

L'architecture CoMED pour la gestion collective de documents électroniques dans l'organisation

Guillaume Cabanac¹, Max Chevalier^{1,2},
Claude Chrisment¹, Christine Julien¹

¹*IRIT/SIG – Unité Mixte de Recherche 5505 CNRS
118 route de Narbonne
F-31062 Toulouse cedex 9*

²*LGC – Équipe d'Accueil 2043
IUT Paul Sabatier Toulouse III
129 avenue de Rangueil – BP 67701
F-31077 Toulouse cedex 4*

{cabanac,chevalier,chrisment,julien}@irit.fr

Résumé :

À l'heure actuelle, de nombreuses organisations souffrent d'une overdose informationnelle : les individus accèdent et conservent quotidiennement une quantité croissante de documents électroniques. De plus, les efforts individuels de recherche mis en œuvre ne sont pas rentabilisés et valorisés par une diffusion adaptée. Pourtant, la gestion rationnelle des documents est une condition nécessaire pour être performant dans les différentes activités liées au cycle de vie du document. C'est pourquoi nous présentons dans cet article une architecture intégrée pour la gestion collective des documents électroniques de l'organisation baptisée CoMED (*Collective Management of Electronic Documents*). Notre proposition, fondée sur l'activité d'annotation, exploite l'interdépendance des activités documentaires pour valoriser l'information introduite dans l'organisation. Son objectif consiste à aider l'individu dans ses activités documentaires, tout en faisant progresser l'organisation dans son ensemble.

Mots-clés : gestion collective de documents, architecture intégrée, activités documentaires, document électronique, annotation.

1. Contexte et problématiques de nos travaux

Comme le proposait [Seletzky, 2002] les organisations modernes peuvent être qualifiées d'« orgaNETisées » car elles reposent de plus en plus sur les Systèmes d'Information (SI) disponibles, qu'ils soient internes (un réseau de l'entreprise, un ERP...) ou externes (Extranet, Internet...). Ces SI sont d'importants vecteurs d'information pertinents pour l'organisation. Dans le même temps, ils provoquent une overdose informationnelle : l'organisation est le plus souvent incapable de traiter de façon optimale toute l'information collectée. Ce problème peut être identifié à deux principaux niveaux : individuel et collectif. Au niveau individuel, les membres de l'organisation peuvent avoir du mal à identifier, à trouver et à stocker l'information pertinente pour leurs activités par exemple. Au niveau collectif, le problème consiste surtout à propager au mieux l'information introduite dans l'organisation (*e.g.* par chacun de ses membres) afin que les individus susceptibles d'être intéressés puissent l'exploiter au mieux. En effet, nous supposons que les besoins en information des différents membres de l'organisation sont proches voire similaires, du moins au regard de leurs activités dans l'organisation. Une solution à ces problèmes permettrait de faire « vivre » l'information qui est trop souvent statique, stérile et dispersée dans les méandres de l'organisation. Dans cet article, nous proposons d'étudier ces problèmes au travers des activités liées aux informations et plus particulièrement liées à leur support : les documents. Ces activités issues du cycle de vie d'un document sont présentées dans la deuxième section, selon leurs caractéristiques individuelles mais également collectives, couvrant ainsi les problématiques exposées. Enfin, nous proposons dans la troisième section une architecture de SI interne à l'organisation lui permettant d'exploiter au mieux ses documents. Cette architecture nommée CoMED (*Collective Management of Electronic Documents*) repose sur un élément central : l'activité d'annotation, ainsi que sur une approche collective – englobant les activités collaboratives et coopératives – permettant l'amélioration mutuelle des différentes activités documentaires. Ainsi, le résultat d'une activité augmente automatiquement la performance des autres activités. Cette architecture est illustrée à partir de documents issus du média pouvant être vu aujourd'hui comme une source privilégiée d'informations : le Web.

2. Les activités documentaires de l'organisation

Dans l'organisation, la gestion des documents peut être perçue comme un facteur de performance ; elle repose sur une optimisation de différentes activités facilitant l'accès aux documents et, par conséquent, aux informations qu'ils contiennent. Ces activités forment le cycle de vie du document *cf.* figure 1 [Sellen et Harper, 2003, p. 203]. Les sections suivantes illustrent les activités identifiées (numérotées de ① à ⑥) qui peuvent être réalisées individuellement mais aussi collectivement, ce qui permet de tirer parti d'un groupe.

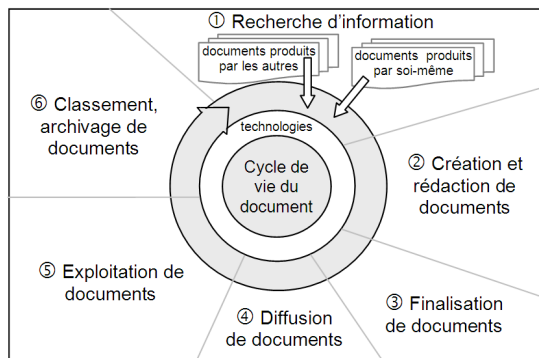


FIG. 1 – Cycle de vie du document [Sellen et Harper, 2003, p. 203]

2.1. Recherche d'information et navigation ①

Le challenge quotidien de l'organisation moderne consiste à fournir à chacun de ses membres des informations utiles pour son activité. Pour ce faire, l'individu s'inscrit dans une activité de Recherche d'Information (RI) ① qui nécessite d'alterner sans réellement s'en rendre compte entre deux tâches distinctes : la recherche et la navigation [Hölscher et Strube, 2000]. Ces tâches peuvent être réalisées individuellement ou collectivement.

