
Activités documentaires des usagers au sein de l'organisation : amélioration par la pratique d'annotation collective

Guillaume Cabanac* — **Max Chevalier*,****
Claude Chrisment* — **Christine Julien***

* Université de Toulouse, IRIT UMR 5505 CNRS
118 route de Narbonne, F-31062 Toulouse cedex 9

** Université de Toulouse, LGC ÉA 2043
IUT Paul Sabatier, 129 avenue de Rangueil, BP 67701, F-31077 Toulouse cedex 4
{Guillaume.Cabanac, Max.Chevalier, Claude.Chrisment, Christine.Julien}@irit.fr

RÉSUMÉ. Cet article traite des activités documentaires réalisées par les individus au sein d'une organisation. Ils sont actuellement contraints à maîtriser une kyrielle de logiciels qui sont spécifiques à une activité documentaire donnée. Les usagers monopolisent alors une charge cognitive importante. Au niveau de l'organisation, les informations pertinentes qui ont été trouvées au prix de coûteux efforts lors de ces activités ne sont toutefois pas partagées, constituant de fait un capital en sommeil à forte valeur ajoutée. Nous proposons une architecture multi-utilisateur et des processus fédérant et améliorant les activités documentaires sur le principe du donnant-donnant, favorisant l'enrichissement d'une activité à partir des autres activités documentaires. Cette proposition repose notamment sur les espaces personnels d'annotations des usagers qui permettent de capitaliser toute information détenue par l'organisation.

ABSTRACT. This paper deals with document-related activities that knowledge workers use to manage within their organization. They have to master various applications, each one being specific to a given document-related activity. This current situation implies a high cognitive load for the user. Regarding the organizational level, although relevant information was costly to find, it is poorly shared thus constituting a dormant capital. A multi-user architecture along with suited processes is presented. Our proposal intends to federate and improve document-related activities based on a win-win principle, fostering the improvement of an activity thanks to the other ones. This proposal makes extensive use of people's personal annotation spaces so as to capitalize on any information entering the organization.

MOTS-CLÉS : Activités documentaires, organisation, travailleurs du savoir, annotation.

KEYWORDS: Document-related activities, organization, knowledge workers, annotation.

1. Introduction et motivations

La démocratisation de l'informatique transforme sans cesse les activités des organisations modernes telles que les entreprises, les laboratoires de R&D et toute communauté, au sens large. Ballay (2002) souligne notamment le fait que nous sommes tous devenus des « travailleurs du savoir », des personnes qui travaillent principalement avec de l'information afin d'en produire. Une étude chiffrée indique que cette catégorie d'employés représentait 31 % de la main d'œuvre des États-Unis en 1995, cette proportion devant continuer de croître de manière significative au XXI^e siècle (Sellen *et al.*, 2003, p. 51). Feldman (2004) observe qu'ils passent de 15 % à 35 % de leur temps de travail en quête d'information. C'est une activité ardue pour de multiples raisons : difficulté de formulation des besoins, sources multiples contenant des informations partielles, fausses, périmées, etc. Selon cette même étude, une requête sur deux n'aboutit pas, ce qui en fait une activité définitivement peu rentable. Dans ce contexte, lorsque l'utilisateur trouve enfin des documents pertinents pour ses activités au prix de nombreux efforts, il les mémorise généralement dans son espace personnel d'information (EPI) afin d'y accéder facilement dans le futur (Jones *et al.*, 2007, p. 10). Les documents y sont généralement organisés hiérarchiquement, dans une arborescence de répertoires par exemple. Étant donné que les individus partagent un intérêt pour les activités de l'organisation, il existe des thématiques et des besoins se recouvrant partiellement voire totalement. Ainsi, l'ensemble de ces EPI constitue un capital informationnel structuré et à haute valeur ajoutée, car issu d'un effort des membres organisationnels lié aux activités qui les motivent. Paradoxalement, on remarque que ces difficultés individuelles sont accentuées au niveau macroscopique. En effet, les pépites d'informations contenues dans les EPI ne sont pas valorisées au niveau collectif : les efforts d'un individu ne sont pas capitalisés afin d'aider ses collègues car ses documents demeurent en sommeil dans son propre EPI. L'organisation est par conséquent doublement perdante quant à son capital informationnel, tant au niveau individuel qu'au niveau collectif.

Cet article identifie les problématiques auxquelles les usagers sont confrontés au cours des activités liées à l'information et à son support (le document) afin d'y apporter une solution pertinente et originale. La section 2 présente ces activités qui forment le cycle de vie du document, en distinguant les niveaux individuel et collectif, de façon à couvrir les problématiques préalablement évoquées. La section 3 détaille notre proposition : une architecture visant à améliorer la gestion de l'information en fédérant et exploitant les activités du cycle de vie du document (Cabanac *et al.*, 2008a). Au cœur de notre proposition figure la pratique commune d'annotation, car elle s'inscrit dans chacune des activités et représente une réelle plus-value apportée aux documents par et pour les travailleurs du savoir (Kidd, 1994). En effet, la trace d'une annotation sur un document reflète les efforts cognitifs de son créateur alors qu'il interagissait avec ce vecteur d'information : finalités de mémorisation, d'argumentation, de correction, etc. Outre cette architecture reposant sur l'annotation, une assistance personnalisée est offerte aux membres organisationnels grâce à six processus définis sur les principes du donnant-donnant et de la non-intrusion. Enfin, les résultats d'une activité documen-

taire sont exploités pour enrichir les autres activités, au bénéfice de l'individu comme du groupe. La section 4 présente les expérimentations réalisées pour valider l'architecture et les processus proposés. Elle décrit également le prototype implantant cette architecture pour en montrer la faisabilité. Ce dernier prend en charge les documents HTML du web en tant que source d'information exploitée par toute organisation de nos jours. Puis, la section 5 discute les implications de la fédération pour l'utilisateur avant de conclure par les pistes de recherche que nous souhaitons considérer à l'avenir.

2. Problématiques des activités documentaires dans l'organisation

Dans cet article, nous considérons les relations entre l'organisation, l'utilisateur en tant que membre organisationnel, ainsi que les documents électroniques qu'il manipule au quotidien pour réaliser ses activités. Or, pour toute organisation, la gestion des documents est un facteur de performance s'appuyant sur l'optimisation des diverses activités liées aux documents et, par extension, à l'information qu'ils contiennent. Sellen *et al.* (2003, p. 203) identifient six activités principales formant le cycle de vie du document. Nous y faisons référence par la suite grâce aux symboles ① à ⑥. Les membres organisationnels réalisent ces six activités individuellement ou collectivement, à l'aide de systèmes (logiciels) spécifiques :

- la recherche d'information ① est une tâche hautement cognitive pour l'individu. En effet, il doit savoir utiliser un outil de recherche, transcrire la représentation mentale de son besoin en une requête, consulter la kyrielle de documents retournés par le système et enfin interpréter les résultats afin d'en extraire les documents pertinents pour satisfaire son besoin initial. Cette activité repose majoritairement sur l'utilisation de systèmes *pull* : moteurs de recherche, signets publiés sur des sites de *social bookmarking* (Hammond *et al.*, 2005) tels que Delicious ou Connotea. Par ailleurs, Cabanac *et al.* (2008b) passent en revue plusieurs approches collectives visant à automatiquement tirer parti du groupe au profit de l'individu, et *vice versa*. Au final, l'activité de recherche d'information qui représente selon Feldman (2004) entre 15 % et 35 % du temps de travail des individus dans un contexte organisationnel n'est que peu rentable car la moitié des recherches échoue ;

- les activités de création, rédaction ② et finalisation ③ de documents sont principalement mises en œuvre grâce à des logiciels de traitement de texte. Il existe également des approches complémentaires comme les wikis qui permettent la rédaction collaborative asynchrone. D'autres approches encore rendent la rédaction synchrone possible : chaque rédacteur voit en temps réel les modifications des autres contributeurs (Swarts, 2004). Une évaluation quantitative de ces activités montre un faible rendement, en partie dû à l'inefficacité de l'activité ① : un nouveau rapport contiendrait en moyenne 90 % d'information recréée (Feldman, 2004) ;

