

# Vers un système d'annotation distribué

**E. DESMONTILS, C. JACQUIN, L. SIMON**

Institut de Recherche en Informatique de Nantes  
2, rue de la Houssinière  
B.P. 92208  
44322 NANTES CEDEX 3

— *Ingénierie des Connaissances* —



**RAPPORT DE RECHERCHE**

**N° 03.01**

**Février 2003**

E. DESMONTILS, C. JACQUIN, L. SIMON  
*Vers un système d'annotation distribué*  
16 p.

Les rapports de recherche de l'Institut de Recherche en Informatique de Nantes sont disponibles aux formats PostScript® et PDF® à l'URL :

<http://www.sciences.univ-nantes.fr/irin/Vie/RR/>

*Research reports from the Institut de Recherche en Informatique de Nantes are available in PostScript® and PDF® formats at the URL:*

<http://www.sciences.univ-nantes.fr/irin/Vie/RR/indexGB.html>

© Avril 2003 by E. DESMONTILS, C. JACQUIN, L. SIMON

# Vers un système d'annotation distribué

E. DESMONTILS, C. JACQUIN, L. SIMON

## Résumé

De nombreux systèmes de partage d'information existent de nos jours mais les spécificités du Web en font des outils extrêmement difficiles à exploiter. Les outils d'annotation visent à améliorer échange, communication et interopérabilité sur le Web. L'objectif de cet article est d'une part, de faire une synthèse des caractéristiques des annotations et des architectures des systèmes d'annotation et d'autre part, de proposer une nouvelle architecture pour un système d'annotation simple d'utilisation, léger, efficace, non-intrusif, évolutif, partagé et indépendant d'une plate-forme.

Catégories et descripteurs de sujets : H.5.3 [**Information Systems**]: Group and Organization Interfaces

Termes généraux : Collaborative computing, Computer-supported cooperative work, Web-based interaction, Synchronous/Asynchronous interaction

Mots-clés additionnels et phrases : Annotation, notation, WWW, Internet, Proxy distribué



# 1 Introduction

De nombreux systèmes de partage d'information existent : cela va du Web aux outils très évolués de travail collaboratif (Lotus Notes...). Cependant, ce qui est possible avec ces derniers ne l'est plus avec le Web. Ces outils visent un groupe très réduit de personnes travaillant souvent ensemble, avec un vocabulaire commun, sur des thèmes proches et donc avec des habitudes spécifiques. Par contre, le Web concerne potentiellement des millions de personnes non seulement ne se connaissant pas mais ayant en plus des centres d'intérêt différents, des habitudes différentes, des cultures différentes... De plus, sur le Web, l'information est fortement distribuée, extrêmement volumineuse, évolutive, volatile, très "bruitée", très hétérogène et souvent très peu structurée. Dans ce contexte, il est nécessaire de proposer des méthodes et des outils pour comprendre, manipuler et partager des documents, pour mettre en place des services pertinents et performants.

Dans ce but sont nés les outils d'annotation qui visent à améliorer l'appréhension des documents HTML ainsi que la communication et l'interopérabilité sur le Web. Deux types de systèmes d'annotation préexistent, l'un repose sur des annotations sémantiques et l'autre sur des annotations libres. Les annotations sémantiques sont des méta-données basées sur des ontologies et ajoutées au document.

Les annotations libres quant à elles, permettent d'associer des notes de lectures aux documents, de partager de l'information, d'effectuer des tâches rédactionnelles en groupe... Grâce aux systèmes d'annotation, le lecteur devient aussi rédacteur. Plus généralement, dans un contexte comme le Web, le système passe du "one-to-many" (un rédacteur et des millions de lecteurs) au "many-to-many" (tout utilisateur du Web est Lecteur/Rédacteur) [21].

Dans cet article, notre objectif est d'une part, de faire une synthèse des caractéristiques des annotations (section 2) et de décrire les différents systèmes d'annotation proposés sur le Web (section 3), et d'autre part, de proposer un nouveau type d'architecture pour les systèmes d'annotation qui pallient aux problèmes rencontrés avec les systèmes existants (sections 4 et 5).

## 2 Caractérisation des annotations

### 2.1 Description des annotations

Une annotation est une information graphique ou textuelle attachée à un document et le plus souvent placée dans ce document. Cette place est donnée par une ancre.

Les annotations font référence à des entités diverses : un ensemble de documents, un document, un passage, une phrase, un terme, un mot, une image...

Les annotations peuvent prendre plusieurs formes comme :

- des icônes (pour décrire des avis en utilisant des étoiles, des points d'interrogation...),
- des symboles de liens (pour décrire des associations, des relations entre mots...),
- des notes textuelles en marge, en bas de page ou en fin de document repérées dans le texte<sup>1</sup> par des icônes (numéros, étoiles...),
- des mises en formes typographiques (surlignage, soulignage, italique...),
- des redécoupages de texte (à l'aide d'accolades, de numérotation de passages...),
- des images, des sons,
- des concepts et leurs attributs (annotations sémantiques)...

