HAL open science

Représenter et modéliser l'activité de recherche d'information experte avec des traces d'activité pour l'apprentissage

Hulin Thibaud

▶ To cite this version:

Hulin Thibaud. Représenter et modéliser l'activité de recherche d'information experte avec des traces d'activité pour l'apprentissage. Les Enjeux de l'information et de la communication, 2010, 2010 (1), pp.83-96. 10.3917/enic.010.0500 . halshs-01153877

HAL Id: halshs-01153877

https://shs.hal.science/halshs-01153877v1

Submitted on 6 Apr 2023

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

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

Représenter et modéliser l'activité de recherche d'information experte avec des traces d'activité pour l'apprentissage

Article inédit. Mis en ligne le xx xxxxx 2010.

Thibaud Hulin

Ancien allocataire-moniteur, de parcours pluridisciplinaire, Thibaud Hulin est actuellement A.T.E.R. à l'université de Franche-Comté en sciences de l'information et de la communication, au Laboratoire de Sémiotique Linguistique Didactique et Informatique, site de Montbéliard. Ses recherches portent sur la gestion des connaissances, l'épistémologie et la communication avec des traces d'activité.

Plan

Représenter et modéliser l'activité de recherche d'information experte avec des traces d'activité pour l'apprentissage

Résumé
Introduction
Partie 1 : Méthodologie
Partie 2 : Résultats, analyse des discours
Partie 3 : Discussion sur les concepts de traces et d'activité
Conclusion
Remerciements
Bibliographie

Résumé

Les études sur l'activité de recherche d'information et les pratiques de culture informationnelle montrent que réinterroger son activité peut en améliorer les résultats. Il est possible de renforcer cette réflexivité en favorisant la réappropriation et la restitution des traces numériques « côté client », au lieu de se contenter d'une indexation « côté serveur ». Pour comprendre les représentations de la recherche d'information, nous avons étudié les discours d'étudiants sur leur propre pratique et obtenu des modèles de leurs recherches. Dans leurs discours, la description objective, « en mirroir » de l'activité est dépassée par la description subjective des concepts étudiés.

Abstract

Studies on the activity of information seeking and information literacy practices show that re-examining search behavior may improve search results. It is possible to strengthen this reflexivity by promoting the recovery and restoration of digital traces client-side, instead of just indexing them "server side". To understand the representations of information retrieval, we studied discourse of some students on their own information-seeking behavior and obtained models of their research. In student accounts, subjective description of the concepts in question overshadowed any objective description of the activity.

Resumen

Los estudios sobre la actividad de búsqueda de información y prácticas de alfabetización informacional, muestran que reconsiderar su propia actividad puede mejorar los resultados de búsqueda. Es posible fortalecer esta reflexividad, mediante la recuperación y restauración de huellas electrónicas desde la perspectiva del usuario, en lugar de simplemente indizarlas desde la perspectiva del proveedor de servicios. Para comprender las representaciones de recuperación de información, hemos estudiado el discurso de algunos estudiantes sobre sus propias comportamiento de búsqueda de información y hemos obtenido modelos de sus procesos de investigación. Al final de cuentas la descripción subjetiva de los conceptos en cuestión han eclipsado cualquier descripción objetiva de este proceso.

Introduction

Les études sur la recherche d'information peuvent se partager en deux groupes plutôt hétérogènes : ou bien elles s'intéressent aux systèmes experts et aux algorithmes techniques (Manning, 2008) ; ou bien, plus rares, elles portent sur la subjectivité de l'usager et tentent de modéliser l'activité de recherche elle-même (Tricot et Rouet, 1998). Ainsi, peu de travaux interdisciplinaires envisagent de permettre à l'usager de réutiliser des informations sur son activité en reliant informatique et psychologie cognitive, ergonomie ou sociologie des usages dans un but éducatif. L'hypothèse sous-jacente aux approches mono-disciplinaires est que l'on suppose la réticence naturelle ou l'incompétence du chercheur d'information à poser un regard éclairé sur le travail de sa machine et à être acteur du processus de conception technique. Une division du travail serait donc nécessaire entre le spécialiste de la machine et l'usager. Cette logique divisionnaire implique la nécessité de construire des systèmes experts qui surveilleraient (*watch*) notre activité afin de mieux l'assister. Or, ces systèmes, qui reposent généralement sur des algorithmes linguistiques et mathématiques en développement, connaissent certaines limites (Tamine-Lechani, 2009). Par exemple, les systèmes de suggestion dans les moteurs de recherche donnent parfois des résultats inattendus, et les publicités dites contextuelles sont parfois très éloignées du problème initial du chercheur d'information, ce qui développe les phénomènes de sérendipité. La difficulté que rencontrent les concepteurs de tels systèmes « orientés contexte » est de parvenir à fournir une information personnalisée, en cohérence avec les recherches passées de l'utilisateur. La réponse donnée par les moteurs de recherche et réseaux sociaux qui dominent actuellement le Web consiste à maximiser la visibilité des données de l'utilisateur, et à minimiser la visibilité du processus de traçage. Opérant un « renversement copernicien » (Koyré, 2003), nous nous intéressons à l'hypothèse inverse, en supposant qu'il est plus aisé de décrire la recherche d'information des utilisateurs en prenant pour pivot le poste de l'utilisateur plutôt que le serveur, c'est-à-dire en épousant un point de vue ergonomique. Par la suite, cette perspective conduira à maximiser la visibilité du traçage et à contrôler le niveau de transparence des données ou de l'activité de recherche d'information pour des tiers. Cependant, au préalable, pour construire de tels systèmes, nous devons isoler les éléments qui font sens dans l'activité de

recherche d'information. En poussant l'utilisateur à interroger son activité passée et son activité future, nous espérons récolter les éléments pertinents.