- il existe principalement deux stratégies pour diffuser les documents ④. La stratégie *manuelle* repose notamment sur l'envoi de courriels, l'envoi d'annonces sur des listes de diffusion thématiques, ou la publication sur l'intranet de l'organisation. Ces actions représentent une charge cognitive importante pour l'utilisateur : pour chaque do-

cument à diffuser, il est confronté à l'identification des destinataires potentiellement intéressés au sein de l'organisation, ainsi qu'au choix de la rubrique adéquate dans l'intranet. Afin de soulager les usagers de tels efforts rédhibitoires, la stratégie *automatique* comprend les moteurs de workflow, les systèmes *push* de recommandation (Montaner *et al.*, 2003), ainsi que des approches fondées sur l'élicitation de réseaux sociaux (Zhang *et al.*, 2005). Cette seconde stratégie souffre de limites liées à la création des profils pour représenter fidèlement les ressources et les besoins des usagers, ainsi qu'à l'évolution de ces profils (Chevalier *et al.*, 2008) ;

– l'exploitation de documents ⑤ correspond notamment à l'activité de lecture active définie par Adler *et al.* (1972, p. 4). Celle-ci permet à l'individu de s'approprier le contenu d'un document grâce à la réflexion critique qu'il mène durant sa lecture. Cette tâche implique de la part du lecteur la création d'annotations informelles telles que des commentaires, des résumés, des marques diverses, etc. Kidd (1994) indique que ces marques, qui lui permettent de modifier le document afin d'extérioriser sa pensée, seront par la suite utiles pour générer de l'information. Sellen *et al.* (2003, p. 63) rapportent également que la pratique d'annotation est importante pour les travailleurs du savoir, car elle leur permet de structurer et d'organiser leur pensée. Bien qu'elle soit habituelle et facile à mettre en œuvre sur papier, c'est une pratique pas ou peu supportée — du moins avec moins de souplesse que sur le papier — dans les environnements informatiques, suscitant de fait la frustration des lecteurs (Sellen *et al.*, 2003, p. 96) lorsqu'ils sont privés de cet outil précieux ;

– l'activité de classement et d'archivage de documents ⑥ nous permet de les stocker dans notre EPI, au sein de notre arborescence de fichiers par exemple. Kaye *et al.* (2006) ont identifié nos principales motivations : retrouver nos documents plus tard, construire un héritage (*legacy*) et les partager. Nous recourons au classement hiérarchique pour planifier nos projets, en les décomposant en sous-projets (Jones *et al.*, 2005; Jones, 2007). Cette stratégie est confirmée par Khoo *et al.* (2007), qui soulignent la fréquence des arborescences décomposées selon un à trois niveaux au moins. L'activité de classement d'un document dans l'EPI est une tâche hautement cognitive pour l'utilisateur. Rucker *et al.* (1997) indiquent que le regroupement d'objets dans un répertoire reflète une cohérence sémantique entre ces objets. Pour ce faire, l'individu doit tout d'abord se construire une représentation mentale du contenu d'un document, puis décider de le conserver ou pas en estimant son utilité future, par anticipation. Le cas échéant, il doit alors identifier le répertoire le plus adapté dans son EPI, éventuellement en créer un nouveau, pour enfin y stocker le document considéré. Cette activité de classement est fréquemment mise en œuvre : Abrams *et al.* (1998) rapportent notamment qu'un usager de signets web en stocke de trois à quatre par session de navigation.

Notre étude des activités documentaires ① à ⑥ met en lumière une pléthore de logiciels existants, qu'ils soient destinés à un seul individu (ex. : moteur de recherche) ou à un groupe d'individus (ex. : plate-forme de rédaction collaborative). Dans toutes ces approches, nous avons identifié deux problématiques interdépendantes. D'une part, les systèmes imposent une *linéarité* et un *cloisonnement* aux activités documentaires : un système de gestion de fichiers ne permet pas de rechercher de l'information, par

exemple. D'autre part, l'utilisateur est contraint à maîtriser plusieurs systèmes *distincts et spécialisés* : au moins un pour chacune des activités. Les activités représentées dans le cycle de vie du document de Sellen *et al.* (2003, p. 203) sont linéaires, alors que l'observation des pratiques documentaires reflète plutôt leur entrelacement. Par exemple, un individu peut rechercher de l'information ①, commencer la rédaction d'un document ② puis continuer à naviguer dans les résultats de sa recherche ① afin d'approfondir sa connaissance de la question. Comme les activités sont cloisonnées, chaque système réalise une seule activité. De par cette spécialisation, il ne peut construire qu'une représentation partielle des usagers réalisant l'activité pour laquelle le système est conçu, ignorant de fait les cinq autres activités du cycle de vie. Par conséquent, toute assistance apportée sur cette base partielle se révélerait être sous-efficace par essence. Pour autant, l'utilisateur et ses besoins restent les mêmes pour l'ensemble des six activités. Une autre contrainte issue de cette spécialisation implique que l'utilisateur doive maîtriser plusieurs systèmes distincts pour couvrir la totalité du cycle de vie du document, ce qui entraîne une charge cognitive importante. Cette surcharge est également accentuée par la nécessité de gérer et de maintenir plusieurs arborescences de documents (fichiers, emails, signets, etc.), impliquant une fragmentation des données dont les usagers se plaignent comme l'indique Jones (2007) dans sa synthèse des recherches sur la mémorisation et l'organisation des informations personnelles. Finalement, ces problématiques suggèrent le fait que les individus, au travers des systèmes qu'ils utilisent, ne tirent pas un bénéfice optimal des activités quotidiennes des autres membres organisationnels. Cela conduit à une faible exploitation et pérennisation de l'information introduite dans l'organisation, résultant d'efforts cognitifs mis en œuvre par chaque personne. Or, ce faible retour sur investissement implique une efficacité limitée de la gestion informationnelle sur le long terme.

Afin de répondre à ces problématiques, cet article présente une approche originale basée sur une architecture multi-utilisateur fédérant les activités du cycle de vie du document. En opposition avec la vision linéaire et cloisonnée des activités documentaires de Sellen *et al.* (2003, p. 203), nous montrons comment leur fédération rend possible l'intégration de processus exploitant les activités des usagers, sur les principes du donnant-donnant et de la non-intrusion. Le but recherché est d'apporter une assistance à chaque personne alors qu'elle réalise ses activités, le résultat d'un individu ainsi aidé étant capitalisé pour aider d'autres membres du groupe par la suite.

3. Gestion personnelle et collective de documents basée sur l'annotation

L'étude du cycle de vie du document a révélé deux problématiques principales : les activités documentaires sont linéaires et cloisonnées. Afin de proposer une alternative à la situation actuelle, nous détaillons dans cette section une architecture fournissant :

- *une assistance personnalisée*. L'architecture proposée repose sur un modèle unifié fédérant les usagers et les six activités documentaires, de façon à remédier aux représentations (profils) partielles et éparpillées dans plusieurs applications. Adossée à des processus présentés dans cette section, cette architecture aide l'utilisateur en situa-

tions de recherche ①, de rédaction ②③, de distribution ④, d'exploitation ⑤ et d'organisation ⑥ d'information. L'architecture proposée capitalise également les résultats produits par un individu réalisant une de ces activités, afin de les réinjecter dans les autres activités qui s'enrichissent mutuellement. Enfin, l'utilisateur est aidé dans son activité globale, ce qui profite notamment à la construction et à l'évolution quotidienne de son espace personnel d'annotations (EPA) ;

– *une assistance collective*. Nous faisons l'hypothèse que les membres organisationnels partagent des intérêts, des connaissances, des tâches et des activités car ils évoluent dans cette même organisation. Au cœur de nos travaux, nous considérons leurs activités afin de leur porter assistance. L'architecture multi-utilisateur proposée intègre donc des processus sur le principe du donnant-donnant, afin d'accroître le retour sur investissement relatif aux EPA des usagers notamment. Jusqu'alors manifestement considérés comme un capital documentaire en sommeil, nous les exploitons en tant que vecteurs d'une forte valeur ajoutée focalisée sur les activités de l'organisation. En retour, ils bénéficient du système en obtenant de l'information relative à leurs activités, cette information provenant alors du groupe dans son ensemble. De fait, toute information introduite dans l'organisation — extraite de sources externes comme produite par ses membres — devient profitable pour l'ensemble du groupe, alors qu'elle végète bien souvent dans les répertoires des membres organisationnels.

L'étude des activités documentaires en section 2 a identifié la pratique d'annotation en tant que partie prenante à trois d'entre elles : la rédaction ②, la finalisation ③ et l'exploitation ⑤ de documents. Ce constat a motivé notre approche visant à placer cette pratique d'annotation au cœur de notre architecture, de façon à couvrir les trois activités restantes. Les sections suivantes définissent le concept d'annotation collective, puis montrent comment ce concept fédère les activités de recherche d'information ①, de diffusion ④ et de classement de documents ⑥ au travers des six processus illustrés dans la figure 1 : ADAPTAFFICHAGE, NAVI, PROTODOC, RÉORG, RECO et VUE UNIFIÉE.