Activité individuelle. En phase de recherche, un individu explicite ses besoins sous la forme d'une requête qu'il soumet à un outil de recherche appelé moteur sur le Web. Ce dernier restitue alors à l'utilisateur une liste de documents triée par pertinence (système) décroissante. Il est à noter que d'autres types de visualisation ont été proposés au travers d'interfaces graphiques évoluées [Chen, 2006]. En parallèle, l'utilisateur navigue dans l'espace des documents disponibles sans avoir à formuler ses besoins mais également sans connaître *a priori* ni le contenu ni l'organisation de l'espace des documents [Agosti et Smeaton, 1996]. Différentes approches ont été proposées afin d'optimiser ces deux tâches complémentaires : la reformulation de requête, le réordonnancement de la liste des résultats voire la personnalisation de cette dernière pour améliorer la tâche de recherche. Concernant la navigation, les propositions suivent deux principaux courants. D'une part, des accélérateurs de navigation tels que *Letizia* [Lieberman, 1995] proposent à l'internaute les liens de l'hypertexte local qui sont jugés les plus pertinents pour la navigation courante. D'autre part, des systèmes proposent des documents intéressants par rapport à ceux consultés durant la navigation de l'utilisateur, en interrogeant un outil annexe *e.g.* *WBI* [Barrett et al., 1997].

Activité collective. On observe dans la vie courante que les individus se rassemblent en groupes sociaux ou organisationnels pour résoudre des problèmes de recherche [Karamuftuoglu, 1998]. Ainsi, des systèmes reposant sur une approche collective permettent à un utilisateur de tirer profit des expériences

de recherche et de navigation d'autres individus qui possèdent des centres d'intérêts proches. Par exemple, pour la tâche de recherche, *Cosydor* [Jeribi et al., 2001] utilise les expériences de recherche des utilisateurs pour aider un individu à formuler ses besoins, en amont de la RI. Dans *IronWeb* [Dussaux et Pécuchet, 2000] les individus peuvent bénéficier des connaissances d'experts en cherchant dans les hiérarchies de signets de ces derniers. De façon plus générale, le logiciel *Human-Links* permet la RI pair-à-pair dans le corpus constitué par les documents d'individus connectés sur le Web lors de la recherche. Par ailleurs, *VR-Vibe* [Benford et al., 1995] représente les résultats des recherches du groupe dans un environnement virtuel en 3D explorable par l'utilisateur. Enfin, l'aspect collectif dans la navigation est également présent dans *Broadway* [Jaczynski et Trousse, 1999] qui utilise des techniques de raisonnement par cas pour anticiper la navigation d'un utilisateur en fonction des navigations d'autres personnes précédemment observées.

2.2. Création ② et finalisation ③ de documents

À partir des ressources trouvées ①, les membres organisationnels produisent une valeur ajoutée qu'ils communiquent en l'explicitant au moyen de rédactions individuelles et/ou collectives de documents à l'aide d'outils adaptés.

Activité individuelle. La rédaction ② de documents est facilitée par des logiciels de traitement de texte comme *OpenOffice Writer* ou *Microsoft Word* qui permettent la mise en page de textes, schémas, graphiques... En phase de relecture, préalablement à la finalisation ③ du document, l'individu peut ajouter des commentaires, reformuler des passages, identifier des erreurs typographiques, etc. grâce à la fonctionnalité correspondante du logiciel utilisé.

Activité collective. La phase de rédaction ② peut être réalisée collectivement grâce à un Wiki (un site Web entièrement modifiable par ses lecteurs), dont l'intérêt pour la création, la relecture et la diffusion des documents est montrée dans [Guzdial et al., 2000]. Similairement, des produits comme *Microsoft Shared Point Services* permettent à plusieurs individus d'éditer de façon synchrone un même document : en phase de finalisation ③, chaque relecteur voit les modifications apportés par l'ensemble des collaborateurs.

2.3. Diffusion ④ manuelle et automatique de documents

L'organisation, par le biais des activités précédentes, doit être capable de gérer un grand nombre de documents générés par ses activités ou bien issus de sources externes. Pour optimiser l'impact des informations contenues dans les documents, il est important que ceux-ci soient diffusés ④ de manière adéquate dans l'organisation. Ces documents diffusés manuellement ou de façon automatique aux différents membres susceptibles d'être intéressés permettront une meilleure exploitation de l'information tout en évitant sa perte dans les méandres de l'organisation.

Activité individuelle. Traditionnellement, la diffusion manuelle d'un document aux différents membres de l'organisation est réalisée grâce au courrier électronique. Cette diffusion est hautement cognitive pour l'individu qui la réalise car elle nécessite une connaissance de l'organisation et de ses membres, ce qui représente une limite non négligeable dans un contexte organisationnel de grande taille. En effet, un membre recevant une information ne pourra la diffuser correctement que s'il connaît les personnes susceptibles d'être intéressées, ce qui est peu probable dans une grande structure. La création de listes de diffusion pointant vers un ensemble préétabli de courriels simplifie le problème sans toutefois le résoudre : l'individu doit connaître les centres d'intérêts de chaque liste. Une autre approche consiste à formaliser les processus métier et leurs caractéristiques (actions, délais associés et leur ordonnancement, intervenants et leurs rôles, données nécessaires et/ou produites) pour spécifier un *workflow* [Marshak, 1994]. Ce dernier achemine automatiquement les documents en fonction d'événements qui déclenchent les règles de distribution définies *a priori*. Cependant, il est difficile de connaître l'ensemble des personnes intéressées par un sujet donné. C'est pourquoi une diffusion plus large est parfois souhaitable au travers de portails organisationnels proposant des techniques de recherche d'information par exemple.

