

De l'(in) utilité de W3 communication et information vont en bateau

Hervé Le Crosnier, Herve.Le_Crosnier@info.unicaen.fr

Le service d'hypertexte en réseau W3 (*World Wide Web*) est en passe de devenir le service fédérateur de l'Internet. Le succès de W3 a déjà transformé les applications de base du réseau :

- **le transfert de fichiers** utilise désormais principalement cette route, qui intègre un interface graphique pour les services de *FTP anonyme*, mais aussi la description des fichiers dans le texte même de services W3, remplaçant le système malhabile des fichiers *ReadMe*.
- **les accès TELNET aux services publics** (bibliothèques, banques de données...) sont maintenant remplacés par des modes de requête par formulaires. L'exemple du service JOCONDE pour l'accès aux banques de données du ministère de la culture en est un exemple type(<http://www.culture.fr/culture/spic.html>). Le développement de normes de requêtes et de transmission des données bibliographiques comme Z39.50 et ISO S&R vont favoriser cette intégration dans le service W3. Des logiciels documentaires proposent maintenant des interfaces pour le service W3 (requêtes en mode formulaire et présentation des résultats sous forme de pages HTML... comme dans *TextoWeb*).
- **les visualiseurs du service W3** intègrent de plus en plus des outils permettant de lire divers formats de fichiers (gif, pdf, jpeg... bientôt les sons et les vidéos) sans avoir à faire appel à des applications externes, facilitant le modèle "sélectionner-cliquer" pour l'utilisateur, et dispensant le néophyte d'une installation complexe. Ceci est aussi sensible pour l'installation de *lecteurs de niouzes* notamment dans le visualiseur *NetScape*. Le service USENET apparaît ainsi pour de nombreux utilisateurs néophytes comme intimement lié au service W3.
- le **courrier électronique** lui-même est concerné, notamment avec l'intégration par le navigateur *NetScape 2.0* d'un outil de courrier à la norme Secure MIME.

On commence donc à voir apparaître un fort recouvrement entre Internet et le service W3. Ceci est encore plus fort pour le grand-public, et même les "amateurs avertis".

Cela nous impose de réfléchir autour de deux thèmes : les raisons du succès de W3, mais aussi les limites du modèle et les moyens à mettre en œuvre pour les outrepasser. Cela nous conduira à conclure sur les nouveaux développements concernant la rédaction des documents hypertextes en réseau et sur la constitution des bibliothèques numériques.

Les raisons d'un succès

À l'origine de W3 on trouve plusieurs volontés qui ont fortement marqué la première génération de services :

- **l'intégration d'images dans le fil même du texte** est à l'origine du succès de W3, notamment dans sa concurrence avec les services GOPHER. Devenue pléthorique, cette utilisation tend à alourdir les services, cachant le contenu derrière le "joliesse" de la présentation. C'est bon pour les financeurs, qui se dispensent souvent de rechercher de l'information, mais souhaite voir valoriser leur signature dans une "ambiance moderne". La multiplication des fonds colorés accentue cette tendance à surcharger le contenu au détriment de l'information. Rappelons à ce titre que la typographie est avant tout un art de la mise en valeur du texte. Une belle typographie s'efface derrière le texte. On ne la remarque pas.
- **la notion d'hypertexte** qui permet d'intégrer les *ancres* dans le contenu même du document. Les services GOPHER, en limitant l'organisation du texte et notamment des pages d'accueil à un seul modèle de menus emboîtés rencontrait les limites spécifiques de ce mode d'accès à l'information : le texte de chaque item était trop court pour être significatif, la hiérarchisation stricte des menus avait tendance à perdre l'utilisateur. Le modèle hypertexte permet divers chemins d'accès à la même information. Il permet aussi de mettre en valeur l'information à laquelle on va accéder, tout en offrant un début d'indication sur ce que l'utilisateur peut espérer trouver.

Dans la théorie générale des hypertextes, on distingue les hypertextes en mode page (un nœud d'information = un écran, modèle *HyperCard*) qui favorise le maintien des repères du lecteur, et un hypertexte en mode texte (le logiciel GUIDE étant le prototype) dans lequel l'écran ne découvre qu'une partie du texte. C'est ce modèle qui est choisi par HTML. Il s'applique bien à la lecture de textes longs (manuels techniques, articles scientifiques...), mais est plus difficile à gérer pour des pages de sommaire, des présentations de produits, des formulaires de requêtes...qui contrairement sont les principaux types d'information que l'on trouve sous W3.