3.1. RÉORG — *Améliorer l'activité de classement et d'archivage* ⑥

3.1.1. *Définition des éléments fondamentaux de l'approche proposée*

Au quotidien, la copie d'un document dans son arborescence personnelle ou la création d'un signet vers celui-ci permettent de conserver (un lien vers) un document intéressant. Or, de nombreuses situations de travail nécessitent davantage la mémorisation des passages d'intérêt (ex. : quelques phrases, une définition, une photo, un horaire de train) avec les notes du lecteur en contexte, que le simple stockage du document dans son intégralité (Kidd, 1994). Ces approches sont alors inadéquates car elles ne permettent pas de désigner des parties de documents, ni de conserver les commentaires de son créateur. Afin d'apporter une réponse à ce problème, nous avons opté pour le concept d'annotation collective défini dans (Cabanac *et al.*, 2007b). Une telle annotation permet notamment de commenter tout ou partie d'un document. Nous discernons trois objectifs principaux pour les travailleurs du savoir : mémoriser et organiser de l'information, prendre des notes et débattre d'un passage avec d'autres usagers. Ces objectifs permettent de raffiner ce concept en trois catégories d'annotation :

– l’*annotation-stockage* étend la notion de signet en permettant la mémorisation de passages de documents, au lieu de leur intégralité. Le créateur d’une telle annotation l’organise dans son EPA, qui reprend la structure des signets : une arborescence de répertoires. Une raison motivant le choix d’une hiérarchie fait écho à l’étude de Jones *et al.* (2005) soulignant le fait que les individus ont besoin de classer leurs informations dans une hiérarchie pour réaliser leurs activités, notamment leurs projets ;

– l’*annotation-remarque* permet de commenter un passage de document sans pour autant obliger l’annotateur à la classer dans son EPA. L’utilité de cette catégorie d’annotation est visible lors de l’activité de finalisation ③ qui génère de nombreuses annotations (corrections de coquilles, indications de typographie, etc.) que l’annotateur n’a pas besoin de conserver dans son EPA. Par contre, il est essentiel que ses collègues et lui-même les voient lorsqu’ils visualisent le document annoté, par la suite ;

– l’*annotation-argumentative* est ancrée sur une partie d’un document et peut faire l’objet de réponses formulées par les lecteurs successifs. Par ailleurs, toute réponse peut également faire l’objet d’autres réponses encore, créant ainsi un fil de discussion similaire à ceux des forums du web ou de Usenet. Le fait de ne pas avoir à préciser le contexte de la discussion (qui est défini par le point d’ancrage dans le document) est un avantage indéniable de cette approche par rapport aux forums. Par ailleurs, les lecteurs peuvent accéder aux discussions relatives à un document donné sans avoir à connaître l’URL de la discussion, alors que cela est nécessaire actuellement avec une approche à base d’un forum indépendant.

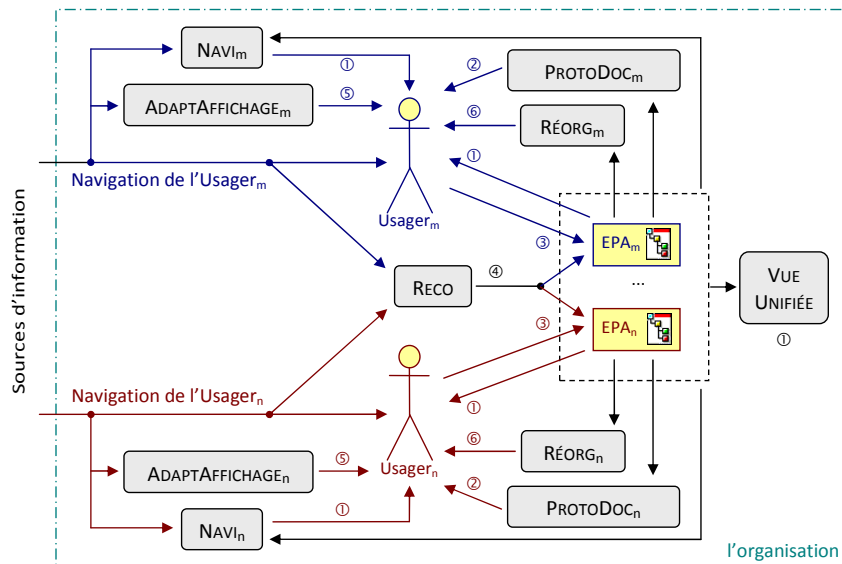


Figure 1. Schéma synoptique de l’architecture proposée représentant une organisation minimale composée de deux individus. Les flèches représentent les flux de données entre les usagers, leurs EPA et les six processus interdépendants

3.1.2. Modélisation des éléments fondamentaux de l'approche proposée

Le diagramme UML de la figure 2 formalise les catégories d'annotation et les EPA, concepts fondamentaux de notre proposition. Par souci de concision, nous ne présentons qu'un faible niveau de détail en cachant intentionnellement les compartiments des classes relatifs aux attributs et opérations. Nous commenterons ce diagramme à l'aide d'un scénario où un Usager visualise une Ressource telle qu'une page web, un planning sur l'intranet de son entreprise, une photo dans son arborescence personnelle. Pour conserver tout ou partie de cette ressource il crée une *AnnotationAncrée*, où la notion d'ancre fait référence au passage du document qu'il désire conserver. Deux situations peuvent alors survenir. D'une part, pour conserver l'intégralité de la ressource, le système emploie un *AncreGlobal* qui correspond à un signet classique en stockant l'URL du document, par exemple. Dans le cas contraire, l'utilisateur ne désire stocker qu'une partie de la ressource : deux phrases non contiguës, par exemple. Pour ce faire, une alternative consiste à modifier la ressource en y intégrant des marqueurs spécifiques référençant le début et la fin de la sélection de l'utilisateur. Cette solution n'est bien entendu possible qu'avec des documents modifiables, situation marginale dans le cas de la consultation de documents publics, notamment sur le web. C'est pourquoi nous avons opté pour une seconde alternative qui consiste à stocker un point d'ancre (*PointAncre*) localisant la sélection de l'utilisateur de façon univoque. Afin de prendre en compte divers formats de documents (HTML, RTF, PDF, etc.), la classe abstraite *PointAncre* doit être raffinée. Pour le format HTML par exemple, la syntaxe *XPointer* exprimant la sélection de l'utilisateur au sein de la structure logique du document (*Document Object Model*) peut être implantée dans une sous-classe de *PointAncre*.

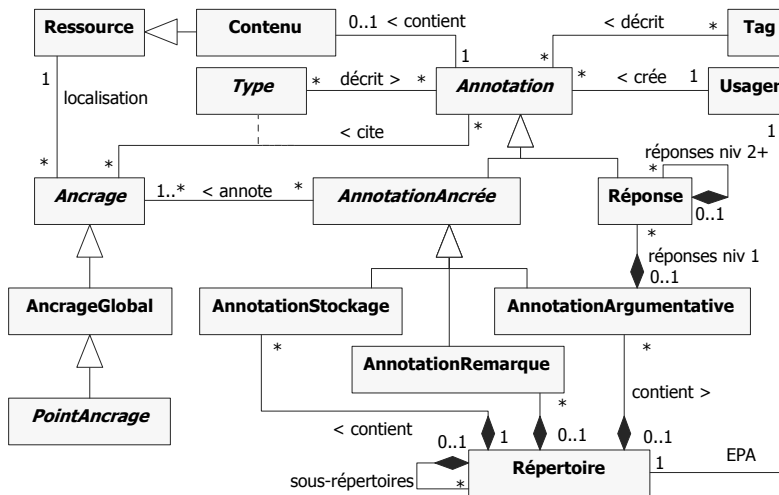


Figure 2. Diagramme de classes UML des concepts d'annotation collective et d'EPA

Le concept d'annotation collective s'articule autour des notions de *données objectives* et d'*informations subjectives* (Cabanac et al., 2007b). Les classes mentionnées

jusqu'à présent représentent les *données objectives* que le système d'annotation déduit du passage sélectionné dans le document. Par la suite, les classes restantes représentent les *informations subjectives* introduites par l'utilisateur qui crée l'annotation. Toute *Annotation* possède un *Contenu*, sans restriction quant au médium ou au format employé : du texte formaté ou un enregistrement audio, par exemple. En tant que sous-classe d'une *Ressource*, tout ou partie d'un *Contenu* peut être annoté à son tour, rendant ainsi possible la création d'annotations imbriquées utiles pour la rédaction ou l'interprétation collective de document. Outre un *Contenu* permettant d'exprimer un commentaire, le créateur de l'annotation peut indiquer des références à d'autres (parties de) ressources. De plus, il peut associer à ces références et à l'annotation elle-même des *Types* par l'intermédiaire des relations « cite » et « décrit ». Cette classe abstraite *Type* décrit l'intention de l'annotateur en fournissant un aperçu de sa sémantique : les sous-classes peuvent modéliser des taxonomies d'objectifs (ex. : commentaire, exemple, question), d'actions (ex. : à faire, à lire), d'opinions (ex. : réfutation, neutre, confirmation) ou de concepts spécifiques au domaine d'application (ex. : partenaire, produit, concurrence dans le domaine de la veille stratégique et technologique). Doter les membres organisationnels de telles taxonomies pourra les aider à décrire à l'aide d'un référentiel commun l'information qu'ils extraient, afin d'en améliorer la compréhension, par exemple. En complément de la classe *Type*, la classe *Tag* permet aux usagers de décrire une annotation à l'aide de leurs propres termes, dans la même optique que les systèmes de *social bookmarking* évoqués en section 2.