Activité collective. Afin de limiter l'effort nécessité par une recherche d'information active, un SI de type PUSH propose automatiquement des documents aux individus (contrairement à l'approche PULL qui nécessite une démarche active de leur part *cf.* section 2.1.). Ainsi, un système de filtrage d'information « amène à l'utilisateur les documents qui vont lui permettre de satisfaire son besoin en information » [Belkin et Croft, 1992] en construisant et en faisant évoluer un modèle d'utilisateur appelé « profil utilisateur ». Il existe plusieurs types de filtres caractérisés par des modèles distincts d'appariement entre l'utilisateur et les documents. Nous pouvons citer *Syskill & Webert* [Pazzani et al., 1996] comme exemple de filtrage cognitif car il recommande un document à un individu lorsque son contenu est similaire au profil de l'utilisateur. Une autre approche qualifiée de filtrage collaboratif consiste à s'appuyer sur une communauté d'utilisateurs comme dans *GroupLens* [Konstan et al., 1997] qui exploite les jugements des individus vis-à-vis des documents pour identifier des groupes d'opinion au sein desquels les documents sont diffusés. Par conséquent, seuls les jugements des utilisateurs doivent être connus du système qui n'a ainsi pas besoin d'indexer les documents, contrairement au filtrage cognitif. Une évolution de ces approches consiste à tirer parti de l'organisation même de la communauté d'utilisateurs exploitée par ailleurs dans le filtrage collaboratif. Nous pouvons citer par exemple [Zhang et Ackerman, 2005] qui présentent des approches exploitant le réseau social (les relations entre individus) en complément des jugements émis par les utilisateurs. De tels systèmes privilégient les documents issus des contacts les plus proches d'un individu, en faisant l'hypothèse qu'il saura davantage apprécier la pertinence de ces sources.

2.4. Exploitation ⑤ et classement de documents ⑥

Chaque membre de l'organisation est un « gestionnaire » potentiel des documents capitalisés collectivement. Il doit donc disposer d'outils adaptés pour gérer et exploiter au mieux les documents, qu'ils proviennent de sources internes ou externes à l'organisation *e.g.* du Web.

Activité individuelle. L'individu consulte ⑤ les documents électroniques sur son écran ou bien sur papier après impression. Outre la lecture, les affordances¹ du papier encouragent le lecteur à personnaliser le texte, à se l'approprier en formulant des annotations. Concrètement, la création d'annotations permet aux individus de matérialiser leur réflexion critique en reformulant, commentant, corrigeant, etc. des passages précis du document qu'ils lisent. Cette activité favorisant l'apprentissage est appelée « lecture active » [Adler et van Doren, 1972] par opposition à la lecture de loisir. En réponse au besoin d'annoter les documents électroniques (une expérience [Sellen et Harper, 2003, p. 95] montre que les individus se déclarent être frustrés de ne pas pouvoir le faire), de nombreux logiciels appelés Systèmes d'Annotation (SA) permettent l'annotation informelle² de tout ou partie d'un document sur le Web. Concernant le stockage ⑥, la hiérarchie thématique *ad hoc* de répertoires est une organisation très répandue *cf.* les systèmes de gestion de fichiers, le classement hiérarchique pour les messageries électroniques. . . C'est également le cas sur le Web où la hiérarchie de signets³ est un véritable espace d'information personnel pour l'utilisateur qui construit une représentation explicite et organisée des documents reflétant ses centres d'intérêt. D'après [Abrams et al., 1998] cette structure évolue de façon assez rapide car un utilisateur y ajoute trois à quatre signets en moyenne lors d'une session de navigation.

Activité collective. La facilité de partage est un avantage important en faveur des annotations électroniques par rapport à leurs homologues papier. Ainsi, le SA *Amaya* [Kahan et al., 2002] permet entre autres de consulter des documents électroniques accompagnés des annotations formulées par leurs lecteurs : l'individu n'est plus restreint au seul point de vue défendu par l'auteur du document car il bénéficie des analyses des lecteurs/annotateurs précédents. Par ailleurs, [Cabanac et al., 2006] propose à chaque lecteur de réagir à des passages d'un document grâce à une annotation de type donné décrivant l'opinion exprimée (confirmation ou réfutation) ainsi que le contenu du commentaire (exemple, correction, question) par exemple. Par la suite, un individu peut réagir à une annotation ou, de façon récursive, à une autre réaction, ce

¹ Propriétés naturelles actionnables entre le monde et un individu *e.g.* les propriétés physiques du papier (fin, léger, poreux, opaque et flexible) suggèrent les actions humaines de saisie, transport, pliage, écriture. . . [Sellen et Harper, 2003, p. 16]

² Dans cet article, nous considérons des annotations *informelles* [Marshall, 1998] également qualifiées de *cognitivement sémantiques* [Zacklad et al., 2003] car leur contenu n'est pas contraint à un vocabulaire contrôlé *e.g.* issu d'une ontologie.

