



## gestion de contenus web

# au tournant des bases de connaissances

jean-pierre cahier

**Davantage collaborative et sémantique, la gestion de contenus web est à un tournant : d'une simple formule de performance industrielle pour le dépôt à la chaîne de contenus en ligne, le champ des outils s'élargit, dans de nombreuses applications, pour aider à capitaliser des bases de connaissances, où les utilisateurs seront bientôt autant contributeurs que consommateurs de contenus liés à leurs métiers.**

Traditionnellement, la gestion de contenus web vient s'intercaler entre, en amont, les techniques de Ged (Gestion électronique des documents) et de création de documents (éditeurs HTML, outils bureautiques) et, en aval, les serveurs d'applications et de personnalisation qui aboutissent à la publication des pages sur les sites. Pour les applications intranet et extranet, professionnelles ou grand public (notamment dans le secteur des médias), la gestion de contenus est le maillon particulièrement critique de la chaîne éditoriale : il s'agit – souvent à flux tendus – de stocker les contenus créés par les rédacteurs, de centraliser les références, d'enchaîner les opérations de préparation et de validation. Il faut aussi souvent associer le contenu à différents modèles de pages, surtout lorsque celles-ci doivent mener à une publication multicanale et multicible, comme lorsqu'on souhaite s'adresser aussi à des utilisateurs nomades utilisant des téléphones Wap ou des assistants portables.

Les outils de gestion de contenus se sont à l'origine constitués autour de problèmes assez techniques, tels que celui de la maintenance et de l'enrichisse-

ment des applications web, rendus difficiles à cause de l'imbrication du logiciel, des gabarits graphiques et du contenu proprement dit. Chacun peut désormais constater que les applications Inter ou intranet comportent rapidement des milliers de documents et de liens, les uns et les autres devant souvent être actualisés. Le fait de devoir ressaisir ou retoucher manuellement le contenu dans le cas d'améliorations d'interactivité des sites – voire, lors de simples modifications de charte graphique et navigationnelle – est vite devenu une corvée intolérable pour les webmasters. De plus, ces derniers n'ont souvent qu'une connaissance superficielle des contenus, ce qui engendre des erreurs. Sur les plus gros sites, une telle entreprise est même devenue matériellement impossible.

### la fonction de référentiel

Dès 1998, les grandes entreprises confrontées à la croissance rapide de leurs intranets ont rapidement été contraintes d'investir dans des outils à l'époque fort coûteux, tels que Documentum, IBM Content Manager, Interwoven... fournissant un référentiel de tous les composants d'une application web, et une maîtrise du cycle de vie des contenus et des outils pour rendre les rédacteurs autonomes par rapport aux développeurs et aux graphistes. La fonction de référentiel, en particulier, joue un rôle clé pour garantir une source unique des informations et éviter les incohérences entre des modifications réalisées par différents rédacteurs. « *Un outil doit aussi supporter la gestion de versions et les annulations multiples, qui sont courantes dans la gestion de contenus collaborative* », précise Arnaud Lefèvre, de la société Nuxeo.

Cette nécessité d'industrialiser la gestion de contenus, en séparant bien l'action des différents métiers, a ensuite gagné des entreprises de toutes tailles et des contenus de tous types. Aux fournisseurs déjà cités sont venues s'ajouter des offres telles que celles de Noheto, Nuxeo, Mayetic,





- • • Stellent, Vignette... si bien que l'on compte aujourd'hui plusieurs dizaines de fournisseurs de solutions de gestion de contenus actifs en France, avec chacun un portefeuille important de clients. Certains éditeurs sont spécialisés sur des segments tels que les contenus "one-to-one" (Broadvision), les contenus de presse (Rosebud), etc. Cette démocratisation des applications en gestion de contenus web s'est poursuivie en 2002, pourtant marquée par le contre-coup du dégonflement de la bulle Internet. La rationalisation du processus de gestion des contenus a gagné les intranets des PME, et des applications extranet, s'est intensifiée dans

Zope, Nuxeo, RedHat...) et un effet d'entraînement considérable dû au déploiement soutenu du standard XML dans les contenus eux-mêmes. XML est en effet le moyen par lequel la sémantique va être exprimée dans les attributs.

## les contenus web plus impliqués dans les activités métier

C'est le besoin d'industrialisation du processus préparant la publication qui, depuis le début, a fondé le recours aux outils de gestion de contenus. Mais

The screenshot shows the homepage of the French Ministry of the Interior. At the top, there is a navigation bar with links for 'Plan du site', 'Téléchargement', 'Liens utiles', and 'Contact'. Below this is a search bar with the text 'Recherche rapide' and 'Tapez un mot-clé'. The main header reads 'Bienvenue sur le site du Ministère de l'Intérieur !' and includes options for 'Imprimer cette page', 'Version texte', and 'Envoyer'. The main content area is divided into several sections: 'A LA UNE' with a headline about a decrease in crime, 'Les autres actualités' with sub-sections on regional intervention groups and intercommunal consolidation, and 'Et aussi...' with news about a national security law and a police school. The left sidebar contains a menu for 'A LA UNE' and 'A VOTRE SERVICE' with various service links. The right sidebar features 'L'événement', 'Annonces des lib', 'Calendrier et commun', 'Zoom', 'CONDUITE HIVERNALE', 'LE MINISTÈRE RECRUTE', and 'CONCOURS'.