En sciences de l'information et de la communication, le problème du lien entre activité passée et activité future a été abordé lors de l'étude sur les sites médiateurs de Gallica, qui articulent des « traces d'usages effectués » et des traces d' « usages anticipés » (Davallon, 2003). Notons que l'étude des traces numériques, leur diffusion et leur centralisation constituent un risque pour la liberté et l'identité de l'individu, comme le montrent de récents travaux (Arnaud et Merzeau, 2009). Cependant, ces traces ont aussi un intérêt pédagogique important, comme le soulignent (Larose et Jaillet, 2009). Pourtant, des études pluridisciplinaires comme celle sur Gallica se font essentiellement « côté serveur », ce qui implique une distance importante avec l'usager réel, réduit à l'indexation de ses traces. Aussi avons-nous réalisé une expérience menée « côté client » en situation de réflexivité renforcée. Nous tentons d'observer la manière dont l'utilisateur parvient ou non à s'approprier son activité passée pour mieux anticiper son activité future, c'est-à-dire en renforçant la réflexivité de son activité. En analysant les discours des utilisateurs à l'aide d'outils automatiques puis par une lecture critique, nous tentons de faire le bilan de leur recherche, puis de mettre en évidence les actions et entités pertinentes pour la réflexivité afin de modéliser les activités de recherche d'information. Ces modèles doivent être compatibles avec la théorie des systèmes à base de traces pour l'apprentissage (Laflaquière, 2009), sur laquelle nous nous appuyons. Après avoir présenté notre méthode et quelques résultats, nous discuterons les limites de notre approche et proposerons quelques éléments pour enrichir une théorie interdisciplinaire de la trace.

Partie 1 : Méthodologie d'observation

Pour analyser les représentations des chercheurs d'information dans le but de leur permettre d'effectuer un retour heuristique sur leur activité, nous supposerons l'existence de quatre séquences fondamentales :

1. la recherche d'information proprement dite ;
2. la réflexion sur l'activité de recherche d'information passée ;
3. l'anticipation de l'activité future à partir de l'activité passée ;
4. enfin, la transmission de l'activité réalisée en direction d'un tiers.

Aussi, les enquêtes que nous avons réalisées se focalisent sur quatre types de questions, chacun d'eux étant lié à une problématique spécifique :

- la description : quels sont les éléments significatifs pour l'utilisateur ; quels découpages fait-il de son activité ? Quel langage emploie-t-il ?
- l'évaluation : sur quoi porte le regard critique ? Quels sont les critères d'évaluation ?
- l'anticipation : comment l'utilisateur opère-t-il de nouveaux découpages ?
- la transmission : comment l'utilisateur invite-t-il à de nouveaux cheminements, comment pose-t-il des repères dans l'espace documentaire ?

Ainsi avons-nous voulu savoir, à l'aide de questions ouvertes, de quelle manière les utilisateurs décrivent leur activité. Le choix de cette méthode se justifie par le fait que le questionnaire réduit les chances d'apprendre des choses inattendues, tandis que l'entretien ne décide pas de manière *a priori* du monde des références symboliques, du système de cohérence interne des informations recherchées. De la sorte, nous évitons un classement en amont des éléments supposés importants. Pour nous assurer que les personnes ne décrivent pas une activité imaginaire, nous enregistrons les actions de l'utilisateur à l'aide d'une vidéo. Enfin, pour donner un sens à l'exercice, nous avons invité les étudiants à constituer une bibliographie dans le cadre d'une recherche documentaire sur une molécule, qu'ils effectuent durant leurs études à l'INSA de Lyon.

Concernant notre méthodologie d'entretiens, nous avons suivi les conseils généraux de (Blanchet, 2005) et (Guittet, 2003). Plus particulièrement, étant donné notre double statut d'observateur et de formateur à la recherche documentaire, nous situons notre approche dans la lignée de l'observation participante (Malinowski, 1989). A la suite de Max Weber, nous nous installons dans le cadre d'une « sociologie compréhensive » qui étudie l'activité comme « un comportement compréhensible » (Weber, 1965, pp. 329-330) ; c'est-à-dire non pas dérivé de « conditions psychiques », ce qui suppose une séparation sans doute arbitraire entre le subjectif et l'objectif, mais d' « expectations » : nous avons tenté de produire des « anticipations nourries subjectivement » au sujet des comportements attendus, selon une « rationalité subjective », et sur la base d' « expériences valables », principe d'une « rationalité objective » (Weber, 1965, p. 334). Ce positionnement nous permet de suivre un processus d'objectivation qui assure le passage du registre procédural (les savoir-faire) au registre déclaratif (les savoir-dire). Derrière notre questionnaire, nous visons les systèmes de représentations, c'est-à-dire les pensées construites, les idéologies comme « organisation d'opinions, d'attitudes et de valeurs » (Adorno, 1950, p. 2). De cette manière, nous nous efforçons de réduire les présupposés, dans notre interprétation des résultats, qui pourraient déformer notre compréhension des activités observées. Pour compléter cette approche, nous avons donc utilisé des outils d'analyse syntaxique à côté des outils conceptuels plus traditionnels.

