWL/RDF. La collecte passe par deux étapes, (a) la description du modèle de la trace première contraignant et structurant la trace collectée et (b) l'instanciation du modèle pour créer la trace en interprétant les sources de traçage de manière *ad hoc*⁸. Étant donné la diversité des sources de traçage et des techniques pouvant être utilisées dans cette seconde étape, le système ATER ne prend en charge que la première étape de construction de l'ontologie du modèle

⁸ En programmant le code permettant de lire les sources et d'instancier le modèle de trace OWL. Le modèle de trace OWL peut être extrait automatiquement à partir des sources de traçage si le format des sources est reconnu. Par exemple, une application qui exporterait une trace en format XMI (XML Metadata Interchange) décrivant les données tracées ainsi que le modèle typant ces données.

de trace⁹. Par contre, une fois la trace première instanciée, le système permet la gestion intégrée et générique des traces : l'utilisateur peut spécifier des transformations de traces, manuellement ou en exprimant des modèles de transformation. Il peut également bénéficier des services de navigation dans les traces.

3.5 Exemple des traces « eLycée » dans ATER

La « classe virtuelle » que propose *eLycée* pour ses enseignements, repose sur un outil de visioconférence, qui fournit notamment des tableaux blancs partagés et un outil de « co-navigation » (*eMediathèque*) : il s'agit pour les élèves de pouvoir naviguer parmi des ressources de façon collective. Les ressources pouvant être des pages *Web*, des textes, des vidéos, du son, etc., elles peuvent être « poussées » à tous par un utilisateur particulier, créant ainsi dans chaque interface un onglet de navigation partagée. L'activité pédagogique est construite d'une part comme une *exploitation de ressources* et d'autre part comme une *production* collective ou individuelle de créations (écrites pour la plupart) nécessitant l'échange et la collaboration des participants. Les activités pédagogiques proposées sont complexes et collaboratives.

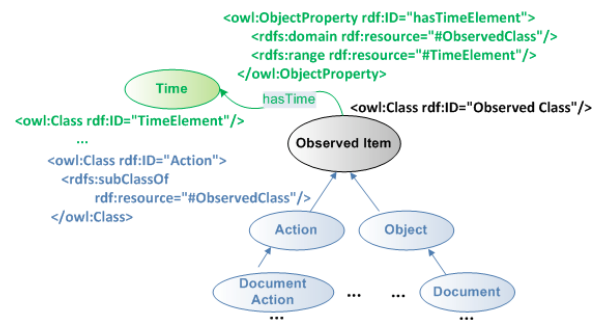


Figure 3 : Ontologie du modèle de trace de *eLycée*

Dans le système ATER, aucun modèle de trace n'est fixé, seul un modèle minimal extensible très générique est imposé, correspondant à la formalisation de très haut niveau proposée dans la partie 3.3. Les systèmes et utilisateurs exploitant le SBT doivent exprimer leur modèle de trace et disposent alors de services de stockage, de transformations et de visualisations.

Dans cette partie, nous allons montrer un exemple de traitement de traces pris charge par le système ATER en donnant deux exemples de traces et de modèles exprimés en OWL, ainsi qu'une transformation exprimée en SPARQL.

Une trace est un ensemble d'observés temporellement situés, pour cela, la description d'un modèle de trace doit

⁹ Il n'est pas envisagé pour l'instant de développer un module de collecte générique paramétrable pour importer différents formats de traces.

essentiellement prendre en compte la description du temps. Nous avons défini dans ATER un schéma minimal de description de modèle de trace montré dans la figure 3, qui correspond au modèle de la trace première (actions et objets essentiellement). La figure 4 montre les classes d'observés d'une trace de plus haut-niveau, dédiée à une activité particulière de co-navigation.

En tant qu'environnement de modélisation de traces, ATER permet de concevoir et de partager facilement des modèles de traces. Il permet également de décrire des transformations à partir des ontologies/modèles de traces.

Les transformations consistent essentiellement à sélectionner des parties de traces de bas niveau pour générer des traces de plus haut-niveau pertinentes au regard de l'activité. Différentes transformations sont possibles, allant de la simple sélection à des filtres plus élaborés, en passant par la réécriture par regroupement de séquences ou motifs d'observés en observés de plus haut niveau d'abstraction.

