

Apprentissage et représentations de la recherche d'information experte avec des traces d'activité

Article inédit. Mis en ligne le xx xxxxx 2009.

Thibaud Hulin

Thibaud Hulin est docteur en philosophie de l'Université de Provence, ancien allocataire moniteur formé à l'Institut d'Ergologie. Après avoir été A.T.E.R. À l'INSA de Lyon en Sciences de l'Information et de la Communication, il est chercheur au LIRIS (Laboratoire d'InfoRmatique en Image et Systèmes d'information) / CNRS UMR 5205 / INSA de Lyon / Université Claude Bernard Lyon I. Ses recherches portent initialement sur la gestion des connaissances et l'histoire de l'information et de la communication.

Plan

Résumé

Introduction

Partie 1 : Méthodologies

Préparation de l'enquête

Objectifs de l'enquête

Hypothèses

Moyens

Conditions de validité et cadre épistémologique

Conception de l'enquête

Population et échantillon

Modes d'accès aux interviewés

Organisation des entretiens

Méthodes d'analyse des résultats

Analyse des significations explicites et implicites avec l'indice d'usage

Partie 2 : Résultats, analyse des discours

Synthèse des résultats

Modélisation des activités de recherche d'information

Pistes pour améliorer la réflexivité de l'activité

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

Conclusion

Bibliographie

RÉSUMÉ

Afin d'assister la recherche experte d'informations, nous proposons de renforcer la réflexivité par la réappropriation de traces d'usages. Pour cela, il est nécessaire de mieux cerner les limites des chercheurs d'information quant à leur capacité à revenir sur leur propre activité. Les résultats que nous présentons sont le fruit d'une expérience menée auprès de jeunes pharmacologues de l'INSA de Lyon. L'analyse de leurs discours met en évidence leur expertise, mais aussi des difficultés pour réutiliser ou partager leur activité. Nous mettons ici à jour les problèmes que posent l'évaluation de la recherche experte. Nous fournissons enfin un modèle pour catégoriser la réflexivité de la recherche d'information. L'enjeu est de pouvoir construire un système expert à base de traces et des formations pour renforcer la réflexivité de cette activité.

INTRODUCTION

Les études sur la recherche d'information se scindent volontiers en deux types: ou bien elles s'intéressent aux systèmes experts et aux algorithmes (Manning, 2008) ; ou bien elles portent sur la psychologie de l'utilisateur et tentent de modéliser l'activité de recherche (Tricot & Rouet, 1998). Ainsi, peu de systèmes experts permettent la réutilisation des données complexes issues de l'activité de recherche elle-même. Par exemple, les systèmes de suggestion interviennent seulement au moment d'entrer une nouvelle requête, et les raisons pour lesquelles tel mot-clé apparaît sont opaques. Pour enrichir les systèmes de recherche, il faudrait donc intégrer des traces d'activité précises qui fassent sens pour l'utilisateur, qui permettent à la fois de décrire, d'évaluer et d'anticiper la recherche. En sciences de l'information et de la communication, ce problème a été posé par exemple 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). Dans notre thématique de la recherche d'information, des utilisateurs experts envisagent le passé et le futur de leur activité, du point de vue descriptif et évaluatif ; enfin ils tentent de transmettre une partie de leur activité, en produisant ici de nouvelles traces. Dans cette étude, nous présentons dans un premier temps les résultats d'une expérience réalisée auprès de jeunes pharmacologues de l'INSA de Lyon. Dans un second temps, nous envisageons les forces et les limites de ces représentations et du rapport au temps de la recherche. Nous proposons un système de catégories pour concevoir un système à base de traces. Pour finir, nous tentons d'apporter quelques éclaircissements aux concepts d'activité et de trace pour la recherche d'information.

PARTIE 1 : MÉTHODOLOGIES

Préparation de l'enquête

Objectifs de l'enquête

Le but de notre démarche est d'analyser les représentations des chercheurs d'information dans le but de leur permettre d'effectuer un retour heuristique sur leur activité. Pour cela, notre expérience interroge quatre types de retour sur l'activité :

- la description : quels sont 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 sur la démarche et les résultats, les éléments pertinents pour cela ?
- 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 ?

