

Annotation collective pour l'intégration des activités documentaires

Unifying Document-related Activities with Collective Annotation

Guillaume Cabanac*

IRIT-SIG/D2S2 – UMR 5505 CNRS

Université Toulouse 3 – 118 route de Narbonne – F-31062 Toulouse cedex 9

cabanac@irit.fr

Résumé

Cet article traite de l'activité d'annotation réalisée sur les documents du Web. Nos travaux visent à unifier autour du concept d'annotation les activités documentaires qui sont actuellement cloisonnées et peu valorisées. Ainsi, les productions et les interactions des individus sont considérées afin d'aider chacun à réaliser ses propres activités. Nos propositions, implantées dans le système d'annotation TafAnnote, sont en cours d'expérimentation.

Mots Clef

Annotation, document, activités documentaires.

Abstract

This paper deals with Web document annotation. Our works aim to harness the annotation concept to unify document-related activities, which are currently partitioned and poorly exploited. Thus, users' interactions and productions are exploited to improve each user's activities. Our propositions, which are implemented in the TafAnnote annotation system, are currently under experimentation.

Keywords

Annotation, digital document, document-related activities.

1 Introduction et motivations

Depuis sa création en 1990, le Web est progressivement devenu une source d'information privilégiée. Dans ce contexte, les activités documentaires couramment réalisées sur les documents papier sont transposées sur leurs homologues électroniques. Ainsi, une kyrielle de logiciels permet de mener à bien chacune des six activités du cycle de vie du document représenté par la figure 1. En effet, une vaste logithèque permet rechercher des documents ① à des fins de production ②③, de diffusion ④, d'exploitation ⑤ et de classement pour une future utilisation ⑥.

L'étude de ces systèmes [3] révèle leur grande spécialisation : un système ne répond généralement qu'à une seule, voire à deux activités. Ce cloisonnement des activités entraîne plusieurs limites, du point de vue de l'utilisateur

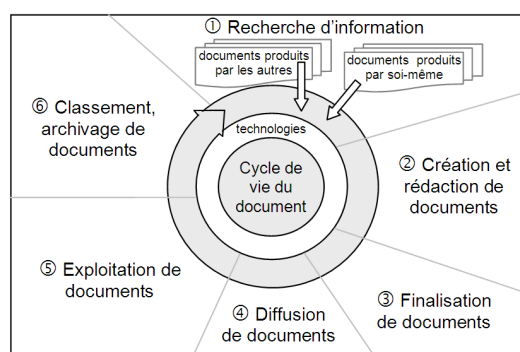


FIG. 1 – Le cycle de vie du document [15, p. 203].

comme du système. D'une part, l'utilisateur doit maîtriser de nombreux systèmes avec lesquels il jongle pour réaliser ses tâches habituelles (qui sont naturellement entrelacées, non cloisonnées). Un scénario courant consiste à chercher avec un navigateur Web, partager ses trouvailles par courriel, questionner des internautes sur des forums et conserver les documents que l'on juge pertinents grâce à des signets. D'autre part, chaque système ne peut construire qu'une représentation partielle des besoins de l'utilisateur, ce qui ne permet pas de répondre correctement à ses besoins ni a fortiori d'anticiper ses demandes.

Nos travaux visent principalement à décroisonner les différentes activités documentaires en les unifiant. Le but de cette approche originale est double : simplifier l'accès et l'appropriation des documents tout en acquérant des connaissances sur l'utilisateur afin d'anticiper ses besoins de manière non intrusive. Notre proposition repose sur une pratique qui couvre et unifie l'ensemble des six activités documentaires : l'annotation (définie dans la section 2) que nous dérivons en annotation collective. Nous détaillons dans la section 3 l'architecture et les fonctionnalités d'un système d'annotation basé sur les annotations collectives qui répond à la problématique posée. La faisabilité de nos propositions est montrée par l'intermédiaire du système TafAnnote¹ (section 4). Ces propositions sont en cours d'expérimentation (section 5). Enfin, la section 6 évoque les principales perspectives de travaux.