Durant l'activité ©, un *Usager* crée donc une *AnnotationAncrée* pour conserver de l'information dont il prend connaissance au cours de ses activités. Il peut ensuite organiser cette annotation selon ses besoins (par thème, par projet, chronologiquement, etc.) au sein de son EPA composé des *Répertoires* qu'il crée et gère. Par la suite, nous détaillons les processus mentionnés dans le schéma synoptique de l'architecture (figure 1) en relation avec les six activités du cycle de vie du document. Ces processus sont modélisés dans la figure 3, présentant un diagramme de classes qui complète le diagramme relatif à l'annotation collective (figure 2). Nous exposons pour chaque cas la raison d'être du processus, ce que l'utilisateur lui fournit en entrée (la plupart du temps de façon non explicite) et ce que le groupe et l'utilisateur lui-même en retirent.

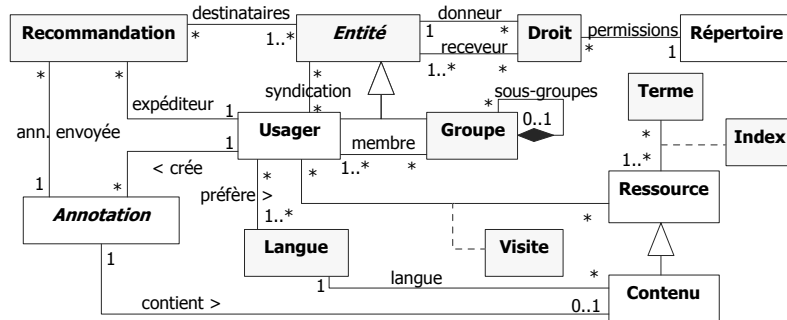



Figure 3. Diagramme de classes UML complétant la modélisation de la figure 2

Le processus RÉORG adjoint à l'activité © assiste l'utilisateur lors de la réorganisation de son EPA. Cette tâche est reconnue hautement cognitive et fréquemment reportée à une date ultérieure (ex : quand j'aurai plus de temps, une fois ce projet fini), comme l'indiquent Abrams *et al.* (1998) et plus récemment Jones (2007). Or, sur une longue période d'utilisation, réorganiser son EPA est indispensable afin de pouvoir accéder à l'information efficacement. Par exemple, certains répertoires « grossissent » quotidiennement lorsque l'utilisateur complète sa connaissance du domaine associé, en y insérant de plus en plus d'annotations pointant vers les passages d'intérêt. C'est dans un pareil cas que le raffinement (éventuellement thématique) du répertoire en sous-répertoires est requis, afin de structurer sa perception de son EPA tout en conservant la faculté d'accéder à ces informations en un temps minimal. C'est pourquoi le processus RÉORG se base sur l'EPA d'un utilisateur pour lui proposer une réorganisation thématique suite à sa demande explicite, qu'il peut accepter partiellement ou en totalité. Ce processus initialement proposé par Chevalier *et al.* (2004) met en œuvre une succession d'algorithmes : l'indexation du contenu des Ressources annotées (Baeza-Yates *et al.*, 1999, ch. 2), une classification ascendante hiérarchique (Jardine *et al.*, 1971) fournissant une arborescence thématique binaire de l'EPA, un seuillage de cette dernière comme proposé par Maarek *et al.* (1996) afin de limiter la profondeur de l'arborescence obtenue, ainsi qu'un étiquetage des répertoires obtenus à l'aide du coefficient du χ^2 pour déterminer les termes les plus discriminants de chaque répertoire.

3.2. ADAPTAFFICHAGE — Améliorer l'activité d'exploitation de documents ©

La section précédente a décrit le rôle dual de la pratique d'annotation. Pour un usage *individuel*, c'est un outil mis en œuvre lors de la réflexion critique, en permettant la prise de notes en contexte (AnnotationRemarque) contrairement au recours à des services tiers (forum, par exemple). Ces notes peuvent être conservées et organisées au sein de l'EPA de l'utilisateur (AnnotationStockage). Par ailleurs, pour un usage *collectif*, les annotations peuvent être partagées à des fins de travail collaboratif : les lecteurs prennent connaissance des commentaires et opinions de leurs prédécesseurs, ils peuvent également participer à des débats en contexte, ancrés dans le document annoté (AnnotationArgumentative). Le fait de créer des annotations est certes utile, toutefois l'affichage de ces dernières peut poser des problèmes liés au passage à l'échelle du dispositif de visualisation. En effet, l'affichage des annotations au sein du document, replacées sur leur point d'ancrage, peut déranger le lecteur dans certains cas. Des observations empiriques montrent que la difficulté de lire un document croît avec le nombre d'annotations affichées. L'utilisation du système Annotea/Amaya (Kahan *et al.*, 2002) confirme ce problème de visualisation, tant les annotations représentées à l'aide de l'icône «  » perturbent la visualisation du contenu initial des documents¹.

Nous proposons deux approches complémentaires pour réduire les efforts cognitifs des utilisateurs. En tant que première adaptation aux besoins de l'utilisateur, la visualisa-

1. cf. <http://www.irit.fr/~Guillaume.Cabanac/annotation/demoAmaya.wmv>

tion proposée ne restitue pas les Contenus qui ne sont pas exprimés dans une Langue spécifiée par l’usager. Bien que très basique, cette stratégie qui évite d’afficher des éléments incompréhensibles à l’usager ne figure pas dans les systèmes d’annotation de notre connaissance. Une seconde adaptation concerne les AnnotationsArgumentatives et les débats qu’elles suscitent. Lorsque de nombreux débats sont ancrés sur un document donné, le lecteur est tenté de tous les consulter, tour à tour. Pour chaque débat, il peut alors déduire mentalement l’opinion globale de ses participants pour évaluer la « validité sociale » de l’AnnotationArgumentative. Bien que nécessaire à la lecture critique et approfondie d’un document, consulter chaque débat et synthétiser leurs opinions demandent un temps non négligeable, ainsi qu’une charge cognitive élevée. En effet, il faut tout d’abord identifier les opinions exprimées dans chaque argument, puis les synthétiser de façon récursive, des feuilles vers la racine de l’arbre représentant le débat. Afin de soulager l’usager de ce fardeau, nous intégrons à l’architecture le processus ADAPTAFFICHAGE basé sur les algorithmes de validation sociale définis dans (Cabanac *et al.*, 2007b). Leur fonctionnement est calqué sur celui des lecteurs, en synthétisant les Réponses de façon à obtenir la validité sociale d’une AnnotationArgumentative : une valeur définie sur une échelle entre *réfuté* et *confirmé* pour représenter l’opinion globale du groupe qui débat eu égard à l’annotation racine. Grâce à ce processus, la seconde adaptation consiste donc à informer les usagers du degré de consensus (resp. controverse) atteint dans chaque débat. Ainsi, ils peuvent se focaliser sur des informations stabilisées, ou bien sur des discussions pour lesquelles les divers participants n’ont pas encore trouvé de consensus.