L'étude porte sur l'activité d'étudiants ingénieurs de l'INSA de Lyon en quatrième année de pharmacologie, âgés de 22 à 25 ans. Nous nous sommes intéressés particulièrement au travail de 12 étudiants répartis en deux équipes. Ce sont des étudiants brillants, formés à la recherche d'information par les professionnels de la bibliothèque de l'INSA, en quatrième année d'ingénieur en pharmacologie. Dans le cadre d'un projet de recherche de deux ans, les étudiants procèdent à une recherche documentaire sur une molécule d'intérêt thérapeutique, ici le rimonabant et l'exénatide, dont ils étudient les propriétés pharmacologiques, la pharmacocinétique et les aspects économiques. Cependant, afin de favoriser une approche qualitative, nous avons réduit le corpus étudié à des interviews réalisées avec cinq étudiants issus de ces deux groupes. Ainsi nous ne présentons ici que les résultats issus des entretiens réalisés en langue française. Ce choix provient du fait que nous nous intéressons moins aux tendances comportementales d'un groupe social qu'à un ensemble de particularités sémantiques invisibles à grande échelle. La possibilité de généraliser ces observations particulières est effectivement

réduite par l'attention aux détails sémantiques, ce qui constitue probablement une limite ou un parti-pris de notre étude.

Pendant les interviews, nous posons des questions ouvertes, de circonstance. Pour cela, nous nous appuyons sur l'outil QQOQCP (dit « 5Ws » en anglais, attribué à tort à Quintilien (Robertson, 1946) et qui provient du rhéteur grec Hermagoras de Temnos) : Qui fait Quoi ? Où ? Quand ? Comment ? Combien ? et Pourquoi ? Ces questions évitent une réponse simple par oui ou par non, et relancent la parole de l'interviewé. Dans la mesure où il s'agit d'un outil présenté en formation à l'étudiant, son emploi favorise un climat de confiance tout en encourageant à approfondir le questionnement. La consigne, qui instruit l'interviewé et définit le thème du discours, a été rédigée en fonction du guide d'entretien. Elle a été présentée en début de séance. Une observation-test a permis de confronter les idées avec les données. Par la suite, nous avons cependant tenu à conserver notre protocole initial. Les relances sont possibles mais réduites au minimum, l'outil « QQOQCP » permettant de structurer l'entretien et d'assurer la communication sans qu'il soit nécessaire de mettre en place des stratégies d'intervention ou des modes de relances plus lourds (paraphrase, commentaire, etc.).

Afin de prendre appui sur des critères les plus objectifs possibles, nous nous appuyons sur un ensemble d'outils informatiques que nous avons développés pour l'analyse textuelle. Ceux-ci dressent des listes de mots épurés des formes contingentes comme le pluriel ou les conjugaisons, et regroupés par classe. Par exemple, vouloir et voulons sont regroupés sous le lemme « voul », par réduction lemmatique et comparaison d'occurrences. Ces outils permettent d'évaluer les concepts et ressources utilisés. Ainsi le calcul d'un « indice d'usage », un pourcentage arrondi à l'unité, pondère la fréquence du mot dans le discours par sa fréquence dans le corpus. Par exemple, lorsque les étudiants décrivent la première étape de leur recherche d'information, ils semblent insister sur leur utilisation de la base de données spécialisée en médecine PubMed afin de valoriser leurs compétences scientifiques. Cependant, le calcul de l'indice d'usage montre que PubMed est autant invoqué que la base de données de presse Factiva dans cette partie : au début de leur recherche, et malgré les apparences, leurs préoccupations sont tout autant scientifiques qu'économiques. Une relecture critique du corpus permet alors de confirmer cette hypothèse posée par l'analyse syntaxique.

L'ensemble des calculs et des codes utilisés sont disponible sous licence libre en (Hulin, 2009), avec les codes sources de l'interview en Bash, Perl et Python, les résultats et les dictionnaires qui ont servis aux calculs. Nous réfléchissons à la possibilité de transférer ces scripts vers la plateforme de textométrie ouverte TXM (Pincemin, 2010) dont nous recommandons l'usage en raison de la mise à disposition du code source et de sa qualité.

Partie 2 : Résultats, analyse des discours

L'analyse des résultats consiste dans la rédaction d'un rapport analytique, disponible en ligne, puis d'une synthèse que nous présentons ici. Pour réaliser le rapport analytique, nous avons procédé à une série de transformations manuelles ou automatiques à partir de l'enregistrement de l'interview. Cette trace brute enregistrée par le microphone connaît alors une série de

transformations, automatisées ou non. Elles ont pour but de faire sens pour l'analyste. Parmi ces traces transformées, nous distinguerons :

- le corpus de textes issu des interviews retranscrites, une trace écrite brute, l'enregistrement ;
- le résultat du transfert de la trace orale vers la trace écrite, le codage de l'interview ;
- le code épuré, lisible par une machine comme par un humain, obtenu après suppression des commentaires et des questions, des divisions en sections, de l'homogénéité de l'écriture de certains termes, etc. ;
- le code obtenu après réduction des mots les plus fréquents à partir d'un dictionnaire de fréquence ;
- le code obtenu à partir de la réduction des mots à leurs racines signifiantes, à partir d'un dictionnaire de lemmes.