*Doctorant en deuxième année de thèse dirigée par Claude Chrisment et co-encadrée par Max Chevalier et Christine Julien.

¹cf. <http://www.irit.fr/~Guillaume.Cabanac/TafAnnote>.

2 La pratique d'annotation

L'essor du Web a démocratisé la publication rapide et peu coûteuse de documents électroniques et les pratiques répandues sur le papier y ont progressivement été transposées [6]. Par exemple, un utilisateur peut ranger un document dans un répertoire et éventuellement l'annoter. Cette section présente la pratique d'annotation de documents électroniques en détaillant ses formes et fonctions.

2.1 Formes et fonctions des annotations


Sur le papier, les lecteurs utilisent une grande variété d'outils pour créer des annotations aux formes diverses². Ainsi, [12] observe une grande expressivité grâce à l'utilisation de crayons à papier, de stylos et feutres de couleur couplée au soulignement, à l'encadrement, au surlignage, au ratissage, au gribouillage, au tracé de flèches, d'astérisques, etc. Ces multiples formes sont utilisées pour répondre à des besoins observables à deux niveaux : pour un usage personnel ou collectif. D'une part, les individus l'utilisent principalement à des fins de lecture active (c.-à-d. pour mener une réflexion critique tout en lisant, en opposition avec la lecture de loisir) et de mémorisation. D'autre part, annoter permet au lecteur qui travaille sur des documents partagés de formuler des remarques et des corrections. Enfin, l'apport des annotations est important aux yeux des lecteurs qui les perçoivent comme une valeur ajoutée apportant un éclairage complémentaire au document original.

L'annotation est une pratique très courante sur le papier, notamment parce qu'elle est facile à réaliser. À l'opposé, les lecteurs sont frustrés de ne pas pouvoir les annoter les documents électroniques [15, p. 95]. Pour faire face à ce besoin, des logiciels appelés systèmes d'annotation (SA) ont été proposés dès l'apparition du Web en 1990. Ils visent à transposer l'annotation papier sur le support électronique tout en exploitant les capacités des ordinateurs modernes.

2.2 Les systèmes d'annotation

Le système Annotea [11] promu par le W3C est très représentatif des vingt systèmes d'annotation que nous avons étudiés en 2005 [2]. Il permet d'annoter tout ou partie d'un document électronique au format HTML. Comme les documents du Web ne sont accessibles qu'en lecture, les annotations ne peuvent pas y être intégrées directement. C'est pourquoi elles sont stockées sur un serveur dédié accessible par le réseau. Afin de pouvoir repositionner l'annotation en contexte dans le document, le système utilise son « point d'ancrage » déduit lors de sa création. Dans Annotea, ceci est réalisé grâce à une expression XPointer³ qui reflète le chemin parcouru de la racine au passage annoté dans la structure logique du document.

En termes de structure, le contenu d'une annotation comprend principalement un commentaire subjectif formulé par son créateur. Ce dernier associe une visibilité à son

annotation : elle peut être privée ou bien visible par les autres utilisateurs du système. Par la suite, les annotations associées à un document visité sont représentées avec une visualisation classique qui reprend la métaphore papier : elles sont affichées en contexte, au sein du document. Dans Annotea, elles sont repérables grâce à l'icône caractéristique  qui représente leur ancrage. Le lecteur peut alors consulter les commentaires de ses prédécesseurs et annoter à son tour le document comme les annotations elles-mêmes. Dans ce cas, cela forme un « fil de discussion » structurant les points de vue et arguments échangés par les annotateurs en contexte.

