

Analyser les traces d'activité pour évaluer les ressources électroniques

HULIN Thibaud
INSA de Lyon – LIRIS CNRS UMR 5205
thibaud.hulin@insa-lyon.fr

Notes biographique : Thibaud Hulin est actuellement A.T.E.R. en sciences de l'information et de la communication à l'INSA de Lyon. Il est titulaire d'un doctorat de philosophie acquis à l'Institut d'Ergologie de l'Université de Provence, formateur en recherche d'information à la bibliothèque de l'INSA, et effectue sa recherche sur les traces d'activité au Laboratoire d'InfoRmatique en Image et Systèmes d'information (LIRIS), à l'Université de Lyon 1.

Mots-clés : évaluation des ressources électroniques, trace, activité.

Introduction

Les études récentes sur le « web log analysis » orientent volontiers l'étude des usages vers l'analyse des fichiers produits par des serveurs de données. Ces approches privilégient les données quantitatives obtenues du côté du fournisseur de service, et tendent à se normaliser avec le projet Counter (Boukacem et Schöpfel 2005). Ces résultats peuvent être nuancés par des analyses qualitatives de l'activité en étudiant les usages généraux des revues électroniques (Broudoux et Chartron 2008) ou les points de vue des auteurs (Nicholas, Jamali, et Rowlands 2006). Pourtant, force est de constater qu'approches qualitatives et quantitatives se rejoignent rarement en raison des objectifs différents : la question de l'évaluation des ressources est finalement isolée des questions à orientation clairement ergonomiques comme celle de la modélisation de l'activité de recherche d'information ((Tricot et Rouet 1998) ou celle de la normalisation des interfaces de bibliothèques numériques (Lompré 2007). Pourtant, l'approche statistique produit elle aussi des connaissances sur l'activité de recherche d'information, et a besoin d'une connaissance précise des usages pour s'affiner. D'où la nécessité de construire une méthodologie scientifique qui puisse récupérer des informations sur les usages du côté de l'utilisateur concret, tout en les rendant disponibles à l'analyse générale. Pour cela, nous proposons ici une méthodologie basée sur l'analyse des traces d'activité plutôt que sur les journaux systèmes.

Notons que cette approche locale tend à renforcer la capacité des bibliothèques à comprendre les pratiques de ses usagers, sans dépendre des données fournies par les éditeurs. L'enjeu est alors de réduire le coût de l'analyse tout en respectant des contraintes éthiques qui supposent pour l'utilisateur comme pour la bibliothèque d'accéder à l'ensemble des traces d'activité. L'intérêt de manipuler ses traces d'activité a donc un intérêt double : assurer une meilleure transparence des investissements, mais aussi favoriser la recherche d'information par une meilleure prise de conscience de l'activité réalisée.

Présentation du modèle

La construction d'un modèle de l'activité de recherche documentaire suppose plusieurs étapes selon (Jermann et coll. 2001) . Nous proposons d'adapter ce modèle computationnel en l'adaptant à notre

problématique de mesure des ressources. Pour cela, nous nous appuyons sur le système à base de traces décrit en (Champin et coll. 2004; Champin, Prie, et Mille 2003; Settouti et coll. 2006; Laflaquière et coll. 2006; Ollagnier-Beldame, Magali 2006; Mille et Prié 2006) qui nous fournit une théorie générale de la trace informatique. À partir de ces travaux, nous proposons ici notre propre approche de la trace dans le cadre de l'évaluation des ressources numériques documentaires. Plusieurs phases sont nécessaires pour construire une évaluation finale à partir de l'analyse du travail des utilisateurs.

Phase 1 : la collecte des données

Dans un premier temps, nous réalisons une étude ergonomique sur les actions possibles de recherche documentaire à l'aide des logiciels appropriés, ici firefox pour la navigation et zotero pour la gestion des références. Ceci permet d'identifier les actions de l'utilisateur en suivant les possibilités de l'interface. En effet il n'est pas possible d'étudier le code directement, qui n'est pas structuré en terme d'actions. L'analyse de ces actions permet de construire une *hiérarchie des actions utilisateurs*. Cette analyse est doublement intéressante car elle permet de :

- définir une classe d'objets appelés « observés » qui possèdent un certain nombre d'attributs et qui définissent un modèle de trace première ;
- définir les emplacements dans le code des logiciels où l'on placera des capteurs afin de journaliser les actions utilisateurs.

Phase 2 : la construction d'un modèle de trace

Une fois la liste des actions obtenues, celles-ci sont organisées en objets et attributs à partir d'un objet général dit « observé », composé d'attributs fondamentaux : un numéro d'identification, un descripteur et une date. De là, on définit des schémas de transformation de ce modèle de trace première. Par exemple, les actions de navigation vont donner lieu à des changements de focus sur telle ou telle ressource électronique. Ces transformations vont aussi donner lieu à des représentations exploitables de l'activité pour l'utilisateur.

Phase 3 : la définition d'un profil évaluatif

Plutôt que de lier des variables aux actions tracées afin d'évaluer simplement les ressources, nous définissons un modèle d'activité normalisé fait d'une séquence d'actions types. Plus ce modèle est répété et suivi, plus la ressource sera évaluée positivement par rapport à ce modèle. Certes le système de calcul est plus complexe, mais il permet d'éviter de supposer qu'une suite d'actions isolées est nécessairement positive vis-à-vis d'une ressource.

