

Rapport sur la thèse d'Aurélien TABARD pour obtenir le grade de docteur en informatique de l'université Paris-Sud, intitulée "Supporting lightweight reflection on familiar information"

Le mémoire remis par le candidat est composé de $XI+121=132$ pages, d'une présentation bien maîtrisée, à la fois efficace et élégante. L'auteur recourt souvent à des photographies, des tableaux, des schémas qui éclairent utilement son propos. Le mémoire est entièrement rédigé en anglais. Lorsque des entretiens avec des sujets sont décrits, les propos sont traduits avec la transcription originale en français en note. Pour autant que je puisse en juger, la rédaction est très convenable.

Le texte s'organise en sept chapitres : une introduction, un état de l'art (12 p.), quatre contributions de 16, 13, 17 et 23 p. respectivement, une conclusion suivie d'une bibliographie d'une bonne centaine de références (10 p., avec, pour chaque référence, les renvois aux pages citantes) et de deux annexes. Les chapitres de contributions s'appuient sur des (co)publications dans des actes de conférences internationales de premier plan (CSCW'08, CHI'07), que j'ai consulté. Leurs textes ont été bien repensés.

La thèse d'Aurélien Tabard a pour objet la gestion d'information personnelle pour les chercheurs en biologie. Plus précisément, l'auteur explore la question très importante de l'intermédiation papier - écran dans le contexte des cahiers de laboratoire. Ce terrain d'observation a déjà fait l'objet de plusieurs études des membres du laboratoire InlSitu et un lecteur extérieur comme moi peut regretter que l'auteur ne situe pas plus explicitement dans l'introduction sa contribution dans la chronologie des travaux de l'équipe, plutôt que par des citations dans le corps du mémoire (chapitre 2 en particulier). En revanche, les points principaux de la thèse et son déroulement concret (alternances régulières entre phases de réflexions théoriques, design et évaluation) sont très clairement introduits.

Le chapitre 2 présente une synthèse des travaux utilisés par l'auteur. Il s'appuie sur une critique, de plus en plus partagée, d'une approche trop techno-centrée de la gestion documentaire (c'est à dire accordant plus d'importance aux moyens qu'aux fins). M. Tabard en analyse les spécificités dans le monde de la recherche ("extreme knowledge workers") et souligne très bien la faillite du workflow ordinaire dans la partie exploratoire du travail du chercheur. La différence faite entre information "personnelle" et "familiale" (utilisée mais non maîtrisée) est soulignée. Les techniques d'hybridation papier/informatique sont ensuite passées en revue. Les travaux les plus récents sont analysés de manière assez détaillés, en particulier, la technique de cahier augmenté Anoto. Je trouve dommage que dans cet état de l'art, l'auteur n'accorde que peu d'importance aux travaux concernant la théorie du document, notamment francophones comme ceux du groupe Pédaque (RTP Document numérique, CNRS) ou E. Souchier, Y. Jeanneret, J. Le Marec (dir.) "Lire, écrire, ré-écrire. Objets, signes et pratiques des médias informatisés", Paris, BPI, 2003.

Le chapitre 3 ("Beyond information management : reflection") est sans doute celui qui mobilise le plus de notions hors du champs de l'informatique classique. Il décrit un travail de terrain effectué avec 10 participants, biologistes à l'Institut Pasteur, afin de mieux cerner leur pratiques documentaires. D'un point de vue méthodologique, l'auteur se base sur la "Grounded Theory" des sociologues Galser et Strauss (pp. 28-29, et annexes), que je ne connaissais pas, mais qui me paraît fructueuse. Le chapitre se termine sur le résultat de cette analyse : le "codage axial" (p. 39) : un graphe d'interaction donnant les relations entre les concepts et les catégories mises en œuvre dans la pratique réflexive qu'autorisent les cahiers de laboratoire. Ce graphe (une deuxième version le complète p. 54) met en évidence le rôle central de l'hypomnemata (Stiegler), support de mémoire au sens le plus large, indiscociable de ses protocoles d'accès (un document, en somme...). En préambule à cette analyse, j'aurais apprécié plus de matière sur la taille et les contenus des

corpus de carnets, une analyse plus quantitative de leur typologie, des indications de fréquence des "liens" effectués entre cahiers papier et logiciels. Les quelques photographies présentées de ces carnets donnent vraiment envie d'en savoir plus.