Le code final a donc subit un ensemble de transformations basées sur le jeu des expressions régulières. Le rapport analytique interprète alors successivement :

- le discours d'ensemble, l'esprit général des interviewés ;
- les différents types d'interviews : description 1, anticipation, description 2, évaluation ;
- les différences d'approches selon les personnes ;
- les écrits destinés à un tiers.

Dans le rapport analytique, nous rédigeons nos conclusions, avec comme exemple et entre parenthèses le nom du lemme et un nombre. Dans la partie sur le discours explicite, il s'agit du nombre d'occurrences de chaque lemme dans l'ensemble de la section de texte. Dans la partie sur le discours implicite, il s'agit de l' « indice d'usage » du lemme, présenté plus haut. Enfin, nous évoquons les termes fréquents dans l'ensemble du discours, mais absents dans la section étudiée.

Les schémas ainsi obtenus ont été abstraits pour construire une représentation générale de l'activité de recherche experte. Il importe d'enrichir un vocabulaire partagé et des éléments structurants communs propres à favoriser la prise de conscience mais aussi le partage de l'activité. En suivant les usages en informatique de modélisation, nous nous appuyons sur le modèle entité-association que nous enrichissons. Ce qui nous conduit à construire une ontologie de l'activité de recherche d'information à partir des éléments suivants :

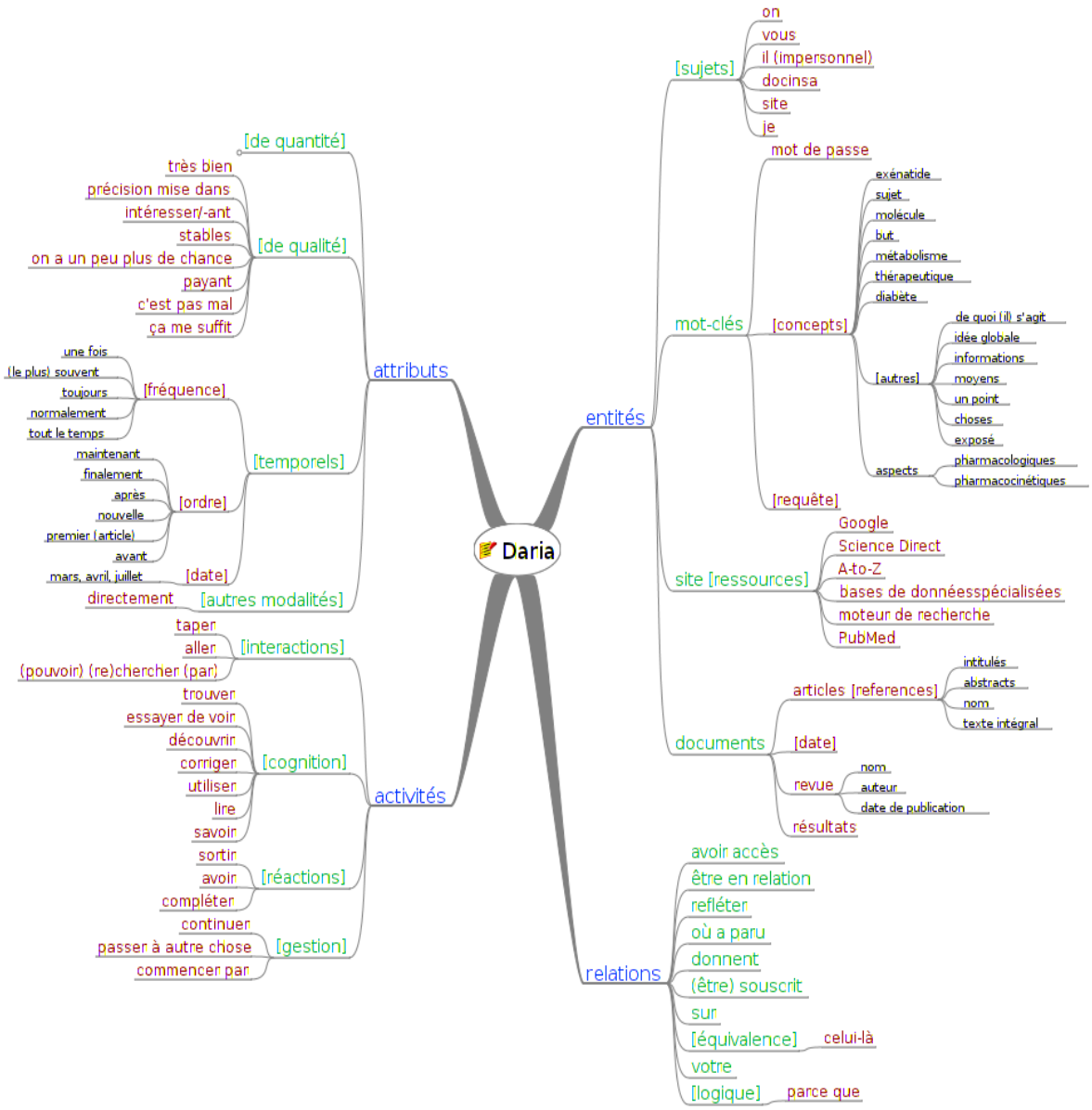
- des activités, qui comprennent les interactions (taper au clavier, aller vers...), des activités cognitives (trouver, lire...), des réactions à un événement de la machine (obtenir un résultat) et des activités d'organisation ou de gestion (commencer, continuer, passer à autre chose, etc.) ;

- des entités désignées nominativement ou connotées, acteur ou objet de l'action, qui comprennent les sites visités (les lieux), les résultats, les mot-clés, les acteurs, les documents, etc. ;

- des attributs ou jugements portant sur les entités, qui concernent des jugements qualitatifs (intéressant, payant), des jugements quantitatifs (pas beaucoup), des précisions temporelles (une fois, après) et d'autres modalités (directement, de manière médiée) ;

- des relations entre entités (« avoir accès », « refléter », « être équivalent »).