Malgré un besoin clairement exprimé par les lecteurs, l'existence de nombreux systèmes d'annotation et la création récente de services d'annotation gratuits sur le Web⁴, l'annotation n'est toujours pas couramment pratiquée sur les documents électroniques. Nous pensons que ceci est majoritairement dû au fait qu'il existe trop de systèmes peu ergonomiques et non interopérables pour une même activité : l'annotation. Par ailleurs, les systèmes permettent d'annoter sans pour autant exploiter pleinement les capacités des ordinateurs. En effet, en dehors de la visualisation en contexte, les annotations électroniques demeurent un capital en sommeil. C'est pourquoi la section 3 synthétise nos contributions qui visent à tirer partie des capacités des ordinateurs pour en faire bénéficier les annotateurs.

3 Un système d'annotation collective

Nos travaux proposent d'enrichir la pratique d'annotation électronique mise en œuvre sur le Web au travers de systèmes d'annotation. L'originalité de notre approche est double. Premièrement, nous proposons d'unifier les activités du cycle de vie du document (figure 1) au travers d'un concept unique : l'annotation collective. Actuellement, ces activités sont cloisonnées et requièrent la maîtrise de systèmes distincts qui ne possèdent par conséquent qu'une vue parcellaire des besoins de l'utilisateur. Deuxièmement, nous considérons les interactions entre les différents annotateurs (pouvant appartenir à différents groupes, communautés ou organisations) contrairement aux approches actuelles où l'utilisateur est isolé. De ce fait, les expériences et connaissances collectives sont exploitées pour aider chaque utilisateur. En retour, l'activité d'un utilisateur aidé produit une valeur ajoutée utilisée au bénéfice du collectif, selon un modèle donnant-donnant. Pour ce faire, nous proposons le modèle conceptuel d'un SA collectif détaillé dans la section 3.1. Les différents processus intégrés qui tirent parti de ce modèle pour améliorer les activités documentaires sont détaillés dans la section 3.2.

3.1 Modèle conceptuel d'annotation

Notre approche nécessite de modéliser des données relatives à trois niveaux de granularité. Le premier niveau correspond à l'annotation, employée pour un usage personnel ou collectif. Le second niveau modélise l'utilisateur qui

²cf. le panel d'annotations recensé par la bibliothèque universitaire de Cambridge : <http://www.lib.cam.ac.uk/marginalia>.

³cf. la recommandation du W3C : <http://www.w3.org/TR/xptr>.

⁴cf. <http://stickis.com> et <http://diigo.com> par exemple.

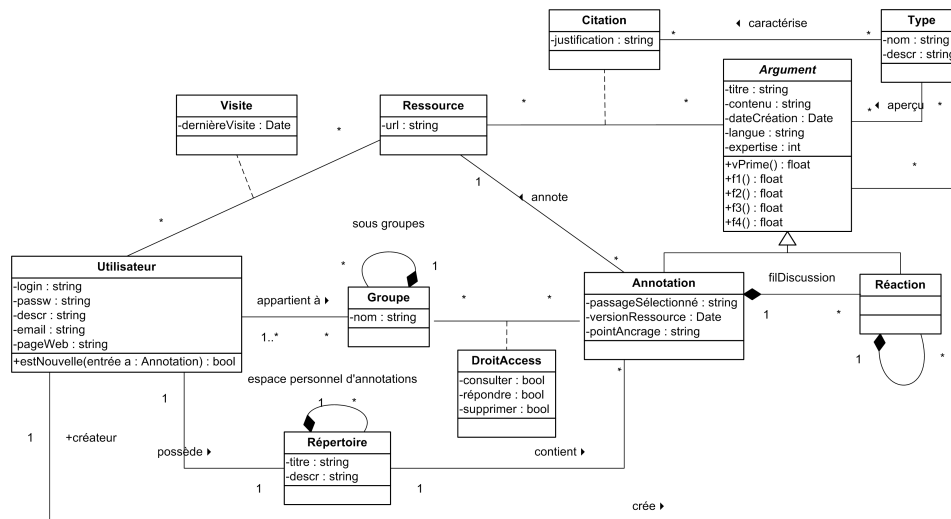


FIG. 2 – Modèle conceptuel du système d’annotation collective [4]

créée et gère de telles annotations. Enfin, le troisième niveau prend en compte l’ensemble des utilisateurs pour mutualiser les résultats de leurs activités afin que le système apporte une aide à chacun d’entre-eux. Ces trois niveaux sont représentés dans un modèle unique illustré par le diagramme de classes UML de la figure 2.