3.3. PROTODOC — Améliorer les activités de création ② et de finalisation ③

Une grande partie de la connaissance d’une organisation est contenue dans les documents qu’elle produit, étant donné que les travailleurs du savoir consacrent beaucoup de temps à la rédaction de documents. Pour ce faire, ils extraient des informations à partir de divers documents avant de les synthétiser en un nouveau document, tel qu’un rapport. Eu égard à l’architecture proposée, l’*AnnotationAncrée* fournit au lecteur un moyen efficace pour collecter des pépites informationnelles et y associer un Contenu pour garder une trace de son interprétation, par exemple. Pour un projet spécifique, tel que l’analyse quotidienne des cours de la Bourse par exemple, l’usager peut créer un Répertoire idoine dans son EPA afin de regrouper l’ensemble des bribes de documents qu’il désire conserver à cet effet. Par ailleurs, la notion de Droit associée aux Répertoires restreint l’accès aux répertoires partagés entre différentes *Entités* collaborant à des tâches telles que la recherche et d’analyse collectives.

Afin d’assister l’activité de création de document, le processus PROTODOC ébauche un proto-document à partir des Répertoires de l’EPA désignés par l’usager. Ce résultat permet notamment de constituer le squelette d’un dossier d’analyse. Pour chaque annotation des répertoires concernés, le processus intègre au proto-document les *Anchages* des Ressources annotées, leur validité sociale, ainsi que les diverses informations fournies par l’annotateur : Contenu, Tags, citations, etc. Par la suite, l’usager peut compléter et retravailler cette ébauche à l’aide du traitement de texte de son choix.

3.4. NAVI & VUE UNIFIÉE — *Améliorer l'activité de recherche d'information* ①

L'architecture proposée améliore les deux modalités au cœur de la recherche d'information : l'interrogation et la navigation. Premièrement, nous proposons de compléter la modalité d'interrogation par la prise en compte des annotations collectives dans (Cabanac *et al.*, 2007b). L'idée est d'exploiter les annotations des lecteurs en tant que contributions reflétant un *feedback* social, dans le but d'améliorer le rappel (en retrouvant davantage de documents pertinents par rapport à la requête) ainsi que la précision des résultats de recherche (en ne retrouvant que les documents pertinents). Au sujet du rappel, le « problème du vocabulaire » formulé par Furnas *et al.* (1987) indique que la requête d'un usager contient rarement (< 20 %) les mêmes termes que ceux constituant les documents pertinents. Or, Fraenkel *et al.* (1999) rapportent de nombreux exemples où le contenu des annotations permet de trouver davantage de documents pertinents, car les termes employés par les lecteurs dans les annotations sont parfois complémentaires de ceux présents dans le passage rédigé par l'auteur du document. De ce fait, la prise en compte des annotations permet d'améliorer le rappel, en retrouvant des documents de façon indirecte, ainsi que la recherche contextuelle en retrouvant des passages annotés plutôt que des documents entiers. Par ailleurs, concernant la précision des recherches, la considération des Contenus des annotateurs permet la désambiguïsation des documents annotés, ainsi que l'intégration de leurs termes dans le processus d'indexation. De plus, la validation sociale des débats (AnnotationArgumentative) offre un moyen de caractériser une Ressource en tant que fiable, controversée, populaire, abandonnée, etc. La prise en compte de tels indicateurs permet alors d'adapter le moteur de recherche aux préférences des usagers.

Deuxièmement, le processus NAVI assiste l'utilisateur de façon non intrusive en exploitant les EPA de l'organisation (résultant des efforts cognitifs de classement de chacun) afin d'en extraire des documents pertinents eu égard à la navigation courante de l'utilisateur. Ces documents identifiés dans le capital documentaire collectif (les EPA) sont alors recommandés à l'utilisateur de façon synchrone à sa navigation. La mise en œuvre de ce processus est originale à deux égards :

1) le processus NAVI exploite la capacité des membres organisationnels à trouver, filtrer et classer des documents pertinents pour l'organisation, étant donné que les recommandations calculées proviennent des EPA de l'organisation. De ce fait, des documents recherchés il y a longtemps puis oubliés dans les répertoires des individus sont automatiquement rentabilisés en les proposant à leurs collègues. Ici nous posons deux hypothèses de travail : *i*) un document contenu dans un répertoire a été jugé intéressant par le propriétaire de l'EPA car il a mis en œuvre un effort cognitif en sélectionnant le répertoire le plus approprié dans son EPA (Rucker *et al.*, 1997). De plus, *ii*) d'autres membres du groupe ayant des intérêts communs peuvent être intéressés par un tel document pour réaliser leurs propres activités ;

2) le second aspect original concerne l'algorithme d'identification des documents à recommander en fonction du document qui est visualisé par l'utilisateur. Les approches classiques se basent uniquement sur le contenu des documents, après une phase d'indexation. Prenant le contre-pied de ces approches, nous avons défini une mesure de

similarité ne nécessitant pas de les indexer : elle est basée sur l'usage des documents (Chevalier, 2002; Cabanac *et al.*, 2007a). Cette dernière évalue à quel point les individus utilisent les documents ensemble pour réaliser leurs activités. Nous postulons que les personnes regroupent dans des répertoires proches — au sein de leur EPA — les documents qu'ils jugent similaires, pour une quelconque raison : un projet commun, leur thématique, leur auteur, etc. La valeur de similarité sur l'usage pour deux documents donnés est alors d'autant plus importante qu'ils sont proches dans les EPA, et que de nombreux individus ont opéré à un tel regroupement. Concrètement, cette mesure repose sur des algorithmes de parcours d'arbre (EPA) détaillés dans (Cabanac *et al.*, 2007a). Il est à noter que les documents sont recommandés à l'utilisateur accompagnés de leurs chemins dans les EPA, de façon à ce que l'individu puisse appréhender le contexte d'organisation et l'intention de classement du propriétaire du document recommandé. L'indication de sa provenance évite que l'utilisateur n'ait l'impression d'accéder au document par « téléportation », ce qui est souvent reproché aux moteurs de recherche selon Teevan *et al.* (2007).

Le processus NAVI recommande des documents introduits dans l'organisation de façon synchrone (durant la navigation) et en fonction d'un besoin précis (selon le document visualisé). Afin d'améliorer l'accès au capital organisationnel en dehors de la tâche de navigation, le processus VUE UNIFIÉE en offre une vision globale, permettant ainsi aux membres organisationnels de prendre connaissance de leur environnement de travail et des ressources capitalisées par le groupe. Concrètement, l'interface multifacette proposée répond à deux types de besoins : opérationnels et stratégiques. Les besoins *opérationnels* émanent des travailleurs du savoir qui cherchent souvent des documents pour trouver des personnes, et *vice versa* selon Hertzum *et al.* (2000). Les besoins opérationnels correspondent également à l'identification des documents utilisés ensemble, à l'identification des thématiques de certaines personnes, à l'accès à leurs EPA... Les besoins *stratégiques* émanent du pilotage de l'organisation, ils visent à identifier des personnes ressources, à établir des fiches de postes pour anticiper le *turnover*, à constituer des groupes de travail pour réaliser des projets demandant des compétences spécifiques... Nous invitons le lecteur à consulter (Cabanac, 2008a; Cabanac, 2008b) pour davantage de précisions, il pourra également manipuler l'interface multifacette à partir de la page web dédiée².

3.5. RECO — Améliorer la distribution des documents ④

L'étude de Feldman (2004) montre à quel point une diffusion d'information limitée au sein de l'organisation mène à des situations contreproductives et coûteuses : méconnaissance des compétences au sein de l'organisation, perte de temps en recherches infructueuses, recréation d'information, etc. C'est pourquoi nous proposons de tirer parti des informations recueillies par les individus, qui demeurent manifestement en sommeil dans le cas contraire. Pour ce faire, nous offrons aux Usagers la possibilité d'émettre des Recommandations *manuelles* à destination d'autres *Entités* de sa

2. cf. <http://www.irit.fr/~Guillaume.Cabanac/MultiFacettes>

connaissance (figure 3). Une fonctionnalité duale permet à chaque Usager de s'enregistrer pour recevoir des notifications concernant les événements sur les *Entités* de son choix. De ce fait, on peut notamment spécifier les répertoires des EPA pour lesquels on désire connaître tout ajout d'annotation. Cette demande proactive de notification est similaire à la syndication de flux RSS sur une page web, cf. (Hammond *et al.*, 2005).