³ Le terme signet a pour synonymes « favori » en français et « bookmark » en anglais.

qui forme une arborescence de réactions dont la racine est une annotation ancrée sur le document. Cette structuration hiérarchique et chronologique des réactions est appelée « fil de discussion ». Concernant le stockage, les membres de l'organisation peuvent centraliser © leurs documents dans un entrepôt de documents [Khrouf et Soulé-Dupuy, 2004] afin d'établir une source commune d'information. Cette vision du partage peut être également remarquée au travers de systèmes de partage de hiérarchies de signets [Kanawati et Malek, 2000]. À contre-pied de la structuration hiérarchique, le phénomène récent de *folksonomy* c'est-à-dire de « classification sociale » consiste à associer aux documents des termes choisis librement appelés *tags*. Ainsi, la maison d'édition *Nature* propose le système *Connotea* [Lund et al., 2005] qui permet de partager des documents « taggués » et de découvrir de nouvelles informations en explorant l'ensemble des documents collectivement constitué.

2.5. Synthèse sur les activités documentaires

Au travers des activités documentaires, nous pouvons constater que la prise en compte de l'expérience, des activités, des contacts, etc. d'un groupe permet de mettre en œuvre plusieurs systèmes distincts visant à exploiter de manière efficace les documents dans l'organisation. Ces systèmes sont basés sur le principe donnant/donnant car un individu œuvre pour un groupe et, réciproquement, un groupe œuvre pour un individu. Cependant, au regard des exemples mentionnés nous observons que les multiples systèmes se concentrent sur une, voire deux activités au maximum. Or, nous pensons que les approches actuelles ne prennent pas suffisamment en compte la réalité. En effet, l'être humain ne réalise pas ces différentes activités linéairement et de façon cloisonnée comme le montre la figure 1 : il peut chercher de l'information ①, commencer à écrire un document ② et chercher à nouveau pour approfondir un point précis ①. Comme conséquence à cette spécialisation, chaque système ne prend en compte qu'une partie de l'activité réelle de l'individu en délaissant les autres activités. Partant de cette constatation, nous proposons dans cet article une approche collective, globale et intégrée nommée CoMED (*Collective Management of Electronic Documents*) qui vise à couvrir l'ensemble des activités documentaires. Notre approche permet de faciliter mutuellement les différentes activités réalisées. Notre proposition repose essentiellement sur l'activité d'annotation qui permet aux membres organisationnels de s'appropriier et de mémoriser les documents qu'ils jugent pertinents. Ainsi, l'activité de classement © qui vise à conserver des documents pertinents au travers des annotations constitue l'élément central de l'approche CoMED.

3. CoMED : gestion collective de documents

La gestion collective de documents doit recouvrir l'ensemble des activités documentaires de l'organisation pour optimiser son efficacité. Cependant, les différents outils examinés dans la section précédente sont principalement li-

mités à une seule activité ; par conséquent, chaque système construit une vision distincte et partielle de l'organisation et de l'individu. C'est pourquoi nous proposons une intégration de toutes les activités documentaires par le biais d'une démarche principalement collective. Pour cela nous avons identifié l'élément fédérateur de notre approche : l'activité de **stockage** ⑥. En effet, nous constatons que toutes les tâches effectives de l'individu sont assujetties au stockage : une personne qui trouve des documents sur le Web ou qui crée des documents – en exploitant notamment des ressources déjà conservées – va vraisemblablement stocker tout ou partie de ces documents. Au regard de l'usage traditionnel fait des documents, nous pouvons souligner que cette activité traduit une appropriation personnelle de ces derniers, qui peuvent être stockés dans un système de fichiers au travers d'une arborescence de répertoires, par exemple.

Nous avons également souligné dans la section précédente que certaines approches se basent sur les hiérarchies de signets pour conserver les documents suscitant un intérêt pour l'individu. Elles représentent de véritables espaces personnels d'informations [Abrams et al., 1998]. Toutefois, ces hiérarchies souffrent de certaines limites dont la sous-information car le contenu des signets est très limité et même parfois peu pertinent. Par conséquent l'utilisateur peine parfois à se souvenir de la raison pour laquelle il a créé un signet vers un document sans avoir à en visualiser le contenu [Maarek et Ben-Shaul, 1996]. Nous pensons que l'individu gagnerait à créer des annotations à la place des signets parce qu'elles permettraient une mémorisation d'information plus riche : alors que le signet désigne obligatoirement un document dans son intégralité, le point d'ancrage de l'annotation est plus précis car défini sur tout ou partie du document (un paragraphe, une phrase, un mot, etc.). De plus, l'annotateur peut spécifier un commentaire et en donner un aperçu en y associant un type *e.g.* confirmation, réfutation, question. D'un point de vue collectif, les annotations permettent les débats au sein de fils de discussions. C'est pourquoi CoMED est basée sur des annotations qui sont exploitées comme vecteurs de stockage des documents pour l'individu. Notre approche est de type donnant/donnant car chaque individu, au travers de ses activités, participe automatiquement aux activités d'un groupe. En contrepartie, il bénéficie à son tour de l'activité des autres membres du groupe. Ainsi, même si l'activité au travers de cette application peut sembler importante, le gain que l'individu peut obtenir n'est pas négligeable. Afin de présenter la démarche adoptée dans CoMED, nous définissons tout d'abord la notion de hiérarchie d'annotations argumentatives, puis les différents processus mis en œuvre dans notre architecture.