Hypothèses

Malgré la qualité de la recherche, l'expérience et les formations dispensées aux experts en recherche d'information, leur représentation de l'activité reste imprécise et instinctive, donc insuffisamment exploitée. L'anticipation d'une

activité ne corresponde pas forcément à la description de cette activité. Pour éviter ce que les ergonomes ont appelé l'écart entre la situation prescrite et la situation réelle (Clot, 2001), les chercheurs d'informations accomplissent une double ruse (*métis* en grec : stratégie, habileté) :

- ils évaluent les résultats plutôt que sur les moyens d'y parvenir ;
- ils cultivent une représentation floue de leur activité.

Au cours de notre étude, il importera de déterminer les éléments qui permettront de pallier à ces insuffisances, tels qu'ils puissent être intégrés dans un système d'aide à la recherche d'information.

Moyens

Nous avons voulu savoir, à l'aide de questions ouvertes, de quelle manière les utilisateurs décrivent leur activité. Un logiciel enregistre les actions de l'utilisateur à l'aide d'une vidéo. Les personnes doivent constituer une bibliographie sur un thème donné.

Tandis que le questionnaire limite les possibilités d'apprendre des choses inattendues par le concepteur, 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.

Ces entretiens visent à constituer une source d'information principale. Les données produites seront confrontées à des hypothèses qui forgent des représentations individuelles, puis un modèle socio-culturel de la recherche d'information.

Conditions de validité et cadre épistémologique

Concernant notre méthode, nous avons suivi les conseils généraux de (Blanchet, 2005). Plus particulièrement, étant donné notre double statut d'observateur et d'enseignant, notre approche se situe 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 » mais d'« attentes » : des anticipations nourries subjectivement à propos de comportements (rationalité subjective) et sur la base d'expériences valables (rationalité objective) (Weber, 1965, p. 334).

Nous nous intéressons ici aux systèmes de représentations, c'est-à-dire aux pensées construites, aux idéologies comme « organisation d'opinions, d'attitudes et de valeurs » (Adorno, 1950, p. 2), une interprétation de la réalité prise pour la réalité, une articulation impensée entre l'être en le devoir-être, ce qui est décrit et ce qui est prescrit. Par notre analyse, nous suivons un processus d'objectivation qui assure le passage du registre procédural (les savoir-faire) au registre déclaratif (les savoir-dire).

Conception de l'enquête

Population et échantillon

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

intéressé 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, très connaisseurs du web, 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, le corpus que nous présentons concerne des interviews réalisés avec 5 étudiants issus de ces deux groupes. Nous ne présentons ici que les résultats issus d'interviews réalisées en langue française. Plutôt qu'une analyse statistique générale censée montrer des tendances, nous avons donc un nombre réduit d'interviewés, ce qui doit nous permettre de réaliser une analyse très précise et de haute qualité sur l'activité concrète.

Modes d'accès aux interviewés

Pour connaître leur activité de recherche, nous avons lu les compte-rendus de recherche documentaire et assisté à leur présentation orale. Nous avons ensuite organisé des séances d'observation individuelles d'une à deux heures.

Le lieu des séances est la bibliothèque de l'INSA où les conditions sont favorables à la concentration des étudiants. Cependant l'espace physique rappelle que l'étudiant est en position d'apprentissage, lequel peut se sentir dominé par l'observateur qui fait parti d'un corps enseignant. Nous avons donc fait le maximum pour maintenir une tension heuristique entre la dimension d'apprentissage, qui suppose traditionnellement un maître (*dominus*) et un élève (*discipulus*), et un climat convivial, qui suppose un sentiment de liberté d'action et de parole.

Organisation des entretiens

Pour neutraliser les conditions déformantes de l'interview, nous nous sommes appuyé sur les outils fournis lors des formations à la recherche d'information, en excluant tout commentaire. Le guide d'entretien présente la séance selon cinq étapes :

1. Accueil et rappel du protocole
2. Première session de recherche d'information, dix minutes
3. Interview sur ce qui a été réalisé et sur ce qui va l'être
4. Seconde session de recherche d'information, dix minutes
5. Nouvelle interview sur ce qui a été réalisé
6. Rédaction d'une note d'une dizaine de lignes, en précisant éventuellement le destinataire (soi-même ou une personne de l'équipe) pour poursuivre la recherche.