### 2.2 Dimensions des annotations

C. C. Marshall [11] propose de caractériser les annotations selon différentes dimensions. Ces dimensions sont des espaces mono-dimensionnels continus. Elles décrivent les propriétés de l'annotation au niveau de sa structuration, de sa fonction et de son rôle dans la communication rédacteur/lecteur. Dans cette section, nous allons décrire ces dimensions en spécifiant les cas typiques, extrêmes, mais il est souvent possible de trouver des exemples d'annotations "intermédiaires". Nous caractérisons aussi les annotations sémantiques en les projetant sur ces différentes dimensions.

---

<sup>1</sup>Ce sont les ancres (comme pour cette note !).

### 2.2.1 Dimensions liées à la structuration de l'annotation

Une première classe de dimensions concerne la structuration de l'annotation. La première dimension proposée décrit le niveau de formalisation de l'annotation : c'est la dimension Formelle - Informelle. Les annotations sont représentées de manière plus ou moins structurée, i.e. plus ou moins basée sur une formalisation rigoureuse. Les notes peuvent aller du renseignement de champs spécifiques (par exemple par réponse à un questionnaire) au texte en langage naturel (mais aussi utilisation de sigles intuitifs...). Les annotations sémantiques, quant à elles, sont par nature formelles (reposent sur une connaissance définie a priori (une ontologie) qui est représentée à l'aide d'un langage spécifique)

Une autre dimension concerne la signification propre de la note, c'est-à-dire entre l'explicite et l'implicite. Une annotation explicite se suffit à elle-même (destinée à une autre personne que le rédacteur ne connaît pas forcément) alors qu'une annotation implicite demande une connaissance complémentaire (table de lecture...) et est destinée à un lecteur instruit des conventions adoptées (souvent le rédacteur lui-même). Ces dernières sont difficiles à interpréter, parfois même, après un certain temps, par son auteur lui-même ! Par exemple, les notes de bas de page sont le plus souvent des annotations explicites alors que les annotations de lecture sont assez souvent une codification propre au lecteur (même si C. C. Marshall a mis en évidence la possibilité de trouver des traits communs dans les comportements d'annotation). Les annotations sémantiques, quant à elles, étant donné qu'elles font référence à une ontologie sont donc implicites.

Il faut noter que certaines annotations peuvent "se déplacer" sur les axes des dimensions au cours du temps. Par exemple, les smiley ("frimousse" en français !) étaient au départ des annotations informelles et sont devenues (à travers tous les livres sur Internet) des annotations formelles (des règles de construction "normalisées" sont proposées...).

### 2.2.2 Dimensions concernant les fonctions de l'annotation

Une classe de dimensions concerne les dimensions décrivant les fonctions des annotations. La première dimension concerne l'utilisation des annotations dans les processus de lecture et de rédaction. Pour aller plus loin que C. Marshall, nous pensons que les annotations ont cinq utilisations majeures :

1. *information, illustration, extension du document* : lorsque le lecteur rédige ses annotations, il devient alors rédacteur<sup>2</sup>,
2. *forum* : en permettant à un ensemble de lecteurs de débattre sur le document,
3. *opérationnalisation de l'information* : les annotations sémantiques permettent une opérationnalisation de l'information contenue dans des documents. En effet, elles sont destinées à être traitées par des machines (par opposition aux annotations libres en langage naturel ou composées de symboles souvent implicites). Leur objectif majeur est de *désambiguïser* le document pour un traitement automatique.
4. *aide au processus rédactionnel* : en permettant d'indiquer des consignes de rédaction (corrections, mouvements d'informations...),
5. *support de lecture* : (la mise en évidence de passages importants...) permettant l'appropriation du texte par le lecteur, l'annotation est alors le reflet de l'engagement par rapport à un texte d'un lecteur qui le personnalise (trace de lecture) afin de faciliter un futur retour.

C. C. Marshall propose aussi une dimension concernant le rôle de l'annotation par rapport au niveau de lecture du document. Certaines notes (hyper-liens par exemple) permettent une lecture extrêmement superficielle du document (surf). À l'opposé, des annotations sont utilisées pour une lecture approfondie d'un texte<sup>3</sup>. Entre les deux, il y a les notes de lecture rapide comme l'utilisation de la typographie (mise en gras ou en italique par exemple) ou le surlignage.

Remarque : D'une certaine manière, les attributs typographiques peuvent être considérés comme des annotations permettant une aide à la lecture, particulièrement le souligné, le surligné, le gras et l'italique.

C. C. Marshall propose aussi une dimension qui concerne la "durée de vie" de l'annotation, c'est la dimension temporelle. En effet, une annotation peut être permanente ou éphémère. Les annotations permanentes sont des

<sup>2</sup>Ce type d'annotation est présente dans beaucoup de systèmes d'annotation sur le Web. C'est une plus-value importante sur le document.