Comme le souligne la section 2, la diffusion manuelle est limitée par divers facteurs liés à l'usager : son réseau social, sa volonté de partager alors que l'information est fréquemment assimilée au pouvoir, les efforts cognitifs en jeu, etc. Pour dépasser ces limites, nous proposons de compléter la diffusion manuelle par une diffusion *automatique* grâce au processus RECO représenté dans la figure 1. Ce dernier vise à exploiter la capacité des individus à rechercher de l'information, de façon non intrusive. En effet, il considère tout document introduit dans l'organisation par un usager comme candidat à la recommandation car d'autres membres possédant des centres d'intérêt similaires peuvent en bénéficier. De façon très synthétique, le processus RECO décrit en détail dans (Chevalier *et al.*, 2004) opère de la façon suivante. Chaque document introduit dans l'organisation (objet de la navigation d'un usager, par exemple) est indexé (classes Terme et Indexation). Puis, le système calcule sa similarité thématique avec les répertoires des EPA des autres Usagers. Pour ce faire, chaque répertoire est doté d'un classifieur construit par extraction des caractéristiques thématiques des contenus du répertoire. Enfin, le document indexé est recommandé dans le répertoire le plus spécifique de l'EPA considéré, à condition que la similarité calculée excède un seuil calculé dynamiquement. Ce seuil permet de limiter le nombre des recommandations de façon à ne pas surcharger l'usager.

4. Expérimentations et développements en cours

Afin de valider l'approche présentée dans cet article, nous adoptons une démarche incrémentale. Cette dernière consiste à tout d'abord évaluer chaque processus de façon indépendante, puis à évaluer l'approche dans son ensemble. Dans cette optique, nous avons pour l'instant évalué trois processus parmi les six proposés : NAVI, RECO et ADAPTAFFICHAGE. L'évaluation des processus NAVI et RECO est exposée dans (Chevalier, 2002; Chevalier *et al.*, 2004). De ce fait, nous ne rapportons que des éléments synthétiques dans cette section. Afin d'évaluer la qualité des recommandations fournies par NAVI durant la navigation, une expérimentation a été conduite à partir des EPA de 14 enseignants-chercheurs, totalisant 4 079 documents (resp. 486 répertoires) avec une moyenne de 291 documents (resp. 34 répertoires) par usager. Cinq individus sélectionnés ont réalisé une navigation prédéfinie par les expérimentateurs. Les résultats de cette expérimentation montrent que la pertinence des recommandations (par rapport au besoin de l'usager) augmente au cours de la navigation. Concernant le processus RECO, différentes stratégies de recommandation d'un document dans les EPA ont été évaluées à partir de la collection TREC 2001 OHSUMED/MeSH.

Par ailleurs, nous avons expérimenté les algorithmes de validation sociale sur lesquels repose notamment le processus ADAPTAFFICHAGE. Sur une période de 16 mois,

d'avril 2007 à août 2008, l'expérimentation visait à évaluer à quel point les résultats de nos algorithmes (Cabanac *et al.*, 2007b) sont proches de la perception humaine du consensus dans des débats argumentatifs. Pour ce faire, nous avons établi un protocole d'expérimentation conforme aux bonnes pratiques observées en psychologie expérimentale (Reips, 2007). Concrètement, un logiciel déployé avec Java WebStart permet à des bénévoles de prendre part à cette expérimentation en ligne³. Contrairement à une expérimentation réalisée sous le contrôle des expérimentateurs dans un laboratoire, notre approche est « écologique » dans le sens où les sujets participent dans leur environnement habituel. Un autre avantage réside dans le fait que les participants sont volontaires, car sollicités à partir de listes de diffusion telles qu'ACM SIGCHI. L'analyse des 121 participations détaillée dans (Cabanac, 2008a) a révélé que les algorithmes proposés correspondent à la perception humaine du consensus dans 80 % des cas.

Outre les expérimentations, nous tenons à mettre en œuvre nos propositions par le développement d'un prototype nommé TafAnnote. Ce prototype représente une « preuve de concept » dans le sens où il démontre la faisabilité de notre approche : fédération et amélioration des activités documentaires par l'annotation collective. Concrètement, il repose sur une architecture client-serveur. Le module client prend la forme d'une barre d'outils intégrée au navigateur Mozilla Firefox, donnant accès aux différentes fonctionnalités : création d'annotation, support des fils de discussion, gestion des EPA, recherche d'annotation, recommandation durant la navigation... Ce composant communique au travers d'une connexion réseau avec le module serveur qui stocke les annotations. Lorsqu'un usager demande l'affichage d'une page web, l'événement correspondant est intercepté pour transmettre immédiatement l'URL demandée au serveur d'annotations. En parallèle, le navigateur récupère le contenu du document, l'affiche et l'indexe en tâche de fond : étapes de segmentation, lemmatisation et dénombrement des lemmes. Enfin, les couples (terme, nombre d'occurrences) résultants sont envoyés au serveur d'annotations. Cette stratégie d'indexation répartie sur les machines clientes disposant du contenu des documents à moindre coût (ils sont déjà récupérés pour être affichés) permet de ne pas saturer le serveur d'annotations qui devrait récupérer et indexer tous les documents visualisés par tous les usagers sinon. Suite à l'envoi de l'URL demandée par l'utilisateur, le serveur d'annotations retourne les annotations correspondantes au module client, qui les intègre dans le document HTML affiché en modifiant le *document object model* (DOM) associé. Ainsi, les annotations sont affichées en contexte, en fonction de leur point d'ancrage respectif. La figure 4 montre un exemple de document contenant deux annotations. Les Types (cf. figure 2) des annotations sont représentés par des pictogrammes visuels qui sont mis en exergue en fonction de la validité sociale de l'annotation associée. De ce fait, le lecteur peut visuellement identifier les annotations consensuelles ou controversées, avant même d'en consulter les fils de discussions associés. Nous invitons le lecteur à consulter (Cabanac, 2008a) qui détaille davantage les aspects relatifs au prototype TafAnnote, disponible en téléchargement sur le site dédié⁴.

3. cf. <http://www.irit.fr/~Guillaume.Cabanac/expe>

4. cf. <http://www.irit.fr/~Guillaume.Cabanac/TafAnnote>

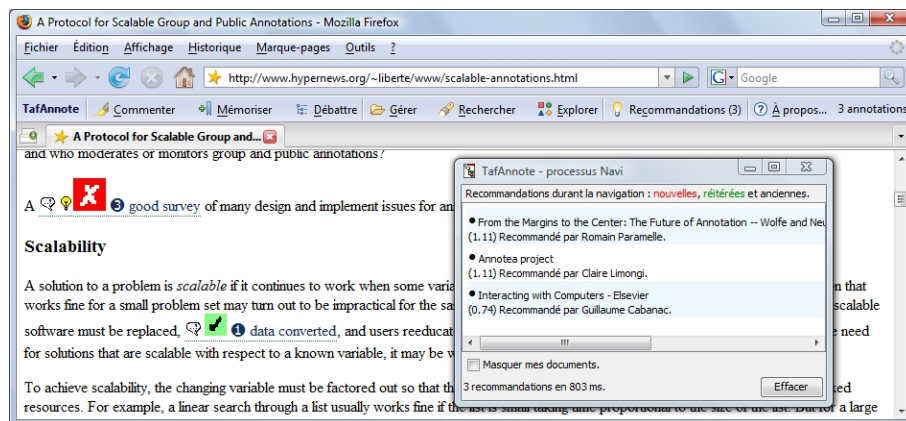


Figure 4. Capture d'écran de TafAnnote intégré dans Firefox. Le document affiché contient trois annotations dont les types et la validité sociale sont affichés. De plus, l'utilisateur a affiché les recommandations de navigation en cliquant sur la barre d'outils

5. Discussion

Notre proposition repose sur la valorisation des EPA des membres organisationnels. En particulier, des documents annotés et organisés par leurs soins sont automatiquement proposés à leurs collègues. La dissémination de ces documents professionnels stockés dans des espaces personnels ne contrevient pas à la législation française, comme l'atteste une recommandation de la CNIL (Bouchet, 2004, p. 25). Ainsi, tout document professionnel issu des EPA peut légalement être exploité par notre approche, exceptés ceux contenant une mention personnelle dans le titre — l'organisation pouvant suggérer l'utilisation de mots-clés spécifiques tel que « perso ». Toutefois, les travaux concernant la gestion d'information collective passés en revue par Lutters *et al.* (2007) montrent que les individus ont besoin de conserver des informations professionnelles privées (ex : choix stratégiques). Ces auteurs rappellent qu'il est essentiel de pouvoir garder le contrôle sur ce qui est échangé, le partage d'information étant flexible, nuancé et contextualisé. Afin de satisfaire ces contraintes et de limiter la résistance des usagers, nous avons proposé dans un premier temps un mécanisme de droits d'accès classique (figure 3). Celui-ci n'est pas complètement adapté aux attentes des usagers qui préféreraient une approche adaptative, ne requérant qu'un minimum d'implication (Lutters *et al.*, 2007). Une piste d'amélioration pourrait s'appuyer sur les actions des usagers pour estimer des seuils de confiance en fonction des échanges observés : l'envoi d'une recommandation de lecture peut indiquer la confiance que l'expéditeur confère au destinataire, concernant ce document et éventuellement ceux qui sont utilisés avec (cf. mesure de similarité sur l'usage, section 3.4). De tels indicateurs pourraient alors être utilisés pour adapter dynamiquement les politiques d'accès aux documents.