Le modèle de la trace transformée (figure 4) décrit des objets plus « proches » de l'activité. Dans la figure Z, un exemple de règle de transformation est proposé ($Production \leftarrow Object.type = \text{« Whiteboard »}$), qui permet d'associer le concept « Production » au filtre décrivant les objets de type « Tableau blanc ».

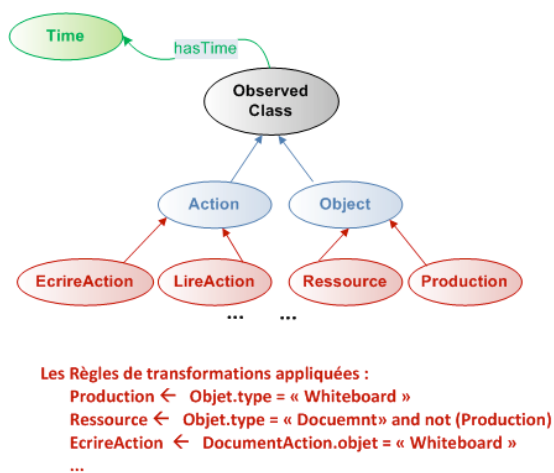


Figure 4 : Ontologie du modèle de la trace transformée

Ainsi, le modèle de trace transformée peut être enrichi et correspondre mieux à ce que l'utilisateur souhaite visualiser dans son activité. Dans l'exemple qui nous intéresse les documents du premier modèle (trace première) sont transformés en « ressources » et « productions » qui correspondent à ce que l'utilisateur manipule réellement du point de vue de son activité. La visualisation de la trace transformée fournit alors à l'utilisateur des moyens de s'emparer de la trace comme véritable trace d'utilisation des documents et des ressources afin de soutenir son activité.

Au niveau implémentation, des langages de manipulations et d'interrogation d'ontologies peuvent être utilisés tels

que OWL QL, SPARQL, etc., Les différentes transformations sont implémentées sur la base de langages d'interrogation de base de connaissances ontologiques afin de profiter de la maturité des outils associés. Les transformations peuvent être combinées pour croiser les résultats et ainsi améliorer la pertinence des traces transformées. L'exemple simple que nous avons présenté ne doit donc pas faire oublier que des règles de transformation plus complexes peuvent être exprimées pour identifier et extraire des inscriptions de connaissances pertinentes.

3.6 Applications en cours

Le système ATER est en cours de développement au LIRIS¹⁰. Une première version *béta* devrait voir le jour et être diffusée à l'automne 2007. Elle permettra de gérer des traces et des modèles de traces, intégrera en standard des transformations et des outils de visualisation minimaux. Une architecture de plugins permettra aux utilisateurs d'ajouter des transformations, des visualisations et des calculs d'indicateurs correspondant à leur besoins. Sur un plan plus technique, une étude des mécanismes d'inférence et d'interrogation de la base de traces RDF/OWL est en cours d'évaluation notamment son exploitation au sein de transformations.

Parallèlement aux développements d'ATER, plusieurs applications sont en cours de développement utilisant le Framework, dans un objectif de validation de celui-ci. Ces applications sont en majorité construites autour d'outils EIAH (environnements informatiques d'apprentissage humain). Autour de l'entreprise eLycée et de l'application eMédiathèque [21], dans un objectif d'analyse de l'activité des apprenants : les traces issues des outils sont intégrées dans le système ATER, puis analysées afin de comprendre l'activité telle qu'elle s'est réellement déroulée, notamment pour les tuteurs.

Autour de l'outil Géonote¹¹ pour l'analyse à base de traces. Géonote permet de gérer les documents de binômes d'apprenants en classe de SVT (Sciences de la Vie et de la Terre), permettant de préparer une sortie géologique sur le terrain puis de rendre compte des différents éléments récoltés et appris au cours de celle-ci. ATER-Géonote permettra à un analyste (professeur de SVT) de modéliser les différents modes d'utilisation de l'outil, et donc les différents modes d'apprentissage.

Autour de l'EIAH AMBRE-Add [22] pour construire un outil d'assistance à l'apprenant au cours de son apprentissage de méthodes additives. ATER est ici utilisé comme un outil extérieur à AMBRE, stockant des traces, les transformant et les fournissant au besoin à l'outil d'assistance. Différents indicateurs (taux d'échecs ou de

¹⁰ L'application est *Open Source* (LGPL), et utilise notamment les technologies JAVA/Eclipse : Rich Client Platform, Eclipse Modeling Framework, GMF, etc.