<sup>3</sup>Cf. rôle des notes dans les œuvres de Tolkien par exemple.

annotations faites pour durer. Dans ce cadre, nous trouverons par exemple toutes les annotations de bas de page dans les documents, les annotations sémantiques... Ce sont généralement les annotations produites par le rédacteur et laissées dans le document final. À l’opposé, les annotations éphémères sont très souvent des annotations d’aide à la rédaction du document. Une spécificité des annotations de documents numériques est que l’annotation n’est pas statique. En effet, sa forme et son contenu peuvent évoluer au cours du temps. L’annotation elle-même peut être placée pour une durée de vie limitée.

### 2.2.3 Dimensions concernant le rôle de l’annotation dans la communication rédacteur/lecteur

Une dimension particulièrement importante concerne la relation lecteur/rédacteur, autrement dit sur le choix des destinataires potentiels de l’annotation. Une annotation peut être soit privée, c’est-à-dire que le rédacteur la destine à lui-même, soit publique. Dans ce dernier cas, plusieurs “degrés” sont possibles : le groupe de travail, l’institution... jusqu’au niveau le plus global (c’est-à-dire le monde entier !).

Le comportement du rédacteur et des lecteurs vis à vis du document et des annotations mène à des systèmes de type “one-to-many” (un rédacteur destine son document à un ou plusieurs lecteurs) ou “many-to-many” (un groupe de lecteurs/rédacteurs travaille en collaboration). Les systèmes d’annotation sur le Web visent à atteindre ce dernier cas.

Remarque : Il existe aussi des annotations, en particulier pour le support de lecture, ajoutées automatiquement au document permettant un accès plus facile que ce soit par la machine ou par l’humain. Le système XLibris [18] par exemple propose de mettre en évidence les mots représentatifs du document en utilisant la méthode statistique TFIDF. Ce même système ajoute une note aux termes déjà annotés. Celle-ci contient le résultat de la recherche des mots annotés par des moteurs de recherche classiques (Altavista...). Google surligne automatiquement les termes du document ayant fait l’objet de la recherche. Certains outils gardent aussi une trace des liens empruntés [20], du temps passé sur les pages ou des événements souris. Ces informations seront ensuite proposées à l’utilisateur et éventuellement elles-mêmes annotées [20].

## 2.3 Exploitation des annotations

Les annotations sont multiples dans leurs formes et dans leurs fonctions. Par rapport au Web, les annotations les plus courantes sont des annotations libres en langage naturel (donc informelles) ou symboliques (surlignage en particulier), publiques (par définition du Web !), plutôt éphémères, explicites et le plus souvent associées à des termes. Elles sont le plus souvent utilisées comme support de lecture (commentaires, recherche de passages, surlignage, cartes de navigation) ou comme médium ayant rôle de forum. Les annotations sémantiques, quant à elles, sont utilisées dans le cadre de la recherche d’information, la classification, le résumé automatique et plus généralement, pour favoriser l’interopérabilité [4].

Des travaux récents, comme [2], proposent d’exploiter les annotations libres afin de capitaliser l’apport d’information issue de ces notes. Les annotations d’un document sont en effet exploitables pour améliorer la gestion du document. Une première utilisation des annotations peut être la constitution d’un résumé du document. En effet, les parties annotées peuvent être considérées comme importantes (positivement ou négativement) pour le lecteur. Aussi, construire un texte à partir de ces “morceaux” s’avère une méthode assez performante pour résumer le texte. Au-delà de l’indexation, dans le même ordre d’idée que le résumé automatique, [2] propose une méthode de classification basée sur les termes annotés.

Les annotations libres sont aussi potentiellement intéressantes dans le cadre de la recherche d’information. En effet, les termes annotés peuvent être représentatifs du document. A cela, on peut ajouter l’exploitation du contenu des annotations dans l’indexation ou même au niveau de la recherche d’informations. Dans ce dernier cas, les annotations sont susceptibles de servir au niveau du retour utilisateur (“relevance feedback”) ou pour étendre la requête (“query expansion”). En effet, l’analyse des passages annotés et des annotations permet de juger la pertinence des documents proposés et de construire d’autres requêtes plus adéquates [7].

### 3 Les systèmes d'annotations libres sur le Web

Les outils d'annotation libre doivent prendre en compte un certain nombre de contraintes et particularités du Web : les acteurs (lecteurs et serveurs Web) sont répartis, les communications se font par le réseau, le système est fondamentalement multi-utilisateurs (utilisateurs par ailleurs très nombreux), le langage de communication (HTTP), les données au format HTML ou XML... Globalement, tous les systèmes respectent le même schéma d'architecture (figure 1) [17] : un intermédiaire "observe" les transactions entre le client Web et les serveurs Web. Cet intermédiaire agit sur la requête, les pages obtenues et, éventuellement, sur les événements issus du navigateur. Cet élément est composé d'un intercepteur qui est chargé de récupérer requête et/ou pages HTML, d'un composeur qui se charge d'associer aux pages les annotations attachées (présentes dans une base de données). Cette combinaison peut dépendre du profil de l'utilisateur.