Le premier niveau modélise les annotations formulées par les individus. Afin de notamment permettre l’exploitation collective de documents ⑤, une annotation peut susciter des réactions au sein d’un fil de discussion. En termes de modèle, une annotation comprend à la fois des données objectives *DO* ainsi que des informations subjectives *IS*. D’une part, le système crée des *DO* pour mémoriser les attributs de l’annotation : son **identifiant**, 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 ainsi que son **point d’ancrage** qui spécifie de manière non ambiguë son emplacement au sein de la ressource annotée. 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, groupes ou publique) qui permet de spécifier la portée de l’annotation, cette donnée renseigne le système quant à la diffusabilité de l’annotation. Nous proposons également 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 [12]. Pour étayer ses remarques, l’annotateur peut également spécifier les **citations** 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.

Le deuxième niveau de modélisation décrit l’utilisateur et son espace documentaire. Une des fonctions de l’annotation est l’appropriation de document. C’est pourquoi chaque utilisateur est doté d’un espace personnel d’annotation (EPA) dans lequel il peut organiser ⑥ les annotations pointant vers les documents qu’il désire mémoriser. Afin de pouvoir gérer un grand nombre d’annotations, celles-ci peuvent être organisées hiérarchiquement, c.-à-d. regroupées en répertoires (les utilisateurs sont familiers avec ce type d’organisation utilisé par les systèmes de gestion de fichiers et de signets par exemple).

Enfin, le troisième niveau prend en compte l’ensemble des utilisateurs et les liens qui existent entre eux. Ainsi, les utilisateurs appartenant à un même groupe peuvent spécifier des droits d’accès sur les annotations de leur EPA. Cette fonctionnalité est particulièrement utile dans le cadre du travail collaboratif lorsque les utilisateurs désirent que leurs annotations ne soient consultables que par les membres d’un groupe de travail spécifique.

3.2 Intégration des activités documentaires

Cette section détaille les différents processus mis en place pour exploiter les annotations collectives tout en unifiant les activités documentaires. Ces processus reposent sur une proposition forte : la « validation sociale » d’une annotation. En effet, interpréter et synthétiser mentalement les différentes annotations d’un fil de discussion requiert un gros effort cognitif. Pour limiter cet effort, nous avons tout d’abord introduit des types d’annotation (réfutation, confirmation, question, etc.) qui apportent une information sur la sémantique et le but de chaque annotation. Ces types sont utiles pour connaître le point de vue d’un participant à la discussion sans pour autant avoir à lire le contenu de son annotation. Malgré cette aide, synthétiser l’opinion globale des annotations d’un fil de discussion faisant intervenir de nombreux niveaux de question-réponse reste une tâche difficile. C’est pourquoi nous proposons de calculer la valida-

tion sociale d'une annotation, c.-à-d. le degré de consensus (exprimé entre « réfutation » et « confirmation ») que suscite une annotation, au regard de ses différentes réponses. À cet effet, nous avons proposé un premier algorithme statistique [2] puis un second fondé sur la théorie de l'argumentation en Intelligence Artificielle [4].

① **Recherche d'information.** Un individu qui recherche de l'information s'appuie sur deux modalités : l'interrogation par requête et la navigation dans l'espace des résultats. En phase d'interrogation, le système prend en compte les annotations de deux façons [5]. Il peut enrichir la liste des documents retrouvés (p. ex. grâce à un moteur de recherche) avec les annotations validées socialement et dont le contenu correspond à la requête de l'utilisateur. Ces annotations jugées pertinentes peuvent être également considérées pour proposer les documents associés, soit par le point d'ancrage, soit par l'intermédiaire des références contenues dans le commentaire de l'annotation.