Pendant les interviews, nous posons des questions ouvertes, de circonstance. Pour cela, nous nous appuyons sur l'outil QQQCP (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 relance 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'interviewer 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 ont été autorisées mais réduites au minimum, l'outil QOQCP permettant de structurer l'entretien et assurer la communication sans qu'il soit nécessaire de mettre en place des stratégies d'intervention ou modes de relances plus lourds (paraphrase, commentaire, etc.).

Enfin, nous avons opté pour une relation dite de « sociabilité amicale » (Simmel, 1981) qui favorise une convention d'égalité entre protagonistes en respectant les convenances et en favorisant la coopération. Ainsi avons-nous mis en place un cadre contractuel de la communication, qui définit les conditions de confidentialité, de publication des résultats, et de ce que l'interviewer peut dire et ne pas dire.

Méthodes d'analyse des résultats

Afin de prendre appui sur des critères les plus objectifs possibles, nous nous appuyons sur un ensemble d'outils informatiques pour l'analyse textuelle par la réduction lemmatique et la comparaison des occurrences. Nous transcrivons le texte enregistré vers un texte écrit que nous mettons en forme selon des règles précises. Puis nous effectuons des transformations sur le code ainsi obtenu jusqu'à obtenir des classes d'occurrences et de discours. Ces transformations éclairent et retardent l'interprétation, renforçant ainsi la rigueur scientifique de notre démarche.

Analyse des significations explicites et implicites avec l'indice d'usage

L'analyse des occurrences permet de mettre en évidence les concepts clés des discours. L'analyse des occurrences en référence à l'ensemble du corpus des interviews est ce que nous appelons l'analyse explicite des discours.

Cependant, l'analyse explicite ne permet pas de mettre en évidence le fait qu'un terme est employé uniquement dans cette section, ce qui devrait attirer l'attention de l'analyste. Pour analyser le discours implicite, nous comparons alors la fréquence d'un lemme dans le corpus, et sa fréquence dans la portion de texte étudiée. C'est ce que nous appelons le calcul de l'indice d'usage.

Pour choisir les mots les plus importants, on observe pour chaque mot :

- le rapport, dans une Section, entre le nombre d'occurrences d'un Mot (sm) et, dans cette même Section, le nombre Total de mots (st) ;
- le rapport, dans une Section, entre le nombre d'occurrences d'un Mot (sm) et, dans un Corpus, le nombre Total de mots (ct).

On prend alors la moyenne de ces deux rapports et l'on multiplie par 100 pour obtenir un pourcentage, ce qui nous donne la formule suivante :

$$Indice = \frac{\frac{sm}{st} + \frac{sm}{ct}}{2} * 100$$

L'analyste effectue alors des hypothèses interprétatives en suivant progressivement les indications fréquentielles qu'il recueille. Pour chaque section du code (au sens d'interview retranscrite), nous analysons d'abord le discours dominant, c'est-à-dire les mots dont l'indice est le plus haut, qu'il organise par thème. Puis il analyse le discours silencieux, c'est-à-dire les mots fréquents dans le corpus mais absents de la section du discours étudiée. Il commence par une réflexion sur l'ensemble du corpus, puis étudie les sections regroupés par thème ou par genre. Enfin, il étudie les discours des individus.

Plutôt qu'une moyenne, on aurait pu insister davantage sur le poids du corpus si l'on juge que les écarts sont négligeables, ou sur le poids du texte si l'on pense que la section est indépendante du reste du corpus. On peut penser que dans un même discours, l'interviewé peut passer d'un discours convenu à un discours original. L'indice d'usage ne fait donc qu'indiquer des écarts de discours : pour apprécier pleinement ces différences, l'analyste relit, réécoute ou visionne à nouveau l'interview ou l'activité enregistrée. L'indice d'usage offre donc une méthode d'analyse à partir de repères précis : il ne remplace pas la qualité de l'interprétation, mais lui fournit une méthode (*methodos*, le chemin) : une piste balisée pour l'analyse rigoureuse.

L'ensemble des calculs ainsi obtenus est disponible en (Hulin, 2009).