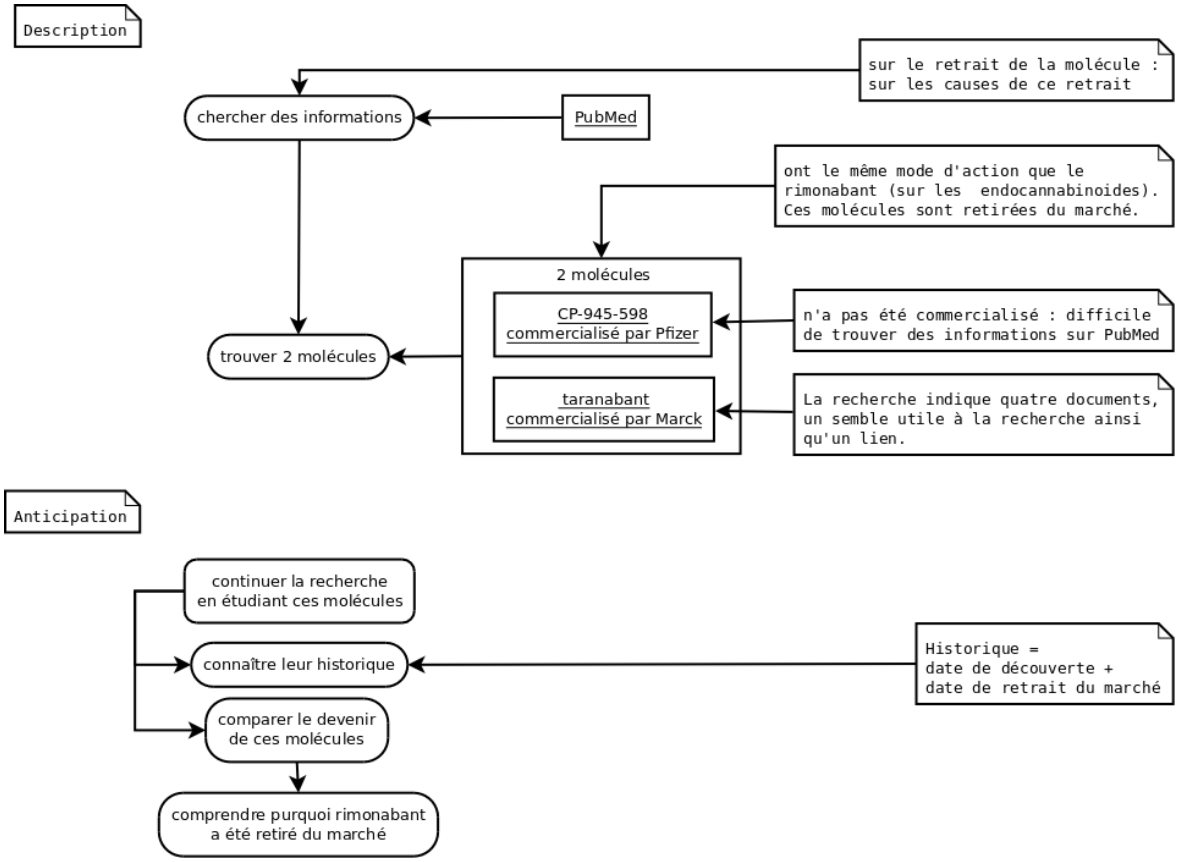
Ainsi tentons-nous de distinguer l'activité traçable du coté du système (actions de navigation, remplissage de formulaires) de l'activité cognitive (lire un document), enfin de l'activité d'organisation de la recherche (gestion et évaluation). Ces règles nous permettent d'analyser un discours conformément à cette ontologie. Nous donnons comme exemple le discours de Daria analysé à l'aide de cette ontologie.



Une fois que nous avons représenté les syntagmes dans le cadre d'une ontologie linguistique à partir du modèle entité/activité/attribut/relation, nous tentons de représenter l'activité de recherche d'information elle-même et non plus les discours. Pour cela, nous dessinons pour les différentes activités des étudiants un schéma inspiré du formalisme UML. Celui-ci a pour but de représenter dans le temps l'activité réalisée, en distinguant cette fois :

- les moments de l'interview : description ou anticipation, indiqués dans des rectangles cornés isolés, à gauche du schéma ;
- les activités d'interaction, représentées par des rectangles aux coins arrondis ;
- les entités concernées, représentées par des rectangles éventuellement imbriqués s'il est possible de déterminer des groupes ;
- des commentaires (rectangles cornés situés à droite de la page pour l'utilisateur, ou en haut à gauche pour l'observateur), qui décrivent les attributs des activités ou des entités ;
- des relations représentées par des flèches, qui indiquent soit un ordre temporel ou logique si elles relient des activités entre elles, soit l'objet rapporté par le commentaire ou l'entité.