3.1. Hiérarchie d'annotations argumentatives

Dans CoMED, un utilisateur mémorise tout ou partie d'un document en créant une annotation. Afin de faciliter l'accès à ces documents qui ont fait l'objet d'un intérêt, nous proposons à chaque individu d'organiser thématiquement les annotations créées dans une Hiérarchie d'Annotations (HdA) personnelle ; une HdA possède une structure arborescente similaire à celle bien connue d'une

hiérarchie de signets, ou plus généralement d'un système de fichiers. Pour choisir le répertoire d'accueil de son annotation, l'individu doit réaliser un effort cognitif [Rucker et Polanco, 1997] qui consiste à sélectionner celui qui contient les annotations les plus similaires à l'annotation candidate. Cet effort cognitif individuel sera exploité pour améliorer les activités du groupe.

En termes de définition, les annotations formulées par les individus contiennent des données objectives *DO* ainsi que des informations subjectives *IS*. En effet, CoMED crée des *DO* pour mémoriser les attributs de l'annotation suivants : son **identification**, l'identité de son **créateur** qui donne accès à ses caractéristiques (identité, courriel, etc.) ; sa **date de création** qui permet d'organiser le fil de discussion chronologiquement (*cf.* section 2.4.) ainsi que son **point d'ancrage** qui spécifie de manière non ambiguë son emplacement au sein de la ressource annotée – différentes techniques proposées sont applicables au contexte des documents semi-structurés *e.g.* *XPointer* [Kahan et al., 2002]. D'autre part, les *IS* sont formulées par les annotateurs ; elles peuvent être omises et comprennent : le **contenu** de l'annotation ainsi que sa **visibilité** (privée ou publique) qui permet de restreindre la portée de l'annotation, cette donnée renseigne CoMED quant à la diffusabilité de l'annotation.

De plus, nous proposons de conserver l'**expertise** de l'annotateur car les individus accordent en général davantage de crédit à l'opinion d'un expert qu'à celle d'un novice [Marshall, 1998]. Pour étayer ses remarques, l'annotateur peut également spécifier la **liste des références** sur lesquelles il s'appuie. Enfin, il peut qualifier la sémantique du contenu de son annotation en y associant **différents types** qui sont, dans notre approche, liés au commentaire (correction, exemple ou question) ou à l'opinion de l'annotateur (confirmation ou réfutation) dans le cas d'une annotation argumentative (c'est une réaction dans le fil de discussion). Par ailleurs, le type jugement (positif ou négatif) permet aux annotateurs de spécifier leur point de vue suggestif. Enfin, une annotation peut susciter des réactions (récursivement annotables) organisées chronologiquement au sein d'un fil de discussion.

3.2. Aide à la gestion collective et intégrée de documents

Nous proposons dans cette section une vue synthétique de l'approche CoMED (*cf.* figure 2). Elle met en perspective le caractère global et intégré de notre architecture de SI qui couvre l'ensemble des activités documentaires. Notre architecture comporte quatre processus interdépendants car le résultat d'un processus est conjointement amélioré par celui des trois autres. Le premier processus (RECO) diffuse ④ les documents capitalisés par l'organisation sur la base des centres d'intérêts des membres ; le deuxième (RÉORG) permet la réorganisation ⑥ thématique des espaces personnels d'annotations construits à l'aide de RECO. Le troisième processus (NAVI) émet des recommandations en fonction de la navigation ① courante de l'utilisateur, en se basant sur les HdA structurées par RÉORG. Le dernier processus (VALID) calcule la validité sociale des annotations au travers des FdD, ce qui permet à l'utilisateur d'estimer

le consensus global suscité par une annotation en phase de création ②, de finalisation ③ ou de lecture active ⑤ pour décider de son traitement (et, de façon indirecte, de sa prise en compte éventuelle par RECO). Nous détaillons ces processus dans les sections suivantes.