¹¹ <http://praxis.inrp.fr/praxis/projets/geomatique/geonote/>

réussite) et des séquences d'observés sont obtenus par interrogation de la base de trace ATER. Les traces transformées alimentent une base de connaissances sur l'activité, qui pourra être traitée par un système d'assistance à base de règles. ATER-AMBRE permettra aussi de capitaliser des modèles de connaissance d'apprenants (les connaissances mathématiques telles qu'elles se révèlent dans les traces) et des profils d'apprenants (connaissances mathématiques des apprenants et modes d'apprentissage).

Autour de la visualisation de traces personnelles d'utilisation d'une machine personnelle (navigation web, mails envoyés et reçus, bureautique, gestion de fichiers, etc.). Un *keylogger* (logiciel espion) a été utilisé pour enregistrer plusieurs jours d'utilisation d'un système, et une trace première a été modélisée. L'objectif de cette application est de s'intéresser plus particulièrement à la présentation de la trace et à l'interaction avec celle-ci.

Toutes ces applications nous permettent d'une part d'appuyer les choix architecturaux pour la conception de ATER, d'autre part d'expérimenter différentes applications visées par le système.

4. Vers une ingénierie de la trace ?

Jusqu'ici nous n'avons finalement présenté que les moyens que nous avons mis en œuvre pour *gérer et traiter* les traces modélisées. Cette présentation n'explique en revanche pas comment sont déterminées ces traces, i.e. comment doit s'effectuer le travail de *modélisation* de ces traces. Le déploiement d'outils, aussi puissants soient-ils ne pourront en rien garantir l'efficacité de notre approche si nous ne pouvons qualifier et encadrer la démarche de modélisation elle-même. Nous proposons ici quelques éléments liés à cette question.

Nous avons vu (cf. 3.1) qu'une M-trace, accompagnée de son modèle devait faire sens pour l'utilisateur autant que pour la machine. Le point crucial est donc de *rendre observable l'utilisation* d'un environnement numérique en tenant compte de ce double point de vue. Cet objectif général est toujours lié à la finalité de la trace, i.e. la réutilisation que l'on souhaite en faire à terme.

Pour connaître une situation, et en particulier pour y déceler une trace, un individu ne pourra en exercer la *critique* [3] que dans la mesure du contexte dans lequel il se trouve. Afin de disposer, dans l'environnement numérique, d'une telle facilité, il est nécessaire de garantir qu'une trace fera sens, donc d'orienter sa modélisation relativement à un type d'activité particulier, formant un cadre de pertinence au-delà duquel la trace peut ne plus faire apparaître d'éléments pertinents au regard de l'activité dans l'utilisation de l'environnement. Par exemple, nous avons produit, dans le cadre du travail sur *eLycée* une modélisation de la trace dans l'optique de rendre l'activité *réflexive* (partie 3.5).

Si l'objectif de la modélisation est de permettre la *génération automatique* d'une trace, ce n'est peut-être pas le plus important. En effet, l'objectif fixé à la modélisation de la trace est de rendre celle-ci, une fois générée, *intelligible en tant que trace, interprétable en tant qu'organisation des inscriptions manipulées, et (ré)utilisable dans le cadre d'un travail intellectuel*.

La démarche exigée par la modélisation de trace est extrêmement proche de la démarche de modélisation en Ingénierie des connaissances (IC) telle que décrite dans [3]¹². Elle doit conjuguer des modèles très différents en tentant d'y faire émerger des éléments critiques. Pour schématiser, nous pouvons dire que la modélisation de trace doit adopter simultanément deux démarches. Une démarche *descendante* en partant d'un *modèle de l'activité* pour atteindre un modèle de l'utilisation de l'environnement dans le cadre d'une activité que l'on veut observable. Une démarche *ascendante* en partant du *modèle de conception* de l'environnement observé (et des éléments techniquement enregistrables que ce dernier définit : les *logs*) pour se hisser à un niveau de description auquel la trace fera sens pour un utilisateur¹³.

Contrairement à certains auteurs qui annoncent que tout ce qui est nécessaire à l'obtention de telles « informations tracées » (en général) peut se limiter à la récupération de *logs* et que le problème fondamental est un problème de visualisation de ces derniers [23], nous pensons qu'une trace en tant qu'inscription de connaissances n'est pas réductible à l'ensemble des *logs* dont elle peut être issue (enregistrements automatiques).