PARTIE 2 : RÉSULTATS, ANALYSE DES DISCOURS

Dans un premier temps nous avons tenté de déterminer les registres d'énonciation dans le but d'identifier des stratégies discursives pour structurer l'analyse finale des entretiens. La trace brute enregistrée par le microphone connaît alors une série de transformations, automatisées ou non, et qui ont pour but de faire sens pour l'analyste.

- Enregistrement : nous avons réalisé un corpus de textes issu des interviews retranscrites, ce qui nous donne une trace écrite brute.
- Codage : transformation de la trace orale vers la trace écrite.
- Nettoyage : nous produisons un code lisible par une machine et par un humain. Par exemple : suppression des commentaires et des questions, division en sections, homogénéité de l'écriture de certains termes, etc.
- Réduction des mots les plus fréquents à partir d'un dictionnaire de fréquence.
- Lemmatisation : la réduction des mots à leurs racines significatives, les lemmes à partir d'un dictionnaire de lemmes.

Le code ainsi obtenu a subi un ensemble de transformations basées sur les expressions régulières. Nous avons utilisé les langages bash, perl et python. Le code source de ces scripts sont publiés sur un espace web (Hulin, 2009) avec

les codes sources de l'interview, les résultats et les dictionnaires qui ont servis aux calculs.

Pour l'analyse des résultats, nous partons des résultats obtenus pour analyser successivement :

- l'analyse d'ensemble, l'esprit général des interviewés
- les différents types d'interviews : description 1, anticipation, description 2, évaluation
- les différences d'approches par personnes
- les écrits

Pour chacune de ces parties, nous citons entre parenthèses un 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, un pourcentage arrondi à l'unité, donc le calcul est indiqué plus haut. Enfin nous évoquons les termes fréquents dans l'ensemble du discours, mais absents dans la section étudiée. Aucun chiffre n'est donc ici donné.

Synthèse des résultats

De manière générale, les descriptions montrent que la théorie pédagogique enseignée est comprise et appliquée ; ils ont d'ailleurs obtenu une note honorable à leur travail de groupe. En revanche, la réflexivité sur l'activité, qui ne fait pas l'objet d'un enseignement, est appréciée de manière différente.

En effet, l'attention se porte beaucoup sur les résultats. Les connaissances sont construites après la lecture des textes et non pendant la navigation : peu de remarques sur les auteurs ou les concepts clés. Le vocabulaire portant sur les actions possibles est réduit, en particulier la logique de la navigation est rarement envisagée.

Les outils utilisés aussi sont rarement, les fonctionnalités avancées sont peu utilisées, et l'usage de booléens est rare ou limité. Les jeunes chercheurs d'information se concentrent sur certains types de documents, en général les articles scientifiques, mais n'interrogent pas toujours la pertinence de ce type de document, ou même l'intérêt de la revue : s'ils utilisent les bases de données spécialisées qui leur sont indiquées, ils leur font souvent confiance, sans jamais interroger le rapport entre le bruit et le silence.

La distinction entre évaluation et description n'est pas toujours bien claire. Lorsque l'évaluation est distinguée clairement, elle ne s'appuie pas sur des critères mesurables.

Comme les étudiants ont suivi la même formation, ils partagent un vocabulaire commun qui leur est rappelé au moment de l'interview. Cependant, passé ce stade, chacun déploie son propre vocabulaire qui se prête difficilement à une mise en commun. Même remarque pour la logique des actions telle qu'elle apparaît dans les notes.

Si l'idée de rédiger une note pour un collègue ou pour soi-même pose rarement un problème, celles-ci se réduisent parfois à des titres d'articles, ce qui correspond à un usage faible de la réflexivité.

Les problèmes que rencontrent les jeunes chercheurs tiennent beaucoup à l'accessibilité d'articles : ils utilisent alors bien les résumés comme le moyen

de savoir si un article est intéressant ou non, c'est-à-dire en fait pertinent. Sa fiabilité n'est pas interrogée. Enfin, le résumé contient en général la thèse de l'auteur et les grandes lignes de sa méthode : nous n'avons guère vu d'étudiants s'en satisfaire, alors que le résumé fournit déjà des pistes scientifiques exploitables.

Ainsi, l'étudiant se contente souvent d'une démarche statique : il ne structure ni ne planifie sa recherche, et l'on est en droit de penser qu'en l'absence d'un tiers, l'usage de la sérendipité doit être plus important encore.