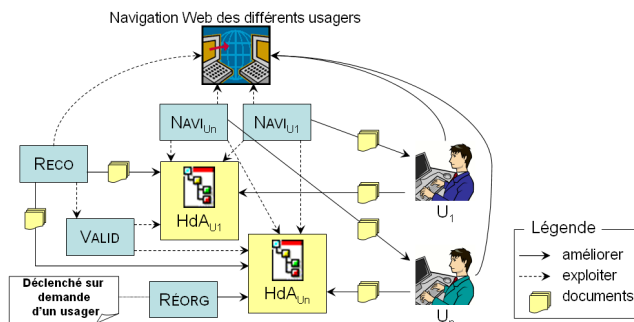


FIG. 2 – Vue synthétique de l'architecture CoMED

3.2.1. Processus de diffusion (RECO) et de classement (RÉORG) de documents issus de la navigation collective

Au sein de l'organisation, les documents trouvés ① par chacun des membres sont rarement diffusés ④ car cela demande une démarche active hautement cognitive *cf.* section 2.3. Ainsi l'effort de recherche n'est pas factorisé, ce qui entraîne une perte d'efficacité et un coût non négligeables pour l'organisation. C'est pourquoi nous proposons le processus de recommandation RECO qui automatise la diffusion des documents dans l'organisation. En entrée du processus, le corpus de recommandation est constitué de l'intégralité des documents visualisés dans l'organisation car ils ont suscité l'intérêt d'au moins un individu. En sortie du processus, des recommandations sont insérées dans les HdA des individus intéressés et plus précisément dans le répertoire le plus adéquat sur la base d'une comparaison thématique. Une telle recommandation est en réalité une annotation dont la granularité du point d'ancrage est maximale : il est défini sur l'intégralité du document. Concrètement, le processus RECO réalise un parcours en profondeur d'abord de chaque HdA pour recommander les documents dans les répertoires les plus spécifiques tout en évitant plusieurs recommandations du même document dans un même chemin de la HdA. Le processus évalue alors une distance thématique entre le document à recommander et chaque répertoire r_i . Pour ce faire, un classifieur $C_i = \langle f_i, \tau_i \rangle$ est associé à chaque r_i , c'est un filtre qui accepte le document d lorsque $f_i(d) \geq \tau_i$. Ce filtre est construit à partir des annotations de type « jugement positif » car nous sommes certains qu'elles représentent un intérêt pour l'utilisateur, contrairement à des annotations qualifiées avec des types différents. Les éléments clés du processus RECO, évoqués sommairement ici, sont détaillés dans [Chevalier et Julien, 2003]. Déclenché à intervalle régulier, il permet à chaque individu de bénéficier de l'activité de recherche collective.

Les documents recommandés sous forme d'annotation sont insérés directement dans les HdA des individus sous la forme d'une annotation. Une telle introduction de recommandations peut amener un utilisateur à vouloir réorganiser sa HdA. Comme cette tâche nécessite un effort cognitif important [Rucker et Polanco, 1997], nous proposons d'aider l'individu à réorganiser thématiquement tout ou partie de sa HdA grâce au processus RÉORG. Ce dernier est basé sur l'algorithme de classification ascendante hiérarchique [Jardine et van Rijsbergen, 1971] couplé à une technique d'étiquetage des classes obtenues et de seuillage permettant à l'utilisateur de spécifier la profondeur de la hiérarchie réorganisée [Chevalier et Julien, 2003].

3.2.2. Processus d'aide à la navigation (NAVI)

La navigation d'un individu profite au groupe (RECO) mais aussi à l'individu lui-même : CoMED lui propose les annotations en rapport avec sa navigation courante (NAVI) qui proviennent des HdA de ses collègues. Ainsi, à contre-pied des approches basées sur l'hypertexte local ou sur des outils tiers *cf.* section 2.1., nous faisons l'hypothèse qu'un utilisateur peut tirer parti, lors de sa navigation, des ressources conservées par d'autres utilisateurs ayant des centres d'intérêt proches de son besoin momentané, caractérisé par sa navigation. C'est pourquoi le processus de recommandation NAVI se base en entrée sur la navigation courante d'un individu pour lui proposer en sortie un ensemble d'annotations pertinentes. Ces recommandations sont issues des HdA des membres organisationnels. Concrètement, le processus se base sur la structuration des HdA (résultant des efforts cognitifs individuels) pour recommander les annotations les plus proches du document en cours de navigation, tout en privilégiant les documents qui appartiennent au plus grand nombre de HdA. Dans cette approche, la proximité est calculée au travers des HdA ; elle est égale à la distance moyenne entre le document à recommander et le document visité. Notons que les annotations à la racine des HdA ne sont pas prises en compte car elles n'ont pas fait l'objet d'un effort cognitif de classement. L'intégralité de ce processus de filtrage est détaillée dans [Chevalier et Julien, 2003].

Les recommandations émises par CoMED lors de la navigation de l'individu lui permettent de découvrir des ressources jugées intéressantes par d'autres membres organisationnels. Ces derniers ont pu annoter les ressources et débattre de points de vue divergents au sein de fils de discussion. De telles annotations argumentatives, véritables valeurs ajoutées aux documents, sont des sources d'informations complémentaires pour l'individu.

3.2.3. Processus de validation sociale d'annotations (VALID)

L'individu engagé dans des activités collaboratives dans l'organisation a parfois besoin d'identifier les propositions validées socialement parmi toutes celles présentes sur un document, c'est-à-dire celles qui font consensus. Par exemple, en phase de rédaction collective ② il est nécessaire d'identifier les

idées qui font consensus au sein du groupe ; en phase de relecture ③ il semble cohérent de prendre en compte les remarques à propos desquelles s'accordent de nombreuses personnes *e.g.* une réfutation confirmée par l'ensemble des relecteurs. Enfin, lors de la lecture de documents ⑤, repérer les annotations qui font consensus permet d'obtenir une valeur ajoutée sensée alors que celles qui suscitent des débats permettent d'identifier les arguments de leurs détracteurs et défenseurs. En fait, évaluer le degré de consensus d'une annotation argumentative *i.e.* sa validité sociale nécessite d'extraire l'opinion (confirmation ou réfutation graduelles) de chaque réaction de son FdD et d'en faire une synthèse mentalement. Cette tâche entraîne une surcharge cognitive qui doit pourtant être évitée car elle perturbe l'activité principale de l'individu : sa lecture [O'Hara et Sellen, 1997]. C'est pourquoi le processus VALID calcule la « validité sociale » de chaque annotation au travers du fil de discussion associé. Cette valeur représente le degré de consensus des participants au débat. Concrètement, nous modélisons les différentes réactions (qui sont des annotations) sous la forme d'un graphe d'arguments reliés par des arcs étiquetés selon le type de chaque argument *e.g.* réfutation, confirmation, correction. Une valorisation de ce graphe détaillée dans [Cabanac et al., 2006] permet de connaître le degré d'accord entre les intervenants. Cette valeur peut être exploitée pour mettre en emphase les annotations qui font consensus et pour décider de leur traitement par RECO. Cela permet aux individus de se focaliser sur les débats qui ont abouti à un accord durant l'exploitation des documents ④ ou sur les remarques les plus appuyées des relecteurs ③ par exemple. À l'opposé, un procédé indiquant les annotations qui ne font pas consensus est également possible.