- **la notion d'hypertexte en réseau** qui permettait de construire un service en se contentant de réaliser des "pages d'accueil", véritables recueils de pointeurs vers les informations produites et proposées par d'autres. Avec la multiplication de ces développements "parasites", l'ensemble du service W3 est devenu largement circulaire, accentuant la perte de repères typique de la navigation hypertexte. Les véritables "producteurs d'information" sont

noyés dans la multiplicité de services et ne peuvent recevoir le "retour aux investissements" de leur travail (nous entendons par ce terme autant le "retour de valorisation" typique de l'investissement universitaire que le "retour financier" que pourraient attendre des services commerciaux). L'information devient un faire-valoir des rédacteurs de "pages d'accueils". Mais cela n'est pas sans rappeler les journaux imprimés dont le contenu rédactionnel n'est qu'un prétexte à la vente d'espace publicitaire.

Cette conception des services W3 comme "recueils de pointeurs" est aussi mise à mal par le succès même du réseau : il devient de plus en plus difficile d'accéder et de lire des documents placés au bout du monde. Le rôle des bibliothèques numériques est alors d'organiser la duplication des documents afin de permettre une lecture fluide des hypertextes en réseau.

- **la notion de déplacement et reproductibilité des pages HTML.** Une page HTML 1.0 était auto-suffisante, emportant avec elle les ancres et les outils de balisage typographique nécessaires. Cette notion permettait d'envisager le "transcodage" de fichiers de texte traditionnels : traitement de texte (*RTF2HTML*), de formateurs de texte (*LaTeX2HTML*). La transformation en hypertexte consistait souvent à la banalisation du mode "sommaire" et à la mise à la disposition du public sur le réseau mondial.

Cette notion est mise à mal par le développement des *CGI-Scripts*. Malgré cette fluidité des premiers documents HTML, nous avons peu assisté à la constitution de "*Bibliothèques numériques*", se fixant comme objectif de dupliquer les documents existant et d'assurer la conservation en mode local (enfin, une localisation par pays, ou par unité géographique ou linguistique). Au contraire, la généralisation de pointeurs vers des "archives d'origine" est devenue le moyen de créer aisément une bibliothèques de journaux électroniques ou de textes numériques. Les *CGI-Scripts* renforcent cette imbrication étroite entre le document et son lieu de production. Il est certes plus valorisant pour un gestionnaire de service W3 de multiplier les images à zone sensibles... Est-ce cohérent avec le développement de documents portables ?

L'ensemble des caractéristiques de la première génération de services W3 a été à l'origine à la fois du succès du paradigme proposé, mais aussi de ses limitations ultérieures, telles qu'elles apparaissent aujourd'hui. Le succès participe de la conception actuelle du multimédia, plutôt conçu comme une "juxtaposition de médias". W3 permet de se placer dans cet univers où se cotoient images, sons, textes et vidéo sans demander l'élaboration d'une grammaire complexe. Les modèles du logo et du siglage, déjà banalisés par la télévision (en haut à droite de chaque écran) et du commentaire devenaient aisés à mettre en oeuvre. La réalisation de "services" a été privilégiée sur la réalisation de "documents". Et souvent, derrière le service on retrouve :

- la version électronique de **la plaquette de présentation** (de l'université, de l'entreprise, de la région...) Peu d'informations complémentaires, et une lisibilité moindre. Sans résoudre pour autant le problème de la mise à jour, souvent parce que les concepteurs n'ont pas attaché assez d'importance au

personnel et aux compétences humaines nécessaires à l'alimentation et au renouvellement d'un service documentaire. Combien de plaquettes obsolètes, d'annonces de congrès déjà tenus, de dates dépassées traînent dans les services ?

- **la collection de pointeurs**, réalisant le vieux rêve de la Renaissance d'une bibliothèque personnelle riche de tous les livres (i.e. les documents électroniques), et dont l'organisation même révélait son propriétaire à lui-même et aux autres (le développement des *ex-libris*, comme marque d'appartenance du document à une collection privée).