Dans le schéma ci-dessous, nous donnons pour exemple la modélisation de l'activité décrite dans la note de Justine.



Dans la phase de description, Justine présente son activité en deux temps : la recherche des informations, et la récolte des résultats, chercher et trouver. L'objet de la recherche sont les deux molécules que l'on peut décrire de

différentes manières. Pendant la phase d'anticipation, Justine segmente son activité future selon des buts : connaître l'historique et le devenir des molécules, et comprendre pourquoi le Rimonabant a été retiré du marché. Cette histoire de l'activité de Justine, ce n'est donc pas le récit des actions passées ou à venir : c'est plutôt l'histoire de l'objet de la recherche, des concepts en construction. Ainsi, le centre de ce que décrit Justine on parlant de son activité n'est pas à proprement parler l'histoire de son activité, mais l'histoire du ou des concepts qu'elle construit. La conséquence que nous en tirons, c'est que favoriser l'apprentissage avec des traces, ce n'est pas forcément représenter une activité « en miroir », selon une logique narcissique. C'est plutôt représenter l'histoire du concept traité, construit progressivement. Ce qui suggère que l'on aurait tort de vouloir distinguer la représentation objective de l'activité de son annotation subjective : pour l'utilisateur, dans l'activité s'entrelacent le sujet et l'objet de la connaissance. Ainsi, tandis que les approches actuelles de l'activité de recherche d'information tendent à représenter cette activité de manière objective, « en miroir » (Li et al, 2010), cette recherche invite plutôt à se réapproprier l'histoire des concepts « en formation », tels qu'ils sont progressivement construits pendant l'activité de recherche d'information.

Partie 3 : Discussion sur les concepts de traces et d'activité

Dans l'objectif de saisir les représentations dans la recherche d'information en situation de réflexivité, nous avons tenté de saisir un ensemble de représentations utiles pour un projet de système d'apprentissage à base de traces. Pour discuter de cette perspective, nous nous intéressons à deux objections qui peuvent nous être faites. La première, c'est que notre étude, qui utilise des outils syntaxiques et de modélisation, oppose deux univers hétérogènes, celui des mathématiques et celui du travail humain. La seconde, c'est que la théorie des traces est une théorie informatique qui ne peut supporter d'analogie avec l'observation directe. Ces deux objections s'appuient sur une thèse implicite de l'hétérogénéité de la sphère technique et de l'activité d'apprentissage.

D'une part, nous devons relativiser l'ambition de ce travail. Certes, la lemmatisation peut être source d'ambiguïtés ou les transcriptions à l'origine d'infidélités ; cependant, nous avons utilisé diverses techniques de contrôle, comme une relecture critique des entretiens, le retour à la vidéo, et le recours à l'expertise de l'observateur et à son expérience en tant que formateur à la recherche d'information. Et surtout, la réalisation de « modèles » n'a pas pour but de décrire une population experte, mais de fournir des éléments d'information utiles. Ainsi, l'étude d'un échantillon réduit d'étudiants n'a pas pour but de décrire une population large d'utilisateurs, mais de trouver des éléments « en micro » qui rendent possible la mise en place d'un système à base de traces. Cette approche « en micro » permet de rompre avec « la dite « mémoire collective », simple discours officiel et normé sur le passé » (Boursier, 2002), et de « saisir une réalité plus profonde » (Ginzburg, 1989) pour renouer avec le possible d'un système d'apprentissage à venir. En effet, une trace peut faire sens pour un usager et non pour un autre : il importe de proposer un système souple, qui ne préjuge pas trop de ce qui sera utile à l'usager.

D'autre part, nous soutenons que théorie de l'activité et modélisation informatique ne sont pas incompatibles. Notre réflexion sur l'activité s'inspire des travaux des ergonomes et des psychologues du travail dans la lignée de (Vytgotski, 1997). Tandis que « l'action a un début et une fin repérables » et identifiables, soumise à une raison et à une décision, l'activité est « un élan de vie et de santé sans borne prédéfinie » qui croise tout ce que nous avons tendance à nous représenter séparément : corps et esprit, individuel et collectif, privé et professionnel, faire et valeurs, etc. (Schwartz, 2003, p. 1). En ce sens, l'activité de recherche d'information ne saurait donc être réduite à un schéma. On sait déjà que, pour un chercheur d'information, les émotions ont un rôle fondamental à jouer dans la performance éducative (Kuhltau, 1993). Pourtant, représenter schématiquement notre utilisation d'un ordinateur peut être utile. (Ollamier-Beldame, 2006) a montré que la visualisation des traces d'activité peut favoriser l'apprentissage. À condition de bien comprendre que l'activité humaine ne se donne jamais en entier, ni ne se divise en tâches ou actions descriptibles. Elle se donne plutôt sous forme de traces, de manière complexe. D'où notre recours à une théorie de la trace d'activité mêlant une dimension subjective, émotionnelle ou cognitive, et une dimension objective ou technique, faites de « modèles » qui ont leur importance parce qu'ils « pointent » l'activité de manière pertinente. Cependant, ils ne sont qu'une représentation possible parmi d'autres. Les modèles sont aussi des traces possédant un certain niveau d'abstraction.