En phase de navigation hypertextuelle, lorsque le lecteur consulte un document d , le système propose les documents proches de d par l'usage. Mesurer la « similarité d'usage » entre les documents est une approche originale qui exploite le travail de chaque individu de façon non intrusive, pour en faire profiter la collectivité. Elle valorise l'effort cognitif réalisé par chaque utilisateur alors qu'il choisit le répertoire le plus pertinent pour un document à classer [13]. Ainsi, les documents conjointement utilisés se retrouvent dans des répertoires « proches » dans les EPA. Par exemple, un lien d'usage est déduit entre d_1 et d_2 lorsque de nombreux utilisateurs classent ces documents de façon proche voire dans le même répertoire.

② **Création et ③ finalisation de documents.** En phase de rédaction d'un document, les annotations en contexte aident les rédacteurs à identifier les passages à corriger, à reformuler, à compléter. . . La validation sociale permet notamment d'adapter la restitution des documents selon les préférences de l'utilisateur qui peut se focaliser sur les propositions qui ont atteint un consensus ou, au contraire, sur celles qui font toujours l'objet de discussion.

④ **Diffusion de documents.** Au sein d'une communauté ou d'une organisation, les documents sont de véritables vecteurs d'information. Afin de réaliser ses tâches, chaque individu a besoin de documents qu'il peut rechercher, classer, transformer et publier. Hormis le fait que ces documents sont pertinents pour cet individu au temps t , ils peuvent également l'être pour d'autres utilisateurs (au même temps t ou dans le futur). Pourtant, [10] montre que les documents sont peu partagés, souvent cherchés plusieurs fois et même recréés. . .

Dans notre approche, chaque utilisateur gère les documents qui l'intéressent dans son EPA. Par ce biais, le système connaît les centres d'intérêts de chaque utilisateur, il peut par conséquent anticiper leurs besoins. Par ailleurs, les annotations collectives sont associées à des documents qui ont suscité l'intérêt d'au moins un utilisateur. Sur cette base, le

système met en œuvre le processus de recommandation détaillé dans [3]. De façon synthétique, un nouveau document est recommandé en l'insérant dans le répertoire le plus pertinent de l'EPA associé. L'utilisateur bénéficiaire peut ensuite consulter ce document, le réorganiser dans son EPA et le recommander à son tour ou bien le rejeter.

⑤ **Exploitation de documents.** Les fonctionnalités détaillées pour les activités ② et ③ sont également adaptées pour l'exploitation des documents. En effet, l'annotation en contexte associée aux fils de discussion argumentatifs et à la validation sociale permet aux lecteurs de s'approprier les documents, de les critiquer, de ne plus être restreint au seul point de vue du rédacteur, etc. Par ailleurs, le système mémorise les documents visités afin de pouvoir indiquer les nouvelles annotations apportées à un document : celles ayant été formulées après la dernière visite de l'utilisateur. Cette fonctionnalité originale répond à une problématique liée au passage à l'échelle des SA : elle permet au système de masquer les anciennes annotations tout en mettant en avant les nouvelles, ce qui réduit l'encombrement visuel caractéristique des documents massivement annotés [5].

⑥ **Gestion de documents.** Dans notre approche, chaque utilisateur possède son EPA où il classe les annotations qu'il crée. À tout moment, il peut réorganiser tout ou partie de son EPA manuellement grâce au processus de réorganisation décrit dans [3]. Ce dernier repose sur un algorithme de classification ascendante hiérarchique couplé à une technique d'étiquetage des répertoires obtenus. Le résultat du processus peut être accepté en totalité ou en partie.

4 Le prototype TafAnnote

Le système d'annotation TafAnnote est un prototype de recherche destiné à montrer la faisabilité de nos propositions. De ce fait, ses fonctionnalités sont en constant développement. L'architecture du système est de type client-serveur : c'est une barre d'outils intégrée au navigateur Web Mozilla Firefox et associée à un serveur d'annotation dédié [1].