Il ne s'agit pas de penser la trace comme une source de d'informations quantitatives¹⁴, mais comme inscription de connaissance, en tant qu'organisation *idiographique* des inscriptions manipulées [3]. Dans ces conditions, une modélisation de trace d'utilisation peut être considérée comme une *démarche particulière de modélisation en IC*.

Pour instrumenter cette « ingénierie de la trace », que nous venons d'esquisser à grands traits, nous avons proposé un cadre conceptuel aussi rigoureux que possible baptisé *Systèmes à Base de Traces* (SBT) afin de penser les traces d'utilisation d'environnements numériques de façon *générique*.

5. Conclusion

Nous avons débuté cet article en pointant le fait que la notion de trace dans les environnements numériques mérit

¹² Où il est expliqué que l'alternative classique d'un travail de modélisation des connaissances est soit de « coller au problème » soit d'y « plaquer un modèle ».

¹³ Ajoutons que la modélisation tiendra nécessairement compte d'autres modèles : modèle de l'activité considérée, modèle d'interaction au sens ergonomique, etc.

¹⁴ Nombre de visites d'un site, fréquences d'utilisation de telle ou telle fonction : types d'information généralement utilisés pour le profilage utilisateur.

taut d'être interrogée, notamment dans le cadre du travail intellectuel reposant sur une manipulation de documents numériques. Nous avons fait valoir l'importance de créer les conditions d'existence de *traces d'utilisation* dans les environnements numériques, traces dont nous avons expliqué qu'elles constituent elles-mêmes des *inscriptions de connaissances*. Cette dernière proposition nous a conduit à considérer la *modélisation* des traces comme l'aspect central d'une approche à base de traces. Après avoir signalé que les approches classiques ne permettaient pas de produire ce type de réflexion générique sur les traces numériques qu'elles utilisent, notamment en utilisant souvent une modélisation *implicite* de la trace, nous avons présenté notre propre des SBTm. Celle-ci repose au contraire des précédentes sur une modélisation *explicite et formelle* des traces. Le cadre conceptuel que constitue le SBTm a été implémenté dans le système ATER, dont nous avons décrit le fonctionnement, et dont nous avons donné diverses applications actuellement en cours, notamment celle qui concerne la plateforme de « classe virtuelle » qui nous a servi d'exemple. Finalement, nous avons expliqué en quoi la modélisation d'une trace d'utilisation relevait d'une démarche de modélisation en Ingénierie des Connaissances, nous permettant de proposer une « ingénierie de la trace ».

Les travaux évoqués dans cet article se poursuivent, tant sur le plan théorique (modèles de traces, transformations, services génériques à base de traces), que pratiques, au sein de plusieurs projets¹⁵.