Ainsi la trace d'activité n'existe pas isolément, elle est toujours la trace de quelque chose. En tant que mémoire de l'activité, le concept de trace est utilisé par de nombreuses disciplines en des sens multiples : informatique, mathématiques, médecine, électronique, linguistique, psychologie, philosophie, droit, sciences physiques... Ce qui unit ces disciplines, c'est cette « présence dans l'absence » (Derrida, 1967), le fait qu'elle renvoie à autre chose qu'elle même. On ce sens, on distinguera la trace d'autres concepts similaires (Serres, 2002) :

- L'indice, par exemple la fumée qui est l'indice du feu, entretient un rapport causal entre un signifié et un signifiant, mais n'a pas l'intention de signifier (Peirce, 1978) ;
- la piste est une série d'indices qui indiquent le parcours d'un objet ou d'un être vivant dans l'espace physique ;
- l'empreinte renvoie directement à l'identité d'un être vivant ;
- enfin la marque renvoie à l'identité de objet désigné par un être vivant.

Ainsi, tous les objets ne sont pas nécessairement des traces. Par exemple, la limite peut-être une trace dans la terre s'il s'agit de préparer les fondations d'une maison. En revanche les contours de la maison une fois construite ne sont pas une trace car ils ne renvoient qu'à l'existence de la maison, non à son absence. Dans le cas de cette étude sur la recherche d'information, nous avons donc distingué ici :

- des traces numériques : saisies à partir d'évènements numériques ;
- des traces écrites : rédaction de notes ;
- des traces orales : enregistrement de discours, interviews ;
- des traces visuelles ou sonores : enregistrements vidéos.

La trace n'est donc pas ce que nous désignons comme trace, mais ce que nous décidons de percevoir comme tel en tant qu'elle nous permet de pointer quelque chose qui n'est plus. Ce pointage est à l'origine d'un savoir. En ce sens, les traces d'interaction sont bien des inscriptions de connaissance, comme l'expliquent (Laflaquière, Prié, Mille, 2008). Les traces cognitives, dites aussi traces mnésiques ou engrammes, comportent donc des dimensions liées à l'histoire, mais aussi au fantasme et à la sensation (Roussillon, 2003). Traces numériques et traces mnésiques ne sont donc pas à opposer, mais à articuler ensemble. Elles articulent l'objet de la connaissance et le sujet qui connaît, rendant indissociables ces deux aspects de l'activité de recherche d'information.

Conclusion

Afin d'améliorer l'apprentissage lors de la recherche d'information experte par la réflexivité, nous avons réalisé plusieurs observations visant à saisir la manière dont de jeunes chercheurs se représentent leur activité de recherche d'information. L'analyse des interviews réalisées permet de fournir un modèle général, mais aussi une série de schèmes ou de découpages utiles pour construire un système à base de traces pour l'apprentissage. Elle montre aussi que les représentations ne portent pas sur l'histoire de l'activité réalisée, mais sur l'histoire de la représentation des concepts étudiés. Nous supposons donc qu'une représentation « objective » de l'activité de recherche d'information n'aide que faiblement l'utilisateur. Nous discutons alors les concepts de trace et d'activité dans la perspective de mieux articuler les dimensions objectives et subjectives de la connaissance. Sur le plan théorique, nous plaidons en faveur d'un concept interdisciplinaire de trace et sur une conception de l'activité compatible avec l'idée de modélisation. Sur le plan pratique, pour construire un système d'apprentissage à base de traces, nous suggérons de ne pas séparer les outils de représentation des outils d'annotation, afin de favoriser cette articulation heuristique du sujet et de l'objet dans la recherche d'information experte.

Remerciements

Je remercie vivement pour leur soutien, écoute ou conseils tous les membres de l'équipe SILEX du LIRIS, en particulier Julien Laflaquière, Alain Mille, Yannick Prié ; les membres du laboratoire ELICO, en particulier Céline Brun-Picard, Ghislaine Chartron, Jean-Philippe d'Erceville, Gérard Régimbeau, Talal Zouhri ; l'équipe pédagogique de DocINSA, en particulier Myriam Goutte et Laurent Quinson ; les étudiants en quatrième année de pharmacologie, en particulier Agnès, Alix, Audrey, Daria, Justine, Marine, Quinlan, Sylvain ; ma famille et mes proches, surtout Mélisande, Michèle et Morgane.

Bibliographie

ADORNO, Theodor. W., FRENKEL-BRUNSWIK, Else, LEVINSON, Daniel J., SANFORD, R. Nevitt (1950), *The authoritarian personality,* New York : Harper and Row.

ARNAUD, Michel, MERZEAU, Louise (2009), Traçabilité et réseaux, *Hermès* n° 53, Paris : CNRS éditions.

BLANCHET, Alain, GOTMAN, Anne (2005), *L'entretien*, Paris : Armand Colin.

BOURSIER, Jean-Yves (2002), « La mémoire comme trace des possibles », *Socio-anthropologie*, n°12, 2002, mis en ligne le 15 mai 2004, [en ligne] http://socio-anthropologie.revues.org/index145.html , page consultée le 21 juin 2010.

CLOT, Yves, FAÏTA, Daniel, FERNANDEZ, Fernandez, SCHELLER, Livia (2001), « Entretiens en autoconfrontation croisée: une méthode en clinique de l'activité », Éducation permanente, no. 1, 146, pp. 17–25.