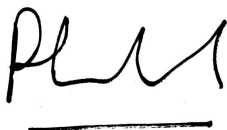
Les chapitres 4 et 5 présentent le logiciel Prism. Résultat d'une longue phase de design participatif (plusieurs réunions dont une regroupant 12 participants, puis 5 utilisateurs pendant 9 mois), Prism joue le rôle de "cahier maître", organisant sur une base quotidienne la production écrite du chercheur sur les différents médias : cahier de laboratoire sur papier augmenté Anoto, saisies de notes par ordinateur, état des documents manipulés dans la session (mail, pages web etc.). Prism est indirectement alimenté en documents par des extensions réalisées sur les navigateurs et éditeurs, mais aussi grâce à un analyseur d'activité basé sur le gestionnaire de "documents récents" du système d'exploitation (pp. 73-77). Prism a ensuite été modifié pour devenir un outil de travail collaboratif en ligne, en s'appuyant sur des mécanismes de flux et de syndication. Le chapitre se termine sur plusieurs remarques intéressantes pour l'évolution future du logiciel, qui témoignent d'une prise de recul de l'auteur (pp. 71-74). Sans doute des techniques de visualisations plus massives pour la navigation dans le cahier maître seraient aussi utiles. Je sais qu'un travail en ce sens a été effectué par l'auteur et d'autres membres de l'équipe (Streamliner, Yuan, Tabard et McKay, 2008) mais l'auteur a choisi de ne pas en faire état dans ce mémoire.

Le chapitre 6 décrit PageLinker, un système de signets Web contextuel. L'auteur remarque à juste titre que les mécanismes d'historique de navigation et de signets (et leurs outils d'éditeurs associés dans tous les navigateurs web) n'ont pas su résister à l'explosion de la masse documentaire en ligne et à la diversité des modes d'accès. Plutôt que de chercher de nouvelles techniques de visualisation/édition, il propose de mieux prendre en compte le contexte qui motive le recours aux signets. Pour cela, il a réalisé une extension du navigateur Firefox qui archive automatiquement (et aussi à la demande) les actions de copier/coller de l'utilisateur entre pages Web. Ces "signets contextuels" sont présentés sur le côté de la page en cours de visite. On pourrait regretter la méthode d'affichage retenue, qui est très textuelle (cf la copie d'écran fig. 49, p. 86). Elle aurait pu être complétée par des techniques de visualisation moins textuelles, déjà industrialisées comme pour les favoris dans la dernière version du navigateur Safari. L'évaluation de PageLinker est très bien justifiée, décrite et discutée (pp. 87-95). Elle indique qu'avec ce procédé, la navigation parasite est réduite et que le taux de revisite des pages est amélioré.

La conclusion s'organise comme de coutume en un bilan et des perspectives. Le bilan me paraît très positif : une bonne analyse des besoins en gestion documentaire (alternance capture / réflexion) des biologistes a permis la réalisation de deux logiciels, qui à mon avis sont tout à fait adaptables à d'autres contextes. Les perspectives évoquées en matière de gestion de l'oubli et de traces d'interaction sont intéressantes.

En conclusion, s'il ne peut épuiser un sujet très riche, le mémoire présenté me paraît d'une grande valeur, tant du point de vue méthodologique que des contributions aux techniques d'interaction. Je donne donc un avis très favorable à la soutenance de thèse de doctorat d'Aurélien Tabard.

fait à Paris, le 4 novembre 2009.



Pierre CUBAUD, professeur des universités <cubaud @ cnam.fr>
Responsable du groupe de recherche "interactivité pour lire et jouer" (ILJ)
Centre d'étude et de recherche en informatique (CEDRIC)
Conservatoire national des arts et métiers, 292 rue saint Martin, 75003 Paris.