La logique de la recherche, en terme d'étapes, de succession, voir de retour sur certaines étapes sous certaines conditions (boucle conditionnelle) n'est pas présente.

La conscience du temps se réduit et les recherches s'éternisent, ce qui est sans doute un problème courant pour tout chercheur.

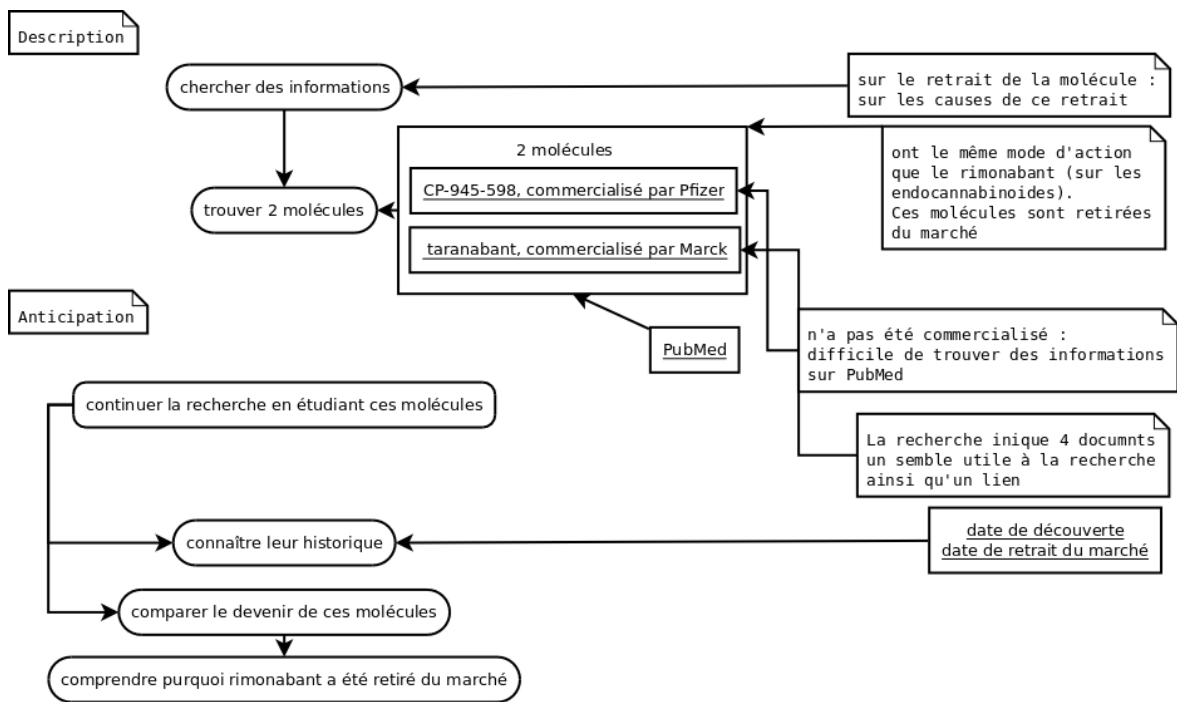
En raison du manque de temps, la recherche est donc souvent faite de manière intuitionnelle et rapide : pourtant, mal menée, mal planifiée, elle est le meilleur moyen de perdre son temps plutôt que d'en gagner.

Les recherches d'informations des étudiants remplissent leur contrat pédagogique, signé entre le formateur et l'étudiant, mais restent faiblement structurées. En effet, les outils pour la réflexivité dont disposent l'étudiant sont destinés à gérer les références bibliographiques. Si l'historique est présent, l'étudiant ne sait finalement pas comment l'utiliser. Il ouvre au mieux plusieurs onglets qui lui permettent d'effectuer un retour sur son activité. Afin d'organiser les concepts qui régissent la réflexivité de l'activité de recherche d'information, nous proposons de partir de l'activité concrète elle-même et de la généraliser en l'organisant en étapes.

En somme, certains des étudiants ont des qualités différentes ou complémentaire : l'un est davantage attentif à structurer sa recherche, l'autre justifie davantage ses étapes, le troisième fait appel à un vocabulaire plus précis... Il importe donc que les étudiants soient également formés à la réflexivité, condition indispensable de la collaboration.