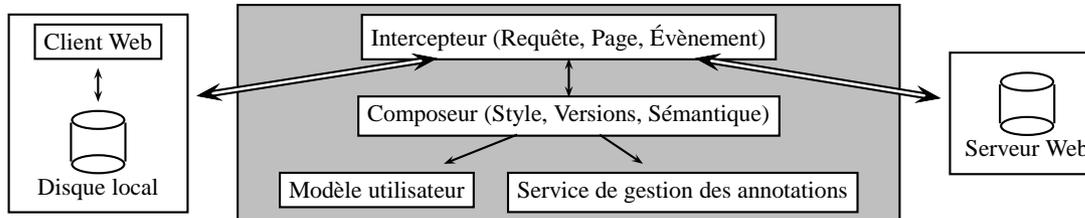


FIG. 1 – Schéma général d'un outil d'annotation sur le Web ([17])

En pratique, les systèmes d'annotations libres sur le Web se divisent en deux grandes catégories : ceux basés sur des serveurs mandataires ("proxy") et ceux utilisant un intermédiaire attaché au navigateur (intermédiaire client).

Les intermédiaires de type serveur mandataire [13, 12, 19] sont des serveurs indépendants du client et des serveurs Web. Dans la configuration la plus standard, le serveur mandataire observe toutes les requêtes du client et gère la page ayant des annotations présentes dans sa base. Avec un tel système, toutes les transactions du client doivent passer par ce serveur posant ainsi un problème de confidentialité. De plus, ces systèmes posent aussi des questions autour du droit d'auteur. En effet, certains sites refusent que leurs pages soient modifiées. Pour éviter ces deux problèmes, [19] (CritLink) propose un système basé sur un proxy facultatif (figure 2). L'utilisateur fait appel au système d'annotation uniquement quand il le désire (confidentialité) et les pages des sites peuvent être consultées sans les annotations (pointillés dans la figure 2)<sup>4</sup>. Ce type de serveur d'annotations pose d'autres problèmes :

- il n'est pas possible d'annoter des documents locaux,
- l'ajout d'annotations passe nécessairement par un dialogue spécifique (parfois assez lourd) et par un rechargement de la page,
- il ne sait pas gérer les pages dynamiques,
- il est lent (goulot d'étranglement) puisque toutes les requêtes doivent passer par lui,
- le document d'origine est forcément modifié (les notes sont insérées dans le document),
- il gère mal les codes Java et Javascript ainsi que les frames.

Les serveurs mandataires ont tout de même un certain nombre d'avantages non négligeables : le partage d'annotations est facile, les outils sont indépendants du système d'exploitation et du navigateur utilisés. De plus, ils sont faciles à installer (et après une session d'annotation, il n'y a rien à désinstaller !).

Les systèmes à base d'intermédiaire client [15, 10, 9, 3, 16] sont des systèmes à installer sur le navigateur (figure 3). Par conséquent, ces systèmes sont totalement dépendants du système d'exploitation et du navigateur utilisés. Dans ces systèmes, le problème du partage des annotations n'est pas aussi simple à gérer. Certains systèmes proposent des protocoles pour les envoyer par courrier électronique ou d'utiliser un serveur d'annotation. Cette dernière solution réintroduit le problème de goulot d'étranglement normalement résolu du fait de la distribution du système d'annotation (en opposition à la centralisation des serveurs mandataires). Dans ces systèmes, la création et la visualisation des annotations sont beaucoup plus souples, quelque soit la structure du document et ce qu'il

<sup>4</sup>Par exemple, si l'on désire accéder à la page "www.www.fr" d'origine, il suffit d'entrer sur son navigateur "http://www.www.fr". Par contre, si l'on désire la page avec les annotations qui y sont attachées, il suffit d'entrer "http://crit.org/http://www.www.fr".

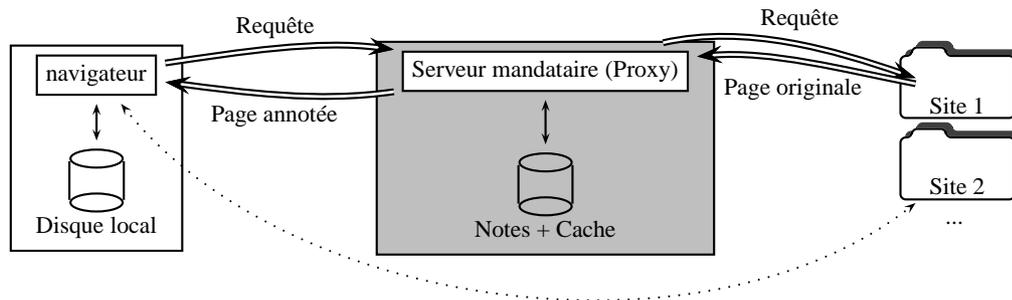


FIG. 2 – Architecture en proxy simple facultatif

contient. Comme il est intégré au navigateur, il a accès à la structure DOM (Document Object Model) du document ainsi qu'aux différents évènements issus du navigateur.