Phase 4 : la représentation des traces pour l'utilisateur

La représentation des traces pour l'utilisateur donne lieu soit à l'aide d'un tableau des actions réalisées soit à l'aide d'une représentation graphique chronologique. L'utilisateur a alors la possibilité de commenter ses traces. La présence d'un attribut « notation » lui permet d'évaluer manuellement les résultats de ses recherches.

Phase 5 : le moteur de calcul de traces

La comparaison de la trace d'activité réelle d'un utilisateur avec une séquence attendue ou prescrite se fait par la comparaison de séquences. L'évaluation se fait par le biais de constantes et de règles logiques qui permettent de calculer des écarts.

Phase 6 : le classement des ressources

Le classement et l'évaluation des ressources utilisées est ici résultat des comparaisons de séquences d'activité à l'aide du moteur de calcul de traces. Le groupement de règles agissent ici comme des automates finis au sens mathématique.

Conclusion

La système à base de traces (SBT) développé par les recherches en informatique de Mille est selon nous le meilleur modèle pour relier des événements informatiques à une activité complexe et continue, comme « élan de vie et de santé sans borne définie »(Schwartz 2003). Appliqué à la question de l'évaluation des ressources électroniques, il indique la possibilité pour les centres de documentation de mieux connaître les usages des ressources en tenant compte de l'assentiment de l'utilisateur qui à son tour peut directement évaluer son travail sur ces outils. Enfin, ce méta-modèle permet d'élaborer des calculs complexes d'analyse de l'activité grâce aux sources récupérées. Bien qu'il existe quelques travaux sur la réflexivité des usages dans la recherche d'information, l'usage d'une réflexion générale sur les traces d'activité nous semble aujourd'hui nécessaire.

Cependant, nous n'ignorons pas certaines difficultés rencontrées. Parmi celles-ci, la complexité de l'activité suppose un modèle *a priori* qui peut tromper sur l'activité cognitive réelle, l'évaluation par l'utilisateur restant une perspective insuffisante pour maîtriser les enjeux heuristiques sur les processus mentaux. A défaut de maîtrise, cette approche a l'avantage d'offrir un niveau de portée de vue sur l'activité important, tout en permettant l'usage de raisonnements mathématiques avancés sur les traces par la construction d'automates (Diekert 1995). Enfin, pour éclairer les problèmes de sens liés à l'interprétation des traces obtenues, une mise en œuvre expérimentale est en cours de réalisation par nos soins.

- Boukacem, Chérifa, et Joachim Schöpfel. 2005. Statistiques d'utilisation des ressources électroniques : le projet COUNTER. *Bulletin des Bibliothèques de France* 50, no. 4: 62-64.
- Broudoux, Evelyne, et Ghislaine Chartron. 2008. *Traitements et pratiques documentaires*. Paris: ADBS.
- Champin, Pierre-Antoine, Julien Laflaquière, Alain Mille, et Yannick Prié. 2004. Musette: a framework for Knowledge Capture from Experience. *12ème Atelier de Raisonnement à Partir de Cas*.
- Champin, Pierre-Antoine, Yannick Prie, et Alain Mille. 2003. Musette: Modelling uses and tasks for tracing experience. Dans *ICCBR*, 3:279-286.
- Diekert, Volker. 1995. *The book of traces*. River Edge, NJ (USA): World Scientific.
- Jermann, Patrick, Amy Soller, M. Muehlenbrock, Martin, et Martinez Monés, Alejandra. 2001. From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. Dans *Proceedings of the First European Conference on Computer-Supported Collaborative Learning*, 324-331.
- Laflaquière, Julien, Lotfi Settouti, Yannick Prié, et Alain Mille. 2006. Trace-based framework for Experience Management and Engineering. *Lecture Notes in Computer Science* 4251: 1171-1178.
- Lompré, Nicole. 2007. Normes ergonomiques et usages des bibliothèques numériques. Dans *Usages et pratiques dans les bibliothèques numériques*, éd. Pappy, Fabrice, 1:25-52. Hermès-Lavoisier. Paris.
- Mille, Alain, et Yannick Prié. 2006. Une théorie de la trace informatique pour faciliter l'adaptation dans la confrontation logique d'utilisation/logique de conception. Dans *13èmes journées de Rochebrune -Traces, Enigmes, Problèmes*, 12. Rochebrune.
- Nicholas, David, Hamid R. Jamali, et Ian Rowlands. 2006. On the tips of their tongues: authors and their views on scholarly publishing. *Learned Publishing* 19, no. 3: 193.
- Ollagnier-Beldame, Magali, M. 2006. Traces d'interactions et processus cognitifs en activité conjointe: Le cas d'une co-rédaction médiée par un artefact numérique. Rapport de thèse en

sciences-cognitives de l'Université Louis Lumière Lyon 2.

Schwartz, Yves. 2003. *Travail & ergologie*. Toulouse: Octarès éd.

Settouti, Lotfi, Yannick Prié, Alain Mille, et Jean-Charles Marty. 2006. Système à base de trace pour l'apprentissage humain. Dans *Colloque international TICE 2006 «Technologies de l'Information et de la Communication dans l'Enseignement Supérieur et l'Entreprise»*, 25-27. INP Toulouse.

Tricot, André, et Jean-François Rouet. 1998. *Les hypermédias : approches cognitives et ergonomiques*. Hypertextes et hypermédias ISSN 1280-7842. Paris: Ed. Hermès.