Modélisation des activités de recherche d'information

Afin de modéliser l'ensemble des activités de recherche d'information, nous avons abstrait et représenté les différentes actions, commentaires et entités visés pour chacune des activités décrites. Nous avons alors dessiné un schéma, inspiré du formalisme UML, qui permette de montrer le dynamisme temporel de ces activités, en terme d'action (représentés par des cercles), de commentaires (rectangles cornés à droite pour l'utilisateur, à gauche pour l'observateur), d'entités visées (rectangles simples), de relations logiques (flèches) et de conditions (losanges). Nous donnons pour exemple la modélisation de l'activité décrite dans la note de Justine :



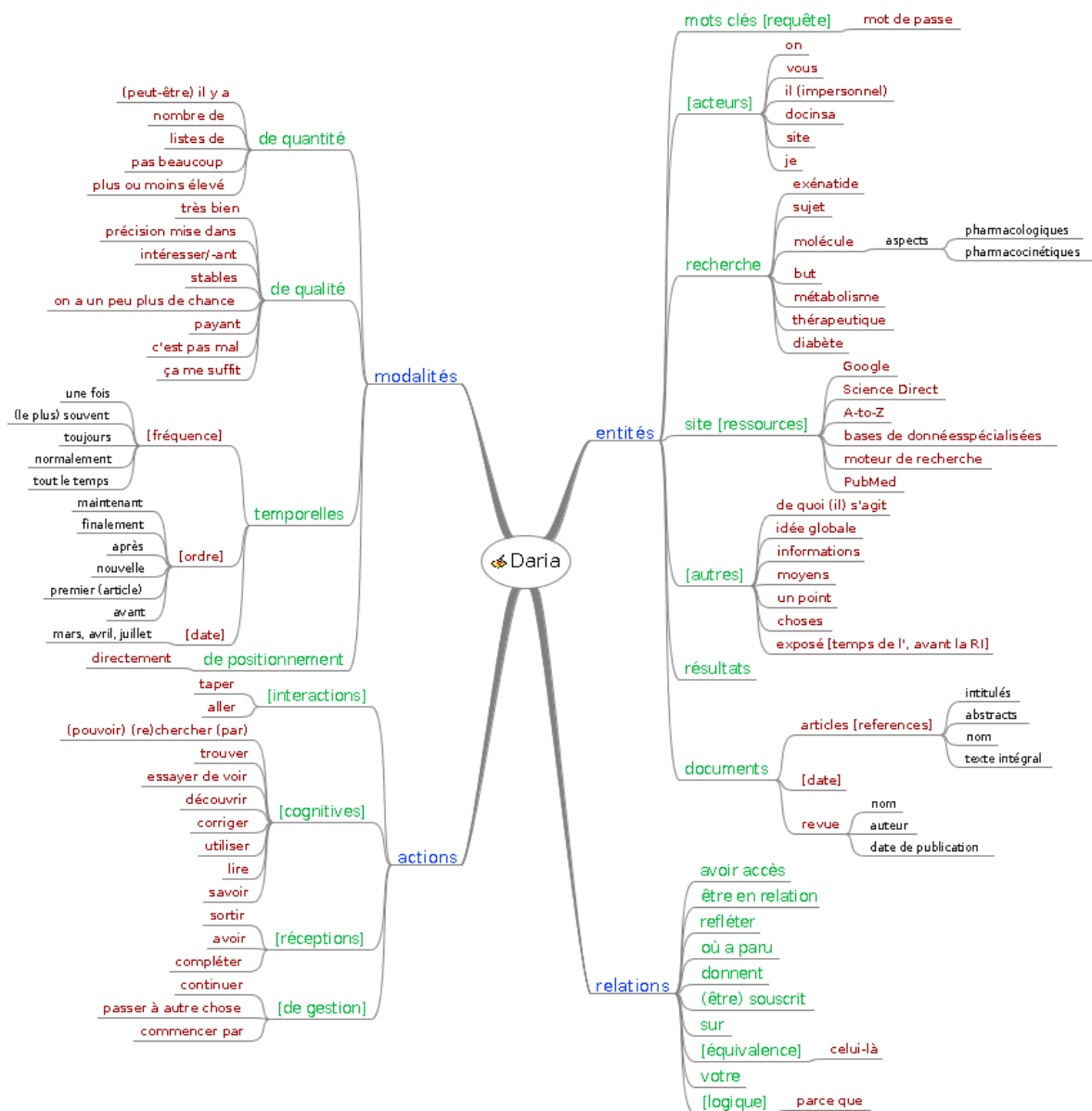
Pistes pour améliorer la réflexivité de l'activité