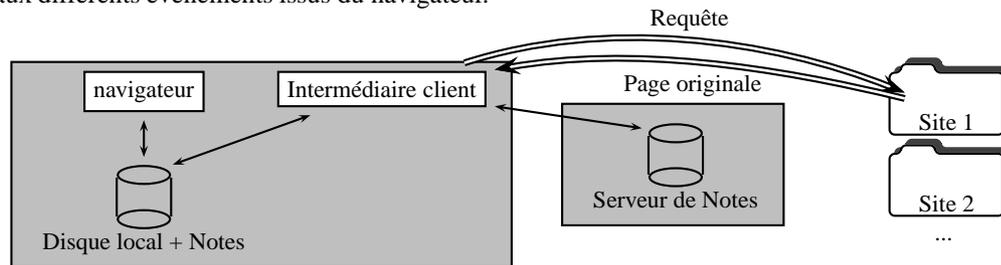


FIG. 3 – Architecture en intermédiaire client

Ces systèmes, et les architectures sous-jacentes, posent un certain nombre de problèmes en dehors des problèmes de confidentialité et de droit d'auteur déjà abordés. Le principal d'entre-eux est le passage à l'échelle ("scalability"). En effet, le problème est d'arriver à gérer des milliers (millions ?) d'utilisateurs et donc de manipuler des millions d'annotations. La solution la plus courante consiste soit à distribuer les serveurs soit à les dupliquer (serveurs miroirs) avec tous les problèmes habituels dans le domaine des bases de données, associés à ces solutions. Un autre problème est l'impossibilité de partager les annotations entre plusieurs systèmes, chacun utilisant son propre formalisme. Les normalisations avenir du W3C devraient résoudre en partie ces problèmes.

Finalement, dans tous les systèmes rencontrés, les annotations sont souvent simplistes et ne tiennent pas compte des objectifs visés, des types d'annotations manipulées (dimensions) et des utilisateurs concernés. Le grand nombre d'échec pour ces systèmes (très peu sont actuellement maintenus ou dépassent l'état de prototype de recherche) montre que ces problèmes sont loin d'être résolus et demandent à être travaillés<sup>5</sup>.

## 4 Une nouvelle architecture d'annotations : le système Dinossys

Les propriétés attendues pour un système d'annotation sur le Web sont : la légèreté, l'efficacité, la transparence, l'indépendance vis à vis de la plate-forme support, le passage à l'échelle [17]. Une autre propriété devient aussi de plus en plus primordiale (par exemple avec l'e-learning), c'est le support du travail collaboratif. En effet, dans ce contexte, des personnes travaillent ensemble, en même temps et le système permet la mise à jour en temps réel des annotations effectuées par chacune d'elles. L'analyse de la section précédente montre que les divers systèmes existants n'ont pas tous les qualités énoncées précédemment. Notre objectif étant de définir une nouvelle architecture respectant les propriétés décrites, nous énonçons ici, nos motivations et nos choix.

1. *Une architecture distribuée* : l'intérêt porté par un utilisateur à une application et très souvent, sans parler de l'intérêt même du logiciel, lié à sa fiabilité et à la «fluidité» des fonctionnalités proposées. En ce qui concerne la fiabilité, elle représente une part importante dans la satisfaction et l'utilisation d'une application.

<sup>5</sup>Pour des comparatifs techniques plus précis (techniques mises en œuvre, présentations des annotations choisies, types d'annotations...), il est possible de consulter[6, 2, 1, 8, 14, 17, 21].

Cette constatation nous a donc motivé dans la mise en place d'un système de proxy distribués géographiquement, spécialisés sur des domaines précis. Cette architecture distribuée apporte une plus grande disponibilité de services en cas d'interruption volontaire ou involontaire d'un proxy et assure une bonne répartition de la charge de travail en permettant la mise en place de nouveaux proxy au gré des sollicitations. Tout ceci garantissant à l'utilisateur final une plus grande fluidité dans l'utilisation de l'application et par conséquent un plus grand intérêt dans son utilisation. la distribution contribue donc à ce que notre système ait les propriétés d'efficacité et de passage à l'échelle.