3.3. Discussion

Notre architecture CoMED souffre de quelques limites que nous discutons dans cette section. Nous faisons l'hypothèse que les individus créent et font évoluer une hiérarchie thématique contenant leurs annotations. La structuration thématique est en effet couramment utilisée, bien que d'autres types d'organisations existent (*e.g.* par auteur, par date de lecture, par tâche) et, dans de telles configurations, notre architecture n'est pas adaptée. De plus, en considérant que les annotations sont organisées hiérarchiquement, nous supposons qu'elles reflètent les centres d'intérêts de l'utilisateur à l'image des hiérarchies de signets [Abrams et al., 1998]. Or, cette supposition devrait être vérifiée par des études comportementales avec des utilisateurs. Par ailleurs, nous ne prenons en compte actuellement que les annotations de type « jugement positif » car nous considérons que seul ce type traduit un intérêt explicite de l'utilisateur. Or, nous devrions nous demander si d'autres types peuvent également indiquer un intérêt. De même, nous n'exploitons pas pour le moment ce qui différencie les annotations des signets (point d'ancrage à granularité ajustable, présence de commentaire, de références sous la forme de liens hypertextes qui étayent un argument *e.g.* article scientifique, de types, notion d'expertise subjective de l'annotateur). L'architecture CoMED proposée dans cet article

fait actuellement l'objet de réalisations logicielles qui sont au stade de prototypes. Nous désirons à terme évaluer nos contributions par des expérimentations en milieu écologique *i.e.* avec d'authentiques utilisateurs. Nous désirons notamment considérer certains résultats d'études comportementales [Beenen et al., 2004] qui montrent, dans des contextes applicatifs précis, qu'un individu participe davantage lorsqu'il sait que ses activités bénéficient à d'autres personnes.

4. Conclusion et perspectives

Cet article présente un état de l'art des systèmes proposés pour faire bénéficier l'individu des compétences collectives de l'organisation au sein de chaque activité documentaire. L'apport de ces systèmes est réel. Toutefois nous remarquons que leur spécialisation pour telle ou telle activité dans le cycle de vie du document est une limite importante. En effet, la caractérisation de l'utilisateur dans une activité ne profite qu'à l'amélioration de celle-ci alors que toutes les autres pourraient en bénéficier car elles sont interdépendantes. C'est pourquoi nous proposons l'architecture originale CoMED (*Collective Management of Electronic Documents*). Notre approche est constituée de processus intégrés et interdépendants qui couvrent les activités documentaires réalisées dans une organisation. L'objectif principal de CoMED vise à diffuser les documents introduits dans l'organisation par les individus afin que toutes les personnes intéressées en bénéficient automatiquement. Ainsi, l'information n'a pas à être cherchée à nouveau par les individus car elle leur est automatiquement proposée en fonction de leurs centres d'intérêt.

En termes de perspectives à notre approche, nous envisageons de construire le réseau social de l'organisation grâce à l'analyse des débats conduits au sein des fils de discussion. Cette connaissance permettrait d'améliorer l'apprentissage des thématiques d'intérêt de chaque membre organisationnel pour affiner la pertinence des recommandations. D'autre part, notre approche ne permet pas à l'utilisateur de consulter la représentation que CoMED a de lui. Pourtant, la transparence du système permettrait aux utilisateurs de mieux comprendre les propositions émises par le système, ils pourraient également mettre à jour les informations qui les concernent ; cela ne peut qu'encourager les individus à participer activement dans CoMED. Enfin, pour que notre système apporte une réelle plus-value à l'organisation dans son ensemble, nous désirons encourager la participation des membres organisationnels en termes de débats constructifs argumentés, de notes de lectures pertinentes, etc.

Références

- Abrams, D., Baecker, R., et Chignell, M. (1998). Information archiving with bookmarks : personal web space construction and organization. Dans *CHI '98 : Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 41–48, New York, NY, USA. ACM Press/A-W.
- Adler, M. J. et van Doren, C. (1972). *How to Read a Book*. Simon & Shuster.