Actuellement, TafAnnote permet d'annoter les documents au format HTML, les points d'ancrage sont exprimés par des expressions XPointer. Le système affiche les annotations en contexte en insérant une icône représentant leur type dans le document visualisé. Des informations connexes renseignent l'utilisateur sur l'expertise de l'annotation et sur le nombre de réponses suscitées. De plus, le calcul de la validation sociale permet d'adapter la restitution selon les préférences de l'utilisateur (p. ex. focus sur les discussions en cours). Enfin, l'utilisateur peut rechercher des annotations en formulant des requêtes et gérer son EPA.

5 Expérimentations

À l'heure actuelle, nous avons défini deux expérimentations. Dans un premier temps, nous avons confronté la mesure de similarité classique basée sur le contenu (mesure du cosinus sur le modèle vectoriel [14]) avec notre proposition de **mesure basée sur l'usage**. Pour ce faire, nous

avons utilisé le corpus de [9] qui comprend 4 079 documents (organisés en 486 répertoires) appartenant à 14 personnes. Des résultats préliminaires montrent que la mesure basée sur l'usage fait apparaître des liens de similarité entre des documents qui ne sont pas similaires sur le contenu. La pertinence de ces liens est actuellement à l'étude.

D'autre part, nous cherchons à vérifier que la **validation sociale** calculée par nos algorithmes [2, 4] correspond réellement à la perception humaine. Pour ce faire, nous avons constitué un corpus de 13 fils de discussion (soit 222 annotations) soumis à l'évaluation humaine via un logiciel d'expérimentation en ligne⁵. Chaque participant à l'expérimentation doit *i*) associer les types du modèle (confirme, réfute, exemple, etc.) aux annotations contenues dans les fils de discussion et *ii*) évaluer mentalement la validation sociale. Cette expérimentation étant actuellement en cours, nous prévoyons d'analyser les données recueillies afin de répondre aux questions suivantes : les individus sont-ils d'accord sur le typage des annotations ? Déduisent-ils la même validité sociale en synthétisant les opinions exprimées dans les annotations ? Arrivent-ils au même résultat que nos algorithmes de validation sociale ?

6 Conclusion et perspectives

Cet article synthétise nos travaux relatifs à l'annotation de documents électroniques. La problématique posée fait référence aux systèmes actuels qui *i*) n'exploitent pas les interactions entre utilisateurs et le résultat de leurs activités et *ii*) sont spécifiques et cloisonnés à une activité documentaire. Notre proposition vise à unifier et intégrer les activités documentaires autour du concept d'annotation ; elle est fondée sur un modèle d'annotation collective et sur des processus identifiés. Le système TafAnnote montre la faisabilité de notre approche ; actuellement deux expérimentations nécessitant des participants sont en cours.

En termes de perspectives, après avoir recensé 70 systèmes d'annotation et assimilables (signets, *tags*, etc.), nous cherchons une corrélation entre les fonctionnalités proposées et les activités documentaires considérées. Cela nous permettra de raffiner le modèle proposé. D'autre part, nous envisageons de définir une API d'annotation afin que TafAnnote puisse être utilisé comme un service par des applications dont l'annotation n'est pas l'objectif principal. Ainsi, tout logiciel (éditeur de texte, lecteur vidéo, etc.) pourrait proposer des services d'annotation de façon intégrée et interoperable. Ces logiciels pourraient eux-même gérer la création et la restitution des points d'ancrage dans leurs propres formats. Actuellement, c'est un challenge qui nous a mené à définir une technique d'ancrage [7, 8] spécifique aux tables multidimensionnelles, qui sont des documents électroniques manipulés par des analystes dans le contexte applicatif des systèmes décisionnels.

Références

[1] G. Cabanac. Annotation de ressources électroniques sur le Web : formes et usages. Rapport de Master 2 Recherche 2IH, IRIT, Université Toulouse 3, France, June 2005.