2. *Support du travail collaboratif* : plusieurs systèmes d'annotations actuels disposent de fonctionnalités de partage d'annotations sous la forme d'envois de mails aux participants. Ces fonctionnalités ne permettent pas des échanges «temps réel» des informations. Dans notre architecture, même si cette fonctionnalité reste présente (notamment pour prévenir le ou les participants non connectés lors d'une session d'annotations collaborative), elle se voit couplée à un système de mise à jour automatique des annotations. Ce dernier sera utilisé lors de sessions d'annotations collaboratives en mode «forum» (i.e. un ensemble de personnes participe en même temps à un échange d'annotations sur un même document) ce qui rend ainsi les échanges plus directs.
3. *Indépendance vis à vis de la plate-forme, transparence et légèreté* : beaucoup de systèmes d'annotations existants sont basés sur des technologies propriétaires. Certains ne sont utilisables que sur une seule plate-forme ou ne fonctionne qu'avec un seul type de navigateur. Pour dépasser ces limitations, s'affranchir des technologies propriétaires et donc ne pas imposer à l'utilisateur un contexte d'utilisation trop différent du sien (i.e. un utilisateur de Netscape sur Linux ne souhaite peut être pas ou n'a pas la possibilité de travailler avec Opéra sur Windows), la partie cliente de l'architecture se compose d'une applet Java. Ce choix technique permet à l'application de fonctionner sur toutes les plateformes dotées d'une machine virtuelle Java telles que Microsoft Windows, Linux, Unix, Macintosh... et avec tous les navigateurs prenant en charge le DOM et le JavaScript comme Internet Explorer, Opéra, Netscape... L'utilisation d'une applet Java pour la partie cliente permet une mise à jour automatique de l'application sur le poste client sans aucune intervention particulière à l'ouverture d'une session d'annotation sur le portail d'accès (il est bien évident que si aucune mise à jour n'est nécessaire, c'est la version en cache sur le poste client qui est utilisée n'engendrant ainsi, aucun téléchargement superflu). Enfin, ce système de mise à jour automatique permet de garantir une utilisation de l'application dans sa dernière version de manière totalement transparente.

## 5 Description d'une session d'annotation à l'aide de Dinosys

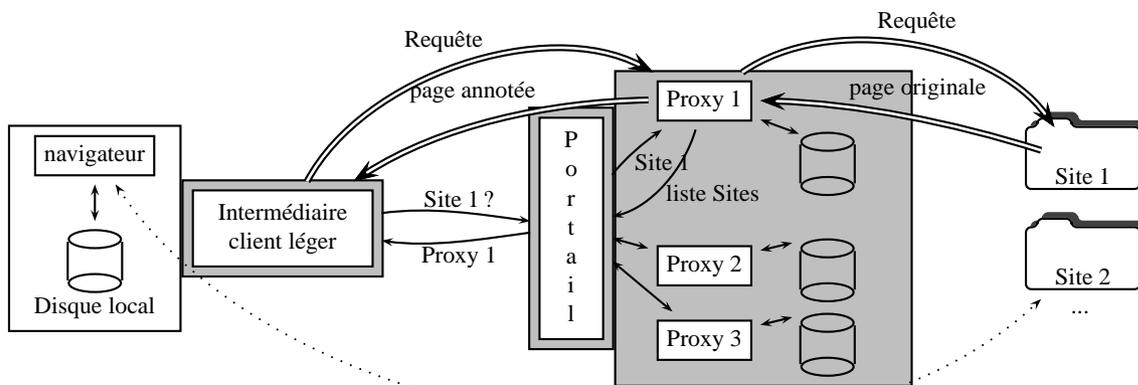


FIG. 4 – Le système Dinosys

Pour utiliser Dinosys (figure 4), il est nécessaire de se rendre à l'adresse du portail via un navigateur internet. Ce portail est chargé d'afficher dans le navigateur l'interface cliente du système (i.e. l'intermédiaire client) qui se présente sous la forme d'une barre d'outils (figure 5). Notons que ce portail se charge également de vérifier la présence ou non de cet intermédiaire sur le poste client et le cas échéant, d'en effectuer un téléchargement

automatique. Ce processus permet donc une installation de l'application pour les nouveaux utilisateurs ainsi qu'un contrôle de version pour les autres totalement transparent. Il est désormais possible à l'utilisateur d'entrer dans la zone d'adresse de la barre d'outils, l'URL du document à annoter et/ou consulter.