Les schémas ainsi obtenus ont été abstraits pour construire une représentation générale de l'activité de recherche experte. Que ce soit pour former les experts à la réflexivité de l'activité ou bien les assister par le biais d'un système expert, nous faisons les trois propositions suivantes :

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. Nous proposons de construire une ontologie de l'activité de recherche d'information à partir des éléments suivants :

- les actions, qui comprennent les interactions (taper au clavier, aller vers...), les activités cognitives (trouver, lire...), les actions de réception (télécharger un fichier) et les actions de gestion (commencer, passer à autre chose, etc.)
- les entités qui comprennent les sites (les lieux), les résultats, les mots-clés, les acteurs, les documents, enfin les autres entités.
- les modalités, qui concernent des jugements qualitatifs, quantitatifs, les précisions temporelles enfin les autres positionnements (de manière médiée, etc.)
- les relations (avoir accès, refléter, être équivalent...)

Nous donnons en exemple le discours de Daria analysé à l'aide de cette catégorisation.



Dans le cadre d'un usage à long terme, on remarquera que ces catégories sont compatibles avec celles fournies par Aristote (Aristote, 2001) sans cependant s'y restreindre, ce qui nous permet de viser une ontologie très solide, voire universelle :

Ontologie de l'activité de recherche d'information	Catégories classiques aristotéliennes
Entité	essence
Jugement quantitatif	quantité
Jugement qualitatif	qualité
Relation	relation
Entité - Site	Quelque part

Jugement temporel	À un moment
Modalité	Dans une position
Relation - avoir	avoir
Action	agir
Action - activité cognitive	pâtir

On peut sans doute discuter des termes et classifications adéquates, car le débat sur les catégories aristotéliennes n'est pas clos. Sans doute l'usage et les styles trancheront-ils.

Ainsi séparons-nous l'activité descriptible (navigation, remplissage de formulaires) de l'activité cognitive, enfin de l'activité évaluative (commentaires). Nous avons maintenant l'ensemble des éléments fondamentaux susceptibles de donner lieu à une formation ou à un système à base de traces.

PARTIE 3 : DISCUSSION SUR LES CONCEPTS DE TRACES ET D'ACTIVITÉ

La perspective d'une recherche d'information réflexive nourrie et se base à la fois sur deux concepts fondamentaux.

Le premier, l'activité, est inspirée des travaux des ergonomes et des psychologues du travail dans la lignée de (Vygotski, 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 corps et esprit, individuel et collectif, privé et professionnel, faire et valeurs, etc. (Schwartz, 2003, p. 1). En ce sens, modéliser une activité ne revient qu'à donner des repères qui engagent à de nouvelles activités. En aucun cas nous ne saurions ici dresser un mode d'emploi, une recette qui fournisse une liste d'actions distinctes ou idéales. Notre approche invite donc à une pédagogie de la réflexivité. Nous ne proposons donc pas un modèle ultime ; nous invitons plutôt le lecteur à poursuivre notre réflexion à l'aide des résultats obtenus. Cette perspective, qui donne le primat à l'activité, nous la qualifierons d'« ergologique ». Elle justifie pleinement l'immersion dans le terrain telle que nous l'avons définie pour saisir les différences fines et opérationnelles. On peut ainsi mieux admettre que description, anticipation et évaluation se mêlent dans l'activité réelle, tandis que ces dimensions sont clairement séparées dans une vision mécaniste en terme de tâches. On sait d'ailleurs que, pour un chercheur d'information, évaluer son activité est à l'origine d'émotions qui ont un rôle à jouer dans la performance (Kuhltau, 1993). Cependant, on a tout intérêt à mieux distinguer les genres à l'intérieur de l'activité, si l'on veut éclaircir et mieux distinguer les concepts en jeu.