DAVALLON, Jean, NOEL-CADET, Nathalie, BROCHU, Danièle (2003), « L'usage dans le texte : les « traces d'usage » du site Gallica », *Lire, écrire, réécrire, Objets, signes et pratiques des médias informatisé*, sous la dir. de SOUCHIER Emmanuel, JEANNERET Yves, LE MAREC Joëlle, p. 47-89, Paris : Bibliothèque publique d'information – Centre Georges Pompidou.

DERRIDA, Jacques (1967), *De la grammatologie*, Paris : Editions de Minuit.

GINZBURG, Carlo (1989), *Mythes, emblèmes, traces. Morphologie et histoire*, Paris : Flammarion.

GUITTET, André (2003), *L'entretien : techniques et pratiques*, Paris : Armand Colin.

HEIDEN, Serge, MAGUÉ, Jean-Philippe, PINCEMIN, Bénédicte (2010) - « TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement », in Sergio Bolasco, Isabella Chiari, Luca Giuliano (eds), *Statistical Analysis of Textual Data -Proceedings of 10th International Conference JADT 2010*, Edizioni Universitarie di Lettere Economia Diritto, Rome, 9-11 juin 2010, Sergio Bolasco (éd.), [en ligne] http://icar.univ-lyon2.fr/membres/bpincemin/biblio/pincemin_al_jadt10.pdf , page consultée le 15 juin 2010.

HULIN, Thibaud (2009), Bash : outils pour l'analyse de données, [en ligne] http://traces.toile-libre.org/blog/index.php?post/2009/09/03/Bash-%3A-outils-pour-l-analyse-de-donn%C3%A9es, page consultée le 9 septembre 2009.

KOYRÉ, Alexandre (2003), *Du monde clos à l'univers infini*, Paris : Gallimard.

KUHLTHAU, Carol Collier (1993), *Seeking meaning : a process approach to library and information services*, London : Libraries Unlimited.

LAFLAQUIÈRE, Julien (2009), *Conception de système à base de traces numériques pour les environnements informatiques documentaires*, thèse de doctorat soutenue le 8 décembre 2009, dir. Y. Prié, Université de Lyon I, [en ligne] http://tel.archives-ouvertes.fr/EC-LYON/tel-00471975/fr/, page consultée le 11 juin 2010.

LAROSE, François, JAILLET, Alain (2009), *Le numérique dans l'enseignement et la formation, analyses, traces et usages*, Paris : L'Harmattan.

LI, Ian, NICHOLS , Jeffrey, LAU, Tess, DREWS, Clemens, CYPHER, Allen (2010), « Here's What I Did : Sharing and Reusing Web Activity with ActionShot », *CHI 2010: End-User Programming II*, April 10–15, 2010, Atlanta : GA, USA.

MALINOWSKI, Bronislaw (1989), *Les Argonautes du Pacifique occidental*, Paris : Gallimard.

MANNING, Christopher D., RAGHAVAN, Prabhakar, SCHÜTZE, Hinrich, (2008), *Inroduction to information retrieval*, New-York : Cambridge university Press.

OLLAGNIER-BELDAME, Magali (2006), *Traces d'interactions et processus cognitifs en activité conjointe : Le cas d'une co-rédaction médiée par un artefact numérique*, thèse de doctorat soutenue le 6 décembre 2006, dir. A. Mille, Université de Lyon.

PEIRCE, Charles (1978), *Ecrits sur le signe*, Paris : éditions du Seuil.

LAFLAQUIÈRE, Julien, PRIÉ, Yannick, MILLE, Alain (2008), *Ingénierie des traces numériques d'interaction comme inscriptions de connaissances*, [en ligne] http://hal.archives-ouvertes.fr/docs/00/38/16/18/PDF/183-195.pdf, page consultée le 9 septembre 2009.

ROBERTSON, Durant Waite Jr. (1946), « A Note on the Classical Origin of "Circumstances" in the Medieval Confessional », *Studies in Philology*, 43:1:6-14, janvier 1946.

ROUSSILON, René (2003), « Historicité et mémoire subjective. La troisième trace », *Cliniques Méditerranéennes*, 67, no. 2003/1, pp. 127–144.

SCHWARTZ, Yves (2003), *Travail & Ergologie*, Toulouse : Octarès.

SERRES, Alexandre (2002), « Quelle(s) problématique(s) de la trace ? », *Traces et corpus dans les recherches en SIC,* Séminaire du CERCOR du 13 décembre 2002, [en ligne] http://archivesic.ccsd.cnrs.fr/sic_00001397.html, page consultée le 9 septembre 2009.

SIMMEL, Georg (1981), *Sociologie et épistémologie*, Paris : Presses universitaires de France.

TAMINE-LECHANI, Lynda (2009), *De la recherche d'information orientée système à la recherche d'information orientée contexte : Verrous, contributions et perspectives,* Mémoire de synthèse présenté en vue de l'obtention de l'habilitation à diriger des recherches, soutenue le 25 novembre 2008, Toulouse : IRIT - Institut de recherche en informatique de Toulouse, [en ligne] http://tel.archives-ouvertes.fr/tel-00355842/, page consultée le 11 juin 2010.

TRICOT, André, ROUET, Jean-François (1998), *Les hypermédias : approches cognitives et ergonomiques, Hypertextes et hypermédias*, Paris: Ed. Hermès.

VYTGOTSKI, Lev Semenovitch (1997), *Pensée et langage*, Paris : La Dispute.

WEBER, Max (1965), *Essais sur la théorie de la science*, Paris : Plon.