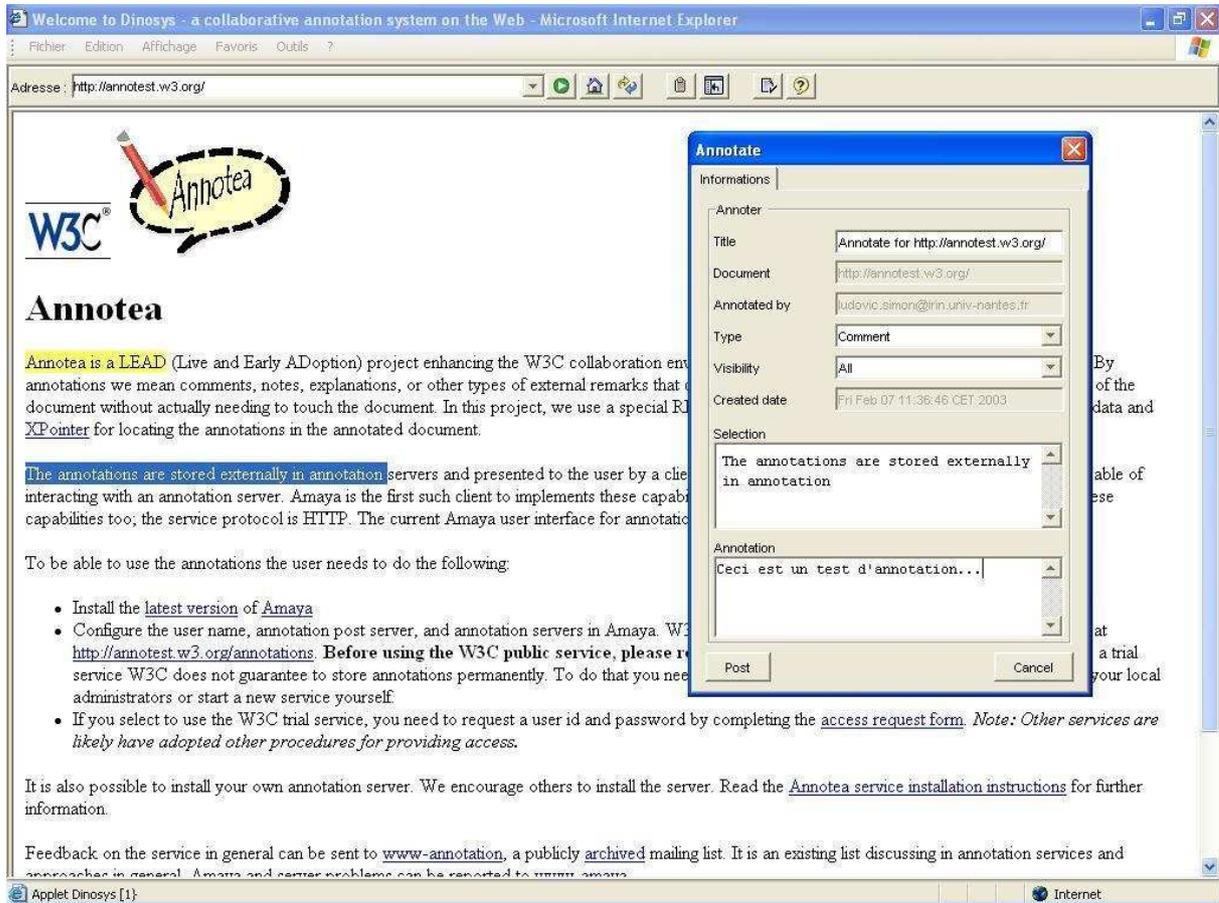


FIG. 5 – Client du système Dinosys

La requête (ici une page du Site No 1) est transmise au portail qui va établir la correspondance entre le site demandé et le proxy qui en a la charge. Une fois la correspondance établie, le portail communique en retour, l'adresse du proxy au client (ici le Proxy No 1). A partir de cet instant, les échanges ne se feront plus qu'entre l'intermédiaire client et le proxy pour ce qui concerne le site consulté. La barre d'outils de l'application, met à la disposition de l'utilisateur, un certain nombre d'outils pour gérer les annotations des documents consultés. Les nouvelles annotations sont transmises au proxy pour y être sauveées dans une base de données et les documents consultés enrichis des annotations le concernant (en respectant les permissions attribuées aux annotations : publiques, privées à un groupe...). Notons que l'utilisateur peut à tout moment repasser en mode "classique" de navigation (i.e. la navigation n'est plus supervisée par le système) en utilisant la barre d'adresse initiale de son navigateur. En ce qui concerne la consultation de sites qui ne sont supervisés par aucun proxy (i.e. nouveau site n'existant pas dans la base des sites supervisés), le portail mettra en oeuvre un algorithme (à définir) d'analyse de charge des proxy, pour transmettre la requête au proxy retenu. Un système d'alerte (envoi de mail, par exemple), avertira l'administrateur du portail dans le cas où la charge de travail de l'ensemble des proxy ne permettrait plus une qualité de service acceptable. L'administrateur pourra alors réagir en fonction et, par exemple, décider d'ajouter un nouveau proxy dans le système.

## 6 Conclusion

Les spécificités du Web en font un outil très difficile à exploiter. Les outils d'annotation sont une voie prometteuse pour l'échange et le partage d'informations. Ils permettent d'espérer atteindre un objectif primordial des concepteurs du Web à savoir un environnement collaboratif où chacun est aussi bien lecteur que rédacteur. Concrètement, les systèmes développés soit se sont soldés par des échecs commerciaux [16, 15] soit, projet de recherche, ont été abandonnés [19, 13, 3]. Nous pensons que ces échecs sont particulièrement dus à deux causes : une mauvaise connaissance des annotations et des utilisations que l'on peut en faire et des architectures inadapées. De plus, ces annotations sont un potentiel considérable afin d'améliorer la compréhension et la manipulation des documents dans le cadre du partage d'information et de l'interopérabilité. Un cas particulier d'annotation se développe actuellement autour du Web Sémantique. Ces annotations un peu particulières, puisque plus orientées vers un traitement automatique des documents demandent encore à être spécifiées et nous pensons que les mettre dans le contexte plus général des annotations est bénéfique. Nous avons présenté une nouvelle architecture pour les systèmes d'annotation qui a comme propriété d'être légère, efficace, non-intrusive, indépendante de la plate-forme, supportant le passage à l'échelle et le travail collaboratif. Cette architecture est distribuée. Elle reprend à la fois la philosophie des systèmes basés sur des serveurs mandataires (mais les proxy sont distribués) et celle des systèmes à base d'intermédiaire client (l'intermédiaire client ici ayant pour rôle la communication avec le portail et avec le proxy qui référence les documents à annoter). Le système est en cours de développement. Nous l'utiliserons comme plate-forme expérimentale pour des applications de e-learning et surtout comme support à notre plate-forme d'indexation sémantique de document [5]. Les annotations apposées à l'aide de Dinosys seront exploitées lors de la phase de détermination des descripteurs des documents.