- [2] G. Cabanac, M. Chevalier, C. Chrisment, and C. Julien. A Social Validation of Collaborative Annotations on Digital Documents. In *International Workshop on Annotation for Collaboration*, pages 31–40, Paris, Nov. 2005. Programme société de l'information, CNRS.
- [3] G. Cabanac, M. Chevalier, C. Chrisment, and C. Julien. L'architecture CoMED pour la gestion collective de documents électroniques dans l'organisation. In K. Zreik and C. Vanoirbeek, editors, *CIDE.9*, pages 237–252, Fribourg, Suisse, Sept. 2006. Europia.
- [4] G. Cabanac, M. Chevalier, C. Chrisment, and C. Julien. Validation sociale d'annotations collectives : argumentation bipolaire graduelle pour la théorie sociale de l'information. In *INFORSID'06 : 24^e congrès de l'INformatique des Organisations et Systèmes d'Information et de Décision*, pages 467–482. Éditions Inforsid, June 2006.
- [5] G. Cabanac, M. Chevalier, C. Chrisment, and C. Julien. Collective Annotation : Perspectives for Information Retrieval Improvement. In *RIAO'2007 : Proceedings of the 8th conference on information retrieval – Large-Scale Semantic Access to Content*. CID, May 2007.
- [6] G. Cabanac, M. Chevalier, C. Chrisment, C. Julien, C. Soulé-Dupuy, and P. Tchienehom. Web Information Retrieval : Towards Social Information Search Assistants. In I. Chen and T. Kidd, editors, *Social Information Technology : Connecting Society and Cultural Issues*, chapter à paraître. IDEA, Mar. 2008.
- [7] G. Cabanac, M. Chevalier, F. Ravat, and O. Teste. Modèle conceptuel pour bases de données multidimensionnelles annotées. In G. Ritschard and C. Djeraba, editors, *EGC*, volume RNTI-E-6 of *Revue des Nouvelles Technologies de l'Information*, pages 119–124. Cépaduès, Jan. 2006.
- [8] G. Cabanac, M. Chevalier, F. Ravat, and O. Teste. Métamodélisation des bases de données multidimensionnelles annotées. In *Revue des Nouvelles Technologies de l'Information (RNTI-B-2) - Entrepôts de Données et Analyse en ligne (EDA'06)*, pages 39–54. Cépaduès, June 2006.
- [9] M. Chevalier, C. Chrisment, and C. Julien. Helping People Searching the Web : Towards an Adaptive and a Social System. In *ICWI 2004 : Proceedings of the 3rd International Conference WWW/Internet*, pages 405–412. IADIS, 2004.
- [10] S. Feldman. The high cost of not finding information. *KM World magazine*, 13(3), Mar. 2004.
- [11] J. Kahan, M.-R. Koivunen, E. Prud'Hommeaux, and R. R. Swick. Annotea : an open RDF infrastructure for shared Web annotations. *Comp. Netw.*, 32(5) :589–608, Aug. 2002.
- [12] C. C. Marshall. Toward an ecology of hypertext annotation. In *HYPertext '98 : Proceedings of the 9th ACM conference on Hypertext and hypermedia : links, objects, time and space—structure in hypermedia systems*, pages 40–49, New York, NY, USA, 1998. ACM Press.
- [13] J. Rucker and M. J. Polanco. Siteeater : personalized navigation for the web. *Commun. ACM*, 40(3) :73–76, 1997.
- [14] G. Salton, A. Wong, and C. S. Yang. A Vector Space Model for Automatic Indexing. *Commun. ACM*, 18(11) :613–620, Nov. 1975.
- [15] A. J. Sellen and R. H. Harper. *The Myth of the Paperless Office*. MIT Press, Cambridge, USA, 2003.

⁵cf. <http://www.irit.fr/~Guillaume.Cabanac/expe>.