Ainsi, le concept de trace est facilement sujet à de nombreuses confusions. Ne ferait-on pas mieux de le réduire à sa dimension technique, et considérer qu'une trace n'est que le résultats d'observés ou d'interactions ? En apparence, cet *a priori* semble aider l'informaticien qui n'aurait plus qu'à se concentrer sur la collecte des évènements numériques, pour retracer l'activité

de recherche. Pourtant, s'il on veut construire des systèmes qui soient utiles, on ne saurait perdre de vue le fait que ce qu'enregistre une machine n'est qu'une partie de l'activité, pas forcément la plus signifiante ou la plus fertile pour la réflexivité. Il importe alors, pour construire un système évolutif, de construire une architecture souple qui intègre des concepts complexes. En tant que mémoire de l'activité, le concept de trace est utilisée par de nombreuses disciplines en des sens multiples : informatique, mathématiques, médecine, électronique, linguistique, psychologie, philosophie, droit, sciences physiques, etc. Nous nous référons ici au concept philosophique de trace de Derrida comme « présence dans l'absence » (Derrida, 1967). On ce sens, on distinguera la trace d'autres concepts comme en (Serres, 2002) :

- l'indice, qui entretient un rapport causal entre un signifié et un signifiant (exemple : la fumée est l'indice du feu), sans intention de signifier (Peirce, 1978) ;
- la piste : série d'indices indiquant le parcours d'un objet ou d'un être vivant dans l'espace physique ;
- l'empreinte, qui renvoie à l'identité d'un être vivant ;
- la marque, qui 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. 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. En ce sens, les traces d'interaction sont bien des inscriptions de connaissance, comme cela est très clairement expliqué en (Laflaquière, Prié, Mille, 2008).

Les traces cognitives, dites aussi mnésiques ou engrammes, comportent donc des dimensions liées à l'histoire, mais aussi au fantasme et à la sensation (Roussillon 2003). Les traces réappropriées sont décontextualisées, mais elles ne peuvent se déraciner complètement de leur origine sémantique, sans quoi elles perdront tout intérêt.

Il importe donc de bien distinguer les espèces de traces ; on gagnera ainsi à relier ensemble des traces d'origines différentes pour enrichir nos activités d'apprentissage.

CONCLUSION

Les résultats de nos travaux fournissent un modèle de catégorisation de l'activité de recherche d'information, du type action/modalité/entité/relation (AMER). Cependant, étant donné son niveau important d'abstraction, nous recommandons de ne pas perdre de vue les activités observées qui sont à son

origine. Pour construire un système à base de traces utile à la recherche d'information, il faut pouvoir se référer aux problèmes liés à l'évaluation que nous avons observés. Enfin, nous avons insisté sur la nécessité de se référer à des concepts riches et complexes de trace et d'activité si l'on veut accompagner utilement l'apprentissage de la recherche d'information.

REMERCIEMENTS

Je remercie vivement pour leur soutien et conseils tous les membres de l'équipe SILEX du LIRIS, en particulier Alain Mille, Yannick Prié et Julien Laflaquière ; les membres du laboratoire ELICO, en particulier Geneviève Lallich-Boidin, Thierry Lafouge, Gérard Régimbeau, Ghislaine Chartron, Talal Zouhri, Jean-Philippe d'Erceville, Céline Brun-Picard ; l'équipe pédagogique de DocINSA, en particulier Myriam Goutte et Laurent Quinson ; les étudiants en 4e année de pharmacologie, en particulier Sylvain Frappart, Daria Markina, Audrey Vogt, Agnès Franco, Marine Croze, Alix Jouffret, Quinlan Liu, Justine Mulet ; ma famille et mes proches, surtout Michèle, Mélisande et Morgane.

BIBLIOGRAPHIE

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., SANFORD, R. N. (1950), *The authoritarian personality*, New York : Harper and Row.

ARISTOTE (2001), *Catégories*, trad. Richard Bodéüs, Paris : Les Belles Lettres.

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

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 informatisés*, 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.

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.

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

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

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

PEIRCE, Charles (1978), *Écrits 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, Jr (1946), « A Note on the Classical Origin of "Circumstances" in the Medieval Confessional », *Studies in Philology*, 43:1:6-14, janvier 1946, [en ligne] <http://www.jstor.org/stable/4172741a>, page consultée le 9 septembre 2009.

ROUSSILON, R., « 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.

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.