En relation avec la notion de droits d'accès, Lutters *et al.* (2007) rapportent à quel point le choix du droit d'accès par défaut (public ou privé) peut affecter le modèle de collaboration sous-jacent. En effet, « privé par défaut » peut conduire les usagers à faire de la rétention non intentionnelle d'information, nuisant alors à la performance des processus d'assistance proposés. Par ailleurs, « public par défaut » peut exacerber la résistance des usagers à adopter le système. Enfin, il nous semble capital d'être en mesure d'identifier et de modérer les passagers clandestins (*free riders*) : les usagers qui parasitent le système en consommant de l'information sans jamais en offrir.

Le dernier point de cette discussion concerne l'approche proposée : la fédération des activités documentaires, qui induit de nombreux avantages mais également des inconvénients critiques au regard de l'utilisateur :

– *avantages de la fédération.* Elle offre une vision globale de l'ensemble des activités en les décloisonnant. La fédération permet d'utiliser les résultats d'une activité menée par un usager (ex : le stockage d'une annotation) pour enrichir ses activités ainsi que celles d'autres membres organisationnels (ex : en leur recommandant le document annoté dans leur EPA *via* le processus RECO). C'est aussi grâce à la fédération que les six processus donnant-donnant sont réalisables. De plus, elle a permis la définition d'un modèle de données unifié, formalisant l'annotation collective et le concept d'EPA, où les usagers stockent et organisent leurs annotations selon leurs besoins. L'interface multifacette exploite également cette unification des données en indexant au préalable les documents des EPA de façon incrémentale (section 3.4). En résumé, c'est au travers de la fédération que notre approche améliore les pratiques documentaires : annotation collective, recommandations synchrone et asynchrone, réorganisation des EPA, exploration du capital humain et documentaire dans une interface multifacette...

– *inconvénients de la fédération.* Les avantages de la fédération sont acquis au détriment de la liberté de l'utilisateur. Jusqu'alors, il devait péniblement jongler avec une kyrielle d'applications, en contrepartie il avait toujours le choix d'en utiliser une plutôt qu'une autre. Dans notre approche, c'est un système unique que l'utilisateur manipule. Étant par conséquent limité dans sa liberté de choix, il faut s'attendre à ce qu'il rejette en bloc l'application proposée, même si le fait qu'elle améliore ses activités soit avéré.

La fédération représente le *moyen* retenu pour arriver à la *fin* consistant à améliorer les activités documentaires. Ce choix résulte de l'étude comparative des possibilités de communication entre les applications couvrant le cycle de vie du document. À l'heure actuelle, à notre connaissance, il n'existe aucune opportunité répandue qui constituerait une solution réalisable. Chaque application gère les données des usagers dans son format (propriétaire), certaines en donnent accès aux travers d'API qui sont parfois trop limitées pour implanter les processus présentés dans cet article. Enfin, au niveau technique, les applications sont loin de partager un protocole universel et indépendant du langage d'implantation : RPC, CORBA, RMI-IIOP, DCOM, XPConnect, SOAP, etc. représentent uniquement la partie visible de l'iceberg. En faisant abstraction de ces limites techniques, la fédération dans un seul système pourrait être avantageusement (pour l'utilisateur) remplacée par une approche fondée sur l'interopérabilité

entre applications, comme le proposent Chevalier *et al.* (2008) au sujet de l'exploitation des profils d'utilisateurs. Avec la maturité du web 2.0 et les avancées quotidiennes du web sémantique, gageons que les applications intégreront à terme de véritables couches d'interopérabilité. Dès lors, nous pourrions mettre en œuvre les processus présentés dans cet article sans que l'utilisateur n'ait à s'adapter à un système fédéré comme nous sommes obligés de l'imposer à cause des limites actuelles.

6. Conclusion et perspectives de travaux

Dans le contexte d'une organisation, nous avons présenté les activités documentaires réalisées au quotidien par les « travailleurs du savoir ». Ces activités formant le cycle de vie du document (Sellen *et al.*, 2003, p. 203) ont été illustrées par des systèmes individuels comme collectifs. À la lumière de cette étude, nous avons identifié plusieurs problématiques. D'une part, la grande diversité de systèmes que les utilisateurs doivent maîtriser pour réaliser les six activités documentaires implique un éparpillement de leurs données, ainsi qu'une surcharge cognitive de leur part. D'autre part, les activités documentaires sont cloisonnées et linéaires, alors que les individus se comportent tout autrement. De ce fait, chaque système est très adapté pour une seule activité, mais il ignore les autres. Cette situation conduit à une représentation partielle des utilisateurs (un traitement de texte « connaît » ses utilisateurs en tant que rédacteurs, mais jamais en tant que chercheurs d'information) et à une assistance forcément sous-optimale. Enfin, les informations introduites dans l'organisation — au prix de coûteux efforts d'élicitation des besoins en information, de recherche, de filtrage, d'analyse, de consolidation, de structuration et de maintenance d'information dans les Espaces Personnels d'Annotations (EPA) — ne sont jamais valorisées, formant ainsi un capital à haute valeur ajoutée, mais paradoxalement en sommeil.

Afin de proposer une solution originale à ces problématiques, nous avons présenté dans cet article une approche basée sur la fédération des activités documentaires dans une architecture multi-utilisateur. Elle vise à améliorer les activités quotidiennes de chaque utilisateur, en lui faisant bénéficier du capital de l'organisation et *vice versa*, sur le principe du donnant-donnant. La conception du système favorise également un enrichissement mutuel des activités : la tâche de navigation, incluse dans l'activité de recherche d'information, exploite le résultat de l'activité de classement, par exemple. Considérer l'organisation comme un capital à haute valeur ajoutée pour que chacun de ses membres en tire bénéfice est un aspect au cœur de notre proposition. Ainsi, nous exploitons l'ensemble des EPA de façon non intrusive, afin d'aider chaque individu pour améliorer le retour sur investissement global.

L'architecture que nous proposons est basée sur un concept fédérateur de l'ensemble des activités documentaires : l'annotation collective, en tant que trace de l'activité intellectuelle des travailleurs du savoir. Cette démarche fait suite aux observations de Kidd (1994) confirmées plus récemment par Sellen *et al.* (2003, p. 63) : ce sont les annotations des individus qui contiennent la réelle valeur ajoutée du document. Ce modèle est complété par les six processus représentés dans la figure 1. Ils

visent à aider l'utilisateur dans les tâches de réorganisation de son EPA (RÉORG), d'exploitation de documents (ADAPTAFFICHAGE), de création de proto-documents à partir des annotations formulées (PROTODOC), de navigation ou de recherche en recommandant des documents provenant de sa propre organisation (NAVI et RECO). Par ailleurs, la VUE UNIFIÉE proposée permet la visualisation et l'exploration de l'organisation selon deux dimensions (les individus et les documents) afin d'en obtenir une meilleure connaissance. Enfin, nous portons une grande importance au développement du prototype TafAnnote afin de démontrer la faisabilité de nos propositions. Ce dernier prend en charge actuellement une majorité des processus présentés dans cet article ; il est en constant développement pour intégrer l'ensemble de nos propositions à court terme. Une partie des propositions de cet article a été évaluée, notamment au travers de la validation sociale qui a fait l'objet d'une expérimentation ayant recueilli les participations de 121 volontaires en ligne et dont les résultats montrent que les algorithmes de validation sociale approximent la perception humaine du consensus à hauteur de 80 %.

De nombreuses perspectives sont envisageables à ces travaux, en particulier concernant l'évaluation de l'architecture dans son ensemble. Pour ce faire, nous envisageons le recours à des méthodologies issues de travaux en sciences sociales, afin d'évaluer le gain en efficacité des membres organisationnels en situation de travail quotidien. Le cas échéant, nous pensons mesurer le compromis réalisé entre les contraintes imposées par notre application et les gains qui en découlent, sur le modèle de l'analyse de Millen *et al.* (2003). Des perspectives moins liées au domaine de l'interaction homme-machine, mais relatives à l'annotation électronique sont également envisagées. Ainsi, nous pouvons considérer la spécification de techniques d'ancrages afin de pouvoir prendre en compte un vaste panel de formats de documents. Concernant l'ancrage des annotations, la recherche de techniques dites « robustes » est également nécessaire pour restituer correctement des annotations sur des ressources évolutives (cas des versions de documents). Enfin, un autre axe de travaux a trait au passage à l'échelle de l'architecture client-serveur proposée, bien que le contexte applicatif de l'organisation s'y prête moins qu'un contexte plus ouvert, de type web. La question des interfaces de restitution des annotations représente également une piste de travail à creuser afin d'améliorer l'utilisabilité des systèmes d'annotation.