Bibliographie

- [1] Serres, A. (2002). *Quelle(s) problématique(s) de la trace ?* Séminaire du CERCOR, 13/12/2002, Traces et corpus dans les recherches en SIC. Sciences de l'Information et de la Communication.
- [2] Chabin, M.A. (2004). *Document trace et document source. La technologie numérique change-t-elle la notion de document ?* dans la Revue I3 4(1), p141-158. 2004I. Troiseu-Pami, Un article intéressant, *Journal de Spirou*, Vol. 17, pp. 1-100, 1987.
- [3] Bachimont B. (2004). *Pourquoi n'y a-t-il pas d'expérience en ingénierie des connaissances ?* Actes de la conférence « Ingénierie des connaissances (IC2004) », p. 55-64, Lyon, 2004.
- [4] B. Bachimont, & S. Crozat. (2004). *Préconisations pour une instrumentation numérique des contenus documentaires : leçons tirées de cinq ans d'expérience dans l'enseignement*. In Actes de la conférence « Ingénierie des connaissances (IC2004) », N. Matta (ed), Lyon. Presses Universitaires de Grenoble.
- [5] Stiegler B, (1995), Annotaiton, navigation, édition électronique : vers une géographie de la connaissance. In actes du séminaire Hypermédias, Education et Formation, p. 27-37.
- [6] Pédaque, R.T. (2005). Le texte en jeu, Permanence et transformations du document. STIC-SHS-CNRS. http://archivesic.ccsd.cnrs.fr/docs/00/06/26/01/PDF/sic_00001401.pdf.
- [7] Mille A. (2006). Raisonner à Partir de l'Expérience Tracée. «Le storytelling : concepts, outils et applications», Eddie Soulier (dir), Traité IC2, Série Informatique et SI, Hermes.
- [8] Laflaquière, J., Settouti, L.S., Prié, Y., Mille, A. (2006). *Trace-Based Framework for Experience Management and Engineering*. Proceedings KES2006 Knowledge-Based Intelligent Information and Engineering Systems, 10th International Conference, Springer publisher, p. 1171-1178, Bournemouth, UK.
- [9] Lieberman, H. (2001). Interfaces that Give and Take Advice, in Carroll, J. (Ed). *Human-Computer Interaction for the New Millenium*, ACM Press/Addison-Wesley, pp. 475-485.
- [10] Beauvisage T., *Sémantique des parcours des utilisateurs sur le Web*. Thèse Sciences du langage. Paris : Université Paris X – Nanterre, 2004, 361p
- [11] Hilbert D., Redmiles F.D. *Extracting Usability Information from User Interface Events*, ACM Computing Surveys, Vol. 32, No. 4, Dec. 2000, pp. 384-421.
- [12] Neal, A. S. & Simons, R. M. (1983). Playback: A method for evaluating the usability of software and its documentation. In *Proceedings of CHI '83*, pp. 78-82
- [13] Tauscher, L. & Greenberg, S. (1997). How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *International Journal of Human Computer Studies, Special issue on World Wide Web Usability*, 47(1). p97-138.
- [14] Mille A., Prié Y., (2006). Une théorie de la trace informatique pour faciliter l'adaptation dans la confrontation logique d'utilisation/logique de conception. Dans 13eme Journées de Rochebrune - Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels, jan 2006, Rochebrune, 12 pp.
- [15] Ollagnier-Beldame, M. (2006). *Traces d'interactions et processus cognitifs en activité conjointe : Le cas d'une co-rédaction médiée par un artefact numérique*. Thèse en Sciences Cognitives de l'Université Claude Bernard (Lyon 1), décembre 2006, 247p.
- [16] Jermenn P.R., Soller A., Muhlenbrock M., (2001). *From mirroring to guiding: A review of state of the art technology for supporting collaborative learning*. In Proc. of European Perspectives on CSCL. Bergen, Norway, pp.324-331
- [17] Escorcia, D., (2006). La participation des composantes métacognitives à la production d'écrits des étudiants en sciences humaines et sociales. 8ème biennale de l'éducation et de la formation : Experience(s), savoir(s), sujets(s), Lyon, avril, 7p.
- [18] Plaisant, C., Rose, A., Rubloff, G., Salter, R. & Shneiderman, B. (1999). The Design of History Mechanism and Their Use in Collaborative Educational

¹⁵ Projet « personnalisation des EIAH » du cluster Rhône-Alpes ISLE, projet ANR Procogec (gestion de connaissances et communautés de pratiques), projet ANR ITHACA (traces de collaboration synchrone).

Simulations, in *Proc of the Computer Support for Collaborative Learning*, pp. 348-359, Palo Alto, CA.

- [19] Settouti, L.S., Prié Y., Mille, A., Marty, J-C. (2006). *Système à base de traces pour l'apprentissage humain*. TICE 06, Toulouse, 2006.
- [20] Champin P.A., Prié Y., Mille A., *MUSETTE: a Framework for Knowledge Capture from Experience*, short paper in *EGC'04, Clermont Ferrand*, jan. 2004.
- [21] CRAM D., JOUVIN D., MILLE A. (2007). Visualisation interactive de traces et réflexivité : application à l'EIAH collaboratif synchrone eMédiathèque. Soumis au numéro spécial "Analyses des traces d'utilisation dans les EIAH". *STICEF, revue francophone de Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*.
- [22] GUIN-DUCLOSSON N., JEAN-DAUBIAS S., NOGRY S. (2002). The AMBRE ILE: How to Use Case-Based Reasoning to Teach Methods. *Proceedings of ITS'2002, Biarritz, Springer*, 782-791.
- [23] Roussel N., Tabard A., Letondal C., (2006). All you need is log. *Proc. WWW2006 Workshop on Logging Traces of Web Activity: The Mechanics of Data Collection*, Mai 2006.