## Références

- [1] BREMER. Web annotations. <http://www.db.cs.ucdavis.edu/bremer/annotations.html>, 2002.
- [2] L. DENOUE. *De la création à la capitalisation des annotations dans un espace personnel d'informations*. Thèse de doctorat, Univ. De Savoie, 2000.
- [3] L. DENOUE et L. VIGNOLLET. Yawas : un outils d'annotation pour les navigateurs du web. In *IHM'99*, Montpellier, France, 22-26 Novembre 1999.
- [4] E. DESMONTILS et C. JACQUIN. Annotation sur le web : notes de lecture. Journées de l'AS Web sémantique, <http://www.lalic.paris4.sorbonne.fr/stic/>, octobre 2002.
- [5] E. DESMONTILS et C. JACQUIN. *Indexing a Web Site with a Terminology Oriented Ontology*, pages 181–197. IOS Press, I.F. Cruz, S. Decker, J. Euzenat, D. L. McGuinness édition, 2002.
- [6] J. GARFUNKEL. Web annotation technologies. <http://look.boston.ma.us/garf/webdev/annotate/software.html>, 1999.
- [7] G. GOLOVCHINSKY. Emphasis on the relevant : Free-form digital ink as a mechanism for relevance feedback. In *ACM SIGIR*, Melbourne, Australie, 1998.
- [8] R. M. HECK, S. M. LUEBKE, et C. H. OBERMARK. A survey of web annotation systems. Rapport technique, Dep. Of Mathematics and Computer Science, Grinnell College, USA, 1999.
- [9] IMARKUP SOLUTIONS. Imarkup. <http://www.imarkup.com/>, 2002.
- [10] M.-R. KOIVUNEN, J. KAHAN, R. SWICK, et E. PRUD'HOMMEAUX. Annotea project. <http://www.w3.org/2001/Annotea/>, 2001. W3C.
- [11] C. C. MARSHALL. Toward an ecology of hypertext annotation. In *ACM Hypertext*, 1998.
- [12] I. OVSIANNIKOV, M.A. ARBIB, et T.H. MCNEILL. Annotation technology. *Int. J. Human-Computer Studies*, 50, 1999. <http://www-hbp.usc.edu/Projects/annotati.htm>.
- [13] Jacob PALME. Web4groups, 1998. <http://www.dsv.su.se/jpalme/w4g/web4groups-summary.html>.
- [14] P. PERRY. Web annotations. <http://www.paulperry.net/notes/annotations.asp>, 2001.

- 
- [15] M. RÖSCHEISEN, C. MOGENSEN, et T. WONOGRAĐ. Commentor. <http://hci.stanford.edu/commentor/doc/>, 1994.
- [16] Inc THIRDVOICE. Thirdvoice 2000, 2000. <http://www.thirdvoice.com>.
- [17] V. VASUDEVAN et M. PALME. On web annotations : Promises and pitfalls of current web infrastructure. In *HICSS*, 1999.
- [18] XLIBRIS, 2002. FX Palo Alto Laboratory, Physical/Digital Documents research program, Fuji-Xerox Co. Ltd.
- [19] K.-P. YEE. Critlink. In *Fifth Foresight Conference on Molecular Nanotechnology*, <http://crit.org/critlink.html>, 1997.
- [20] R. ZEILIGER. Nestor : The web browser and cartographer. <http://www.gate.cnrs.fr/zeiliger/nestor.htm>, 2001. CNRS.
- [21] R. ZOHAR. Web annotation - an overview. Rapport technique, Dept. of Electrical Engineering, Israel Institute of Technology, 1999.





# Vers un système d'annotation distribué

**E. DESMONTILS, C. JACQUIN, L. SIMON**

## Résumé

De nombreux systèmes de partage d'information existent de nos jours mais les spécificités du Web en font des outils extrêmement difficiles à exploiter. Les outils d'annotation visent à améliorer échange, communication et interopérabilité sur le Web. L'objectif de cet article est d'une part, de faire une synthèse des caractéristiques des annotations et des architectures des systèmes d'annotation et d'autre part, de proposer une nouvelle architecture pour un système d'annotation simple d'utilisation, léger, efficace, non-intrusif, évolutif, partagé et indépendant d'une plate-forme.

Catégories et descripteurs de sujets : H.5.3 [**Information Systems**]: Group and Organization Interfaces

Termes généraux : Collaborative computing, Computer-supported cooperative work, Web-based interaction, Synchronous/Asynchronous interaction

Mots-clés additionnels et phrases : Annotation, notation, WWW, Internet, Proxy distribué