7. Bibliographie

- Abrams D., Baecker R., Chignell M., "Information Archiving with Bookmarks: Personal Web Space Construction and Organization", *CHI'98: Proceedings of the conference on Human factors in computing systems*, ACM Press, New York, NY, USA, p. 41–48, 1998.
- Adler M. J., van Doren C., *How to read a book*, Simon & Shuster, NY, 1972.
- Baeza-Yates R. A., Ribeiro-Neto B. A., *Modern information retrieval*, ACM Press/Addison-Wesley, 1999.
- Ballay J.-F., "Nous sommes tous des travailleurs du savoir", *L'Expansion Management Review*, vol. 107, p. 94–101, December, 2002.
- Bouchet H., La cybersurveillance sur les lieux de travail, Technical report, CNIL, Paris, France, 2004. electronic edition <http://www.ladocumentationfrancaise.fr/rapports-publics/044000175/>.

- Cabanac G., Fédération et amélioration des activités documentaires par la pratique d'annotation collective, Thèse de doctorat, Université de Toulouse, France, December, 2008a.
- Cabanac G., "Interface multi-facettes d'accès au capital documentaire de l'organisation", *INFORSID'08 : 26^e congrès de l'Informatique des Organisations et Systèmes d'Information et de Décision*, Éditions Inforsid, p. 69–84, June, 2008b.
- Cabanac G., Chevalier M., Chrisment C., Julien C., "An Original Usage-based Metrics for Building a Unified View of Corporate Documents", in R. Wagner, N. Revell, G. Pernul (eds), *DEXA'07: Proceedings of the 18th International Conference on Database and Expert Systems Applications*, vol. 4653 of *LNCS*, Springer, p. 202–212, September, 2007a.
- Cabanac G., Chevalier M., Chrisment C., Julien C., "Collective Annotation: Perspectives for Information Retrieval Improvement", *RIAO'07: Proceedings of the 8th conference on Information Retrieval and its Applications*, CID, May, 2007b.
- Cabanac G., Chevalier M., Chrisment C., Julien C., "Exploiting the Annotation Practice for Personal and Collective Information Management", in Lopistéguy *et al.* (2008), p. 55–66, May, 2008a.
- Cabanac G., Chevalier M., Chrisment C., Julien C., Soulé-Dupuy C., Tchienehom P., "Web Information Retrieval: Towards Social Information Search Assistants", in T. Kidd, I. Chen (eds), *Social Information Technology: Connecting Society and Cultural Issues*, IGI Global, chapter 16, p. 218–252, March, 2008b.
- Chevalier M., Interface adaptative pour l'aide à la recherche d'information sur le Web, Thèse de doctorat, Université Toulouse 3, France, December, 2002.
- Chevalier M., Chrisment C., Julien C., "Helping People Searching the Web: Towards an Adaptive and a Social System", *ICWI'04: Proceedings of the 3rd International Conference WWW/Internet*, IADIS, p. 405–412, 2004.
- Chevalier M., Julien C., Soulé-Dupuy C., "Profils usagers pour la recherche d'information — Pertinence de leur usage ?", in Lopistéguy *et al.* (2008), p. 81–93, May, 2008.
- Feldman S., "The high cost of not finding information", *KM World magazine*, vol. 13, n° 3, p. electronic edition <http://www.kmworld.com/Articles/PrintArticle.aspx?ArticleID=9534>, March, 2004.
- Fraenkel A. S., Klein S. T., "Information Retrieval from Annotated Texts", *J. Am. Soc. Inf. Sci.*, vol. 50, n° 10, p. 845–854, 1999.
- Furnas G. W., Landauer T. K., Gomez L. M., Dumais S. T., "The Vocabulary Problem in Human-System Communication", *Commun. ACM*, vol. 30, n° 11, p. 964–971, 1987.
- Hammond T., Hannay T., Lund B., Scott J., "Social Bookmarking Tools (I): A General Review", *D-Lib Magazine*, vol. 11, n° 4, p. electronic edition <http://dx.doi.org/10.1045/april2005-hammond>, April, 2005.
- Hertzum M., Pejtersen A. M., "The information-seeking practices of engineers: searching for documents as well as for people", *Inf. Process. Manage.*, vol. 36, n° 5, p. 761–778, 2000.
- Jardine N., van Rijsbergen C. J., "The Use of Hierarchic Clustering in Information Retrieval", *Information Storage and Retrieval*, vol. 7, n° 5, p. 217–240, 1971.
- Jones W., "How People Keep and Organize Personal Information", in Jones *et al.* (2007), chapter 3, p. 35–56, 2007.

- Jones W., Phuwanartnurak A. J., Gill R., Bruce H., “Don’t Take My Folders Away!: Organizing Personal Information to Get Things Done”, *CHI’05 extended abstracts on Human factors in computing systems*, ACM Press, New York, NY, USA, p. 1505–1508, 2005.
- Jones W., Teevan J., *Personal information management*, University of Washington Press, WA, USA, 2007.
- Kahan J., Koivunen M.-R., Prud’Hommeaux E., Swick R. R., “Annotea: an open RDF infrastructure for shared Web annotations”, *Comp. Netw.*, vol. 32, n° 5, p. 589–608, August, 2002.
- Kaye J. J., Vertesi J., Avery S., Dafoe A., David S., Onaga L., Rosero I., Pinch T., “To Have and to Hold: Exploring the Personal Archive”, *CHI’06: Proceedings of the conference on Human Factors in computing systems*, ACM Press, New York, NY, USA, p. 275–284, 2006.
- Khoo C. S., Luyt B., Ee C., Osman J., Lim H.-H., Yong S., “How users organize electronic files on their workstations in the office environment: a preliminary study of personal information organization behaviour”, *Information Research*, vol. 11, n° 2, p. electronic edition <http://informationr.net/ir/12-2/paper293.html>, January, 2007.
- Kidd A., “The marks are on the knowledge worker”, *CHI’94: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, New York, NY, USA, p. 186–191, 1994.
- Lopistéguy P., Tricot A. (eds), *INFORSID/PeCUSI’08 : 2^e atelier Prise en Compte de l’Usager dans les Systèmes d’Information*, Éditions Inforsid, May, 2008.
- Lutters W. G., Ackerman M. S., Zhou X., “Group Information Management”, in Jones *et al.* (2007), chapter 14, p. 236–248, 2007.
- Maarek Y. S., Ben-Shaul I., “Automatically Organizing Bookmarks per Contents”, *Computer Networks and ISDN Systems*, vol. 28, n° 7-11, p. 1321–1333, 1996.
- Millen D. R., Fontaine M. A., “Improving Individual and Organizational Performance through Communities of Practice”, *GROUP’03: Proceedings of the international conference on Supporting group work*, ACM Press, New York, NY, USA, p. 205–211, 2003.
- Montaner M., López B., de la Rosa J. L., “A Taxonomy of Recommender Agents on the Internet”, *Artif. Intell. Rev.*, vol. 19, n° 4, p. 285-330, 2003.
- Reips U.-D., “The methodology of Internet-based experiments”, in A. N. Joinson, K. Y. A. McKenna, T. Postmes, U.-D. Reips (eds), *The Oxford Handbook of Internet Psychology*, Oxford University Press, New York, NY, USA, chapter 24, p. 373–390, 2007.
- Rucker J., Polanco M. J., “Sitseer: personalized navigation for the Web”, *Commun. ACM*, vol. 40, n° 3, p. 73–76, 1997.
- Sellen A. J., Harper R. H., *The myth of the paperless office*, MIT Press, Cambridge, MA, USA, 2003.
- Swarts J., “Cooperative Writing: Achieving Coordination Together and Apart”, *SIGDOC’04: Proceedings of the 22nd annual international conference on Design of communication*, ACM Press, New York, NY, USA, p. 83–89, 2004.
- Teevan J., Capra R., Quiñones M. P., “How People Find Personal Information”, in Jones *et al.* (2007), chapter 2, p. 22–34, 2007.
- Zhang J., Ackerman M. S., “Searching For Expertise in Social Networks: A Simulation of Potential Strategies”, *Group’05: Proceedings of the international conference on Supporting group work*, ACM Press, New York, NY, USA, p. 71–80, 2005.