Les inconvénients de ce type de collection sont multiples :

- le pointeur est plus un instrument de "distinction" que d'information (*my favorite home pages*)
- la maintenance est hasardeuse (déplacement des pointeurs, suivi des publications...)
- la cohérence de la collection et son adéquation à un objectif fixé est souvent difficile à justifier (et c'est tout à fait normal pour des "centres d'intérêts personnels", par opposition à la pratique collective des bibliothèques du XXème siècle).

Les avantages sont à rechercher dans la découverte de documents par un spécialiste d'un micro domaine de la connaissance, ce que peut difficilement concevoir une bibliothèque, surtout dans un univers où la publication n'est pas régulée comme dans le domaine du livre imprimé (annonce des nouveautés, dépôt légal permettant d'avoir connaissance de "tout ce qui paraît" - contrôle bibliographique universel)

Ce foisonnement est certainement nécessaire pour assurer la vitalité d'un service naissant comme W3 (et plus généralement l'Internet). On voit maintenant apparaître deux types de projets qui sont certainement plus longs à mettre en œuvre, mais qui assurent à plus long terme la maintenance de l'information électronique:

- la rédaction de textes qui intègrent dans leur conception même les possibilités de W3. De nombreux cours et ouvrages de formation sont ainsi proposés sur le réseau qui permettent de tester en direct les options proposées. L'utilisation pleine et entière du réseau offre alors des possibilités nouvelles, différentes de l'univers du livre.

Un exemple de cette démarche est le manuel de conception de services de l'Université de Yale (*Web Style manual*, par Patrick J. Lynch. <http://info.med.yale.edu/caim/StyleManual_Top.HTML>).

Cette orientation est aussi sensible dans les journaux électroniques. La première génération de journaux électroniques était diffusée par messagerie et devait se contenter de l'ASCII, et reprenait le modèle traditionnel des articles textuels. Les journaux électroniques utilisant W3 sont au centre des nouveaux projets. Ils permettent :

- d'offrir des informations qui ne trouvent pas leur place dans le texte principal d'un article (les annexes scientifiques, ou les listings en informatique).
- d'offrir la possibilité de consulter les documents d'origine qui ont permis la rédaction d'un article. Le journal de la Suisse romande *L'Hebdo*, propose ainsi une rubrique "documents" : *Les textes originaux sur les sujets d'actualité qui ne peuvent pas prendre place dans «L'Hebdo», vous les trouverez ici. Vous avez ainsi un accès direct aux sources de l'information* (Whebdos : <<http://www.hebdo.ch/>>).
- d'offrir la possibilité d'associer des documents dans d'autres médias aux textes, comme par exemple des démonstrations dans le Journal of Artificial Intelligence Research : (JAIR, <http://www.cs.washington.edu/research/jair/home.html...>).
- la constitution de *bibliothèques numériques* qui proposent les services d'une bibliothèque dans l'univers du réseau, notamment : la description documentaire et catalographique des documents, la maintenance des archives, la duplication des documents et plus généralement la garantie démocratique de l'accès aux documents pour tous.

Les modèles des bibliothèques numériques sont encore en chantier. Les moteurs de recherche (*Webcrawler*) ou les bibliothèques réalisées manuellement (*Yahoo*) sont des premières réponses. Dans ce domaine, imaginons que nous sommes face à une pléthore de documents d'un nouveau type (on peut évaluer à plusieurs centaines de milliers de documents électroniques originaux, représentant plusieurs millions de pages) sans que les procédés de duplication, de conservation, de catalogage et de mode de consultation aient été définis. Gardons-nous de croire que ces opérations sont inutiles. Que deviennent les archives de "journaux électroniques" quand leur "éditeur" a disparu ? que deviennent les collections de pointeurs d'un chercheur ou d'un laboratoire quand les sujets de préoccupation changent ? Comment se reproduit dans l'univers électronique le modèle de validation de la science mis en place depuis deux siècles ?