- Agosti, M. et Smeaton, A. F. (1996). *Information Retrieval and Hypertext*. Kluwer Academic Publishers, Dordrecht.
- Barrett, R., Maglio, P. P., et Kellem, D. C. (1997). How to Personalize the Web. Dans *CHI '97 : Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 75–82, New York, NY, USA. ACM Press.
- Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., et Kraut, R. E. (2004). Using Social Psychology to Motivate Contributions to Online Communities. Dans *CSCW '04 : Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 212–221, New York, NY, USA. ACM Press.
- Belkin, N. J. et Croft, W. B. (1992). Information Filtering and Information Retrieval : Two sides of the Same Coin? *Communications of the ACM*, 35(12) :29–38.
- Benford, S., Snowdon, D., Greenhalgh, C., Ingram, R., Knox, I., et Brown, C. (1995). VR-VIBE : A Virtual Environment for Co-operative Information Retrieval. *Computer Graphics Forum*, 14(3) :349–360.
- Cabanac, G., Chevalier, M., Chrisment, C., et Julien, C. (2006). Validation sociale d’annotations collectives : argumentation bipolaire graduelle pour la théorie sociale de l’information. Dans *INFORSID'06 : 24^e congrès de l'INformatique des Organisations et Systèmes d'Information et de Décision*, pages 467–482. Éditions Inforsid.
- Chen, C. (2006). *Information Visualization : Beyond the Horizon*. Springer.
- Chevalier, M. et Julien, C. (2003). Interface adaptative et coopérative pour l’aide à la Recherche d’Information sur le Web. *Information - Interaction - Intelligence*, 3(2) :47–73.
- Dussaux, G. et Pécuchet, J.-P. (2000). Création collective de bases de connaissances sur le web : Indexation par l’usage des documents. Dans *CIDE 2000 : 3^e Colloque International sur le Document Électronique*, pages 185–203. Europa.
- Guzdial, M., Rick, J., et Kerimbaev, B. (2000). Recognizing and Supporting Roles in CSCW. Dans *CSCW '00 : Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 261–268, New York, NY, USA. ACM Press.
- Hölscher, C. et Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks*, 33(1-6) :337–346.
- Jaczynski, M. et Trousse, B. (1999). Broadway : A Case-Based System for Cooperative Information Browsing on the World-Wide-Web. Dans Padget, J. A., éditeur, *Collaboration between Human and Artificial Societies*, volume 1624 de *Lecture Notes in Computer Science*, pages 264–283. Springer.
- Jardine, N. et van Rijsbergen, C. J. (1971). The Use of Hierarchic Clustering in Information Retrieval. *Information Storage and Retrieval*, 7(5) :217–240.
- Jeribi, L., Rimpler, B., et Pinon, J.-M. (2001). Système d’aide à la recherche et à l’interrogation de bases documentaires, fondé sur la réutilisation d’expériences. Dans *INFORSID'01 : 19^e congrès de l'INformatique des Organisations et Systèmes d'Information et de Décision*, pages 443–463.

- Kahan, J., Koivunen, M.-R., Prud'Hommeaux, E., et Swick, R. R. (2002). Annotea : an open rdf infrastructure for shared web annotations. *Computer Networks*, 32(5) :589–608.
- Kanawati, R. et Malek, M. (2000). Informing the Design of Shared Bookmark Systems. Dans *RIAO 2000 : actes de la conférence Recherche d'Information Assistée par Ordinateur*, pages 170–179, Paris, France.
- Karamuftuoglu, M. (1998). Collaborative Information Retrieval : Toward a Social Informatics View of IR Interaction. *Journal of the American Society for Information Science*, 49(12) :1070–1080.
- Khrouf, K. et Soulé-Dupuy, C. (2004). A Textual Warehouse Approach : A Web Data Repository. Dans Mohammadian, M., éditeur, *Intelligent Agents for Data Mining and Information Retrieval*, chapitre 7, pages 101–124. Idea Publishing Group.
- Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., et Riedl, J. (1997). GroupLens : Applying Collaborative Filtering to Usenet News. *Communications of the ACM*, 40(3) :77–87.
- Lieberman, H. (1995). Letizia : An Agent That Assists Web Browsing. Dans *IJCAI'95 : International Joint Conference on AI*, pages 924–929.
- Lund, B., Hammond, T., Flack, M., et Hannay, T. (2005). Social Bookmarking Tools (II) : A Case Study - Connotea. *D-Lib Magazine*, 11(4).
- Maarek, Y. S. et Ben-Shaul, I. (1996). Automatically Organizing Bookmarks per Contents. *Computer Networks*, 28(7-11) :1321–1334.
- Marshak, R. T. (1994). Workflow White Paper : an Overview of Workflow Software. Dans Bierman, B., éditeur, *Workflow'94*. The Conference Group.
- Marshall, C. C. (1998). Toward an ecology of hypertext annotation. Dans *HYPertext '98 : Proceedings of the 9th ACM conference on Hypertext and hypermedia*, pages 40–49, New York, NY, USA. ACM Press.
- O'Hara, K. et Sellen, A. (1997). A Comparison of Reading Paper and On-Line Documents. Dans *CHI '97 : Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 335–342, New York, NY, USA. ACM Press.
- Pazzani, M. J., Muramatsu, J., et Billsus, D. (1996). Syskill & Webert : Identifying Interesting Web Sites. Dans *AAAI/IAAI*, volume 1, pages 54–61.
- Rucker, J. et Polanco, M. J. (1997). SiteSeer : personalized navigation for the web. *Communications of the ACM*, 40(3) :73–76.
- Seletzky, S. (2002). *L'entreprise orgaNETisée*. InfoPro. Dunod.
- Sellen, A. J. et Harper, R. H. (2003). *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, USA.
- Zacklad, M., Lewkowicz, M., Boujut, J.-F., Darses, F., et Détienne, F. (2003). Formes et gestion des annotations numériques collectives en ingénierie collaborative. Dans Dieng, R., éditeur, *IC2003*, pages 207–224, France. PUG.
- Zhang, J. et Ackerman, M. S. (2005). Searching For Expertise in Social Networks : A Simulation of Potential Strategies. Dans *GROUP '05 : Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pages 71–80, New York, NY, USA. ACM Press.