La duplication des documents électroniques occupe une grande part des préoccupations concernant l'accès universel à la documentation. Cette conception est fondatrice pour les bibliothécaires. Le fait de traiter des documents multiples les distingue des archivistes qui échantillonnent dans une production d'unicats, et des conservateurs de musées qui collectionnent des pièces uniques. Le syndrome d'Alexandrie vient aussi rappeler la nécessité de dupliquer l'information. Les incidents géopolitiques de ces dernières années renforcent cette conception (destruction de la Bibliothèque de Sarajevo, embargo des Etats-Unis sur les documents et les logiciels traitant de la cryptographie...).

Or cette préoccupation n'est que rarement prise en compte par les concepteurs de services W3. Si les premiers documents HTML emportaient avec eux leurs ancres et constituaient un document "auto-suffisant", la multiplication des appels de scripts (*Imagemap, formulaires, génération automatique de pages W3 en fonction du contexte...*) tend au contraire à localiser fortement un document. La capacité à annoter en réseau un document de base, mise en avant dès l'origine de l'hypertexte par Ted Nelson et son projet *Xanadu*, est aussi un élément à prendre en compte dans la duplication des documents. De cette figure apparaît la nécessité de concevoir des modèles sociaux de duplication entre sites producteurs d'information. L'heure est révolue ou un utilisateur pouvait dupliquer sur son propre poste de travail (ou son serveur) un document hypertextuel HTML, en raisons des multiples "appels de scripts". Il faut aujourd'hui élaborer des règles de transfert d'un ensemble de documents constituant un "service", et établir la réciprocité des échanges (les annotations faites sur un serveur en Europe doivent aussi être reportées sur le serveur des États-Unis et réciproquement).

La constitution de bibliothèques numériques est devenu un enjeu d'organisation du cyberspace. Car en sens inverse, le succès de W3 se traduit aussi par la difficulté à accéder à des documents à certaines heures, les liens physiques étant saturés. Or la lecture hypertexte nécessite une fluidité des changements de pages. Le suivi des liens est devenu aujourd'hui une source d'impatience devant la lecture sur écran, qui tend à dévaloriser l'ensemble du service W3 et ses possibilités nouvelles. Que dirions-nous si pour la lecture de chaque article de journal imprimé il nous fallait passer par le rythme du prêt-inter bibliothèques ?

Pour accélérer la conception de bibliothèques numériques, il faut aussi concevoir que le réseau est un support à des "documents", et pas seulement des "services". Cette idée commence seulement à percer. Nous avons longtemps, notamment en France, été façonnés par le modèle des services vidéotex. Dans ce modèle, un service est entièrement organisé et nourri en information par un concepteur unique. Ce n'est plus le cas dans les services W3. Cela permet d'envisager la collaboration. Non pas une collaboration parasite, où chacun se repose sur un autre pour la production documentaire, le tissage de liens et l'archivage, mais bien une collaboration coopérative : une organisation du partage des tâches de collecte, de description et de mise à disposition des documents. Ce modèle coopératif est certes plus long à mettre en œuvre : il faut définir des champs de compétences, des degrés de complétudes des collections souhaitées, des règles de description tant du contenu (indexation documentaire) que de la forme des documents (catalogage descriptif). mais il est le seul à partir des besoins actuels et futurs des lecteurs et non de l'intérêt des offreurs de services. C'est une conception qui doit émerger par comparaison avec le fonctionnement, les missions et les principes des bibliothèques, plutôt que par la pâle copie du modèle de la télévision ou du vidéotex.

Un autre aspect du développement des services W3 est aussi devenu à l'ordre du jour : jusqu'à présent HTML reste un mode assez fruste de mise en page mais surtout repose sur une grammaire hypertextuelle élémentaire : l'appel de document par activation d'une ancre. On est encore en deçà des nombreuses recherches sur une rhétorique de l'hypertexte de la fin des années 80 (Roland Dachelet, *Hypertexte et hypermédia : documents, informations, connaissances*. In : *Le document électronique, Cours INRIA*,

11-15 juin 1990.). Un typage sémantique des liens est devenu indispensable pour offrir une réelle lecture hypertexte qui ne soit pas simplement un jeu de "marche-avant, marche-arrière" avec des sommaires intermédiaires. Les divers types de liens d'un logiciel d'hypertexte comme GUIDE sont des exemples des développements nécessaires pour faciliter la lecture. Ainsi le "lien d'annotation" de GUIDE renvoie l'ouverture d'une petite fenêtre comportant la remarque, ou la référence bibliographique.

De même, le discours actuel sur les informations électroniques fait souvent appel à la notion de versions, à l'idée d'une "maintenance" d'un document... Il faut là aussi se poser des problèmes avec un peu de recul. De ce point de vue, tous les documents ne se valent pas. La maintenance d'une page de sommaire est une nécessité technique : s'assurer que les pointeurs sont toujours valides (dans le temps) ou efficaces (pour des documents dupliqués, les pointeurs font ils appel au serveur qui répondra le mieux pour un utilisateur local). Le suivi d'un document technique, d'un manuel de procédure... est aussi dans l'ordre des choses.. même si toutefois la conservation des anciens manuels est un outils précieux pour ceux qui, dans quelques années ou dans quelques siècles, se pencheront sur l'évolution des techniques en cette fin du XXème siècle. La modification d'un document porteur de sens, de point de vue, d'expérience est plus problématique. Ce qui change dans le temps, c'est la connaissance. Celle d'un environnement social et scientifique, celle d'un individu donné... Mais ce mouvement de la connaissance se construit à partir de référents stables, que sont les documents publiés à un moment donné. Les peintres pratiquaient le "vernissage" des toiles afin de s'interdire toute retouche. Les imprimeurs apposaient "l'achevé d'imprimer". Il convient d'élaborer de même un rite de publication sur le réseau. Afin que des points stables soient offerts à la lecture, à la critique, à la relecture... et parfois aussi à la réhabilitation. Qu'un auteur change d'avis, ou souhaite intégrer des critiques, des remarques et des apports de collègues, rien de plus normal, rien de plus souhaitable. Il peut alors publier, soit une "nouvelle édition", soit un autre article qui montre que son point de vue change, évolue, se peaufine... La gestion des version est un problème lié à certains types de documents et ne peut pas être généralisée. Pour des motifs cognitifs, pour la clarté des débats et par respect pour l'avenir comme pour le passé.

En marge de ces questions portant sur l'unité documentaire, son statut dans l'univers des publications en réseau, on trouve aussi des questions portant sur l'architecture même de ce document électronique de base. Dans sa version actuelle, HTML est avant tout un langage linéaire. Les informations se succèdent. Même l'appel de documents associés (images, sons, vidéos...) se déroulent suivant un mode séquentiel (appel du document, chargement puis lecture avec un logiciel spécifique ou intégré). Or nous voyons apparaître des systèmes permettant la diffusion en direct du son (Real Audio <<http://www.realaudio.com/>>, dont on peut entendre un exemple dans le service de France-Info (<<http://www.publicis.fr/rf/>>). Dès lors, il devient nécessaire de concevoir des architectures de documents réellement multimédia. Une telle architecture devra intégrer une gestion du déroulement séquentiel du son et de la vidéo avec la lecture par accès direct du texte et des images fixes. Le modèle n'est plus seulement SGML, mais des normes du type de *HyTime*.

Comment de tels documents existant sur un réseau vont-ils pouvoir être réellement conservés, dupliqués, offerts à l'accès par plusieurs utilisateurs simultanément, mais

aussi par plusieurs accès successifs d'un même "lecteur" (rôle des bibliothèques numériques) ? On le voit les enjeux sont largement ouverts. L'expérience acquise aujourd'hui dans la rédaction de véritables documents HTML, portant de l'information nouvelle et l'incorporant dans un modèle de lecture innovant devient alors une garantie de développement des futurs services d'hypertexte en réseau. Il restera ensuite (et parallèlement) à définir et maîtriser un modèle de description documentaire et catalographique de ces nouveaux documents électroniques. Un projet plus ambitieux et novateur que la simple numérisation de textes du passé, qui constitue encore aujourd'hui l'horizon de nombreux projets. Un projet coopératif, qui implique de revoir la notion de document, de collection et de bibliothèque. Le réseau est alors un nouvel outil pour penser : de nouveaux moyens, mais aussi un nouveau regard sur les méthodes et l'organisation sociale de l'univers de la connaissance.

Caen, le 2 octobre 1995
Hervé Le Crosnier
Université de Caen