La gestion de contenus du nouveau site web du ministère de l'Intérieur, qui permet un enrichissement direct par de nombreux contributeurs non-techniciens, est entièrement assurée par des logiciels libres (Zope et Nuxeo-CPS).

les départements d'entreprise, les catalogues de commerce électronique. Ce mouvement rapide a bénéficié de la spirale entre l'évolution des technologies des logiciels (l'EAI\*, moins chère, et depuis peu la maturité des logiciels libres dans ce domaine avec des produits de gestion de contenus chez

à côté de ce besoin, on voit depuis une période récente se rajouter un nouvel enjeu majeur de ces outils, provoquant l'apparition de fonctionnalités nouvelles pour rendre les outils plus intelligents et plus collaboratifs. Selon une enquête récente Dauphine-Cegos\*\* sur l'utilisation des nouvelles

technologies en France, les trois-quarts des entreprises possèdent désormais un ou plusieurs sites Internet, mais la proportion est désormais inversée dans leur contenu entre, d'une part, les informations générales et, d'autre part, les contenus spécifiques aux métiers et les applications vécues par les utilisateurs comme des bases de connaissances utilisées dans l'exercice de leur activité. Même si ce glissement s'effectue très progressivement, il signe la fin de la phase des sites vus uniquement comme supports de communication. Surtout, il représente aussi un changement important dans les exigences sur la façon de gérer le contenu. Par exemple, 56 % des salariés interrogés estiment disposer de telles bases de connaissances dans leur intranet, en termes d'accès à l'information, et 47 % estiment qu'il peuvent disposer de ces bases pour les alimenter en contenus nouveaux. Pour Claire Beyou, spécialiste de gestion de connaissances chez Cegos et co-réalisatrice de cette étude, « il faut rester prudent quant au fait que ces bases soient réellement alimentées et permettent la recherche de contenus. De plus, les utilisateurs évoquent sous le terme "bases de connaissances" des contenus très différents, du simple répertoire partagé au système structuré offrant des fonctions de dépôt, publication et recherche d'informations. Mais les pratiques de capitalisation commencent à avoir une réelle existence ».

De fait, les entreprises interrogées dans cette étude estiment que lorsque les collaborateurs trouvent une solution à un problème, ils alimentent en contenus une base de connaissances dans 12 % des cas (sous forme d'alimentation systématique) et dans 20 % des cas (alimentation plus occasionnelle). Même si ces populations de contributeurs en contenus restent encore minoritaires, ces chiffres traduisent un début de participation des acteurs du terrain à l'alimentation directe en contenus, du moins dans les entreprises qui encouragent cette attitude.

### un réflexe quotidien

Le cas du projet Copernic, débouchant sur le nouveau portail "Impôts" du ministère des Finances, destiné à un large public, est caractéristique de cette tendance : pour le lancement de cette nouvelle version du portail Impots.gouv.fr en mars 2003, ce sont dans un premier temps 100 à 150 employés de cette administration qui sont mobilisés comme contributeurs de contenus pour créer puis enrichir en permanence les contenus très détaillés de ce service (expertise réglementaire, formulaires, conseils, Faq - Foire aux questions -, dialogue avec les admi-

nistrés...) ; et ce chiffre devrait monter à une vitesse de croisière de 600 contributeurs dans les prochains mois.

En mobilisant directement les employés compétents et en leur permettant de valoriser leur expertise sur le Web, l'administration espère ainsi fournir au contribuable une finesse d'expertise et une qualité d'information (garantie actualisée en matière de réglementations) que n'aurait jamais atteint un service de communication ou même un bataillon de webmestres aguerris. Pour cela, l'application qui a été conçue avec le concours de la société Cosmosbay, autour d'outils Noheto, s'appuie sur une variété de dispositifs fonctionnels, essentiellement collaboratifs (circuits de discussion interne, de validation...) qui vont rendre possible la cohérence du système et l'appropriation de la gestion de contenu dans le quotidien de centaines d'employés.

Selon Paul Perdrieu, de Noheto, « à partir du moment où contribuer au Web devient un réflexe quotidien dans le cadre des métiers de nombreuses personnes, les fonctions des outils doivent être optimales pour aider ces professionnels dans tout le processus, par exemple à mettre en place rapidement les circuits de collaboration interne pour vérifier ou valider l'information, etc ». Dans les rôles mis en place pour Copernic, il y a par exemple des experts, mais aussi des personnes qui ont un rôle davantage marketing pour rediriger l'information, d'autres qui ont un rôle plus pédagogique afin que le contribuable comprenne bien les changements réglementaires, etc. « La gestion de contenus prend en compte un souci de proximité et doit être beaucoup plus fine au niveau sémantique, pour aider les acteurs à traduire et à classer les informations au bon endroit, de sorte que le client final les retrouve. » C'est là que l'on se rend compte de la maturité qui a été acquise en quelques années pour les outils de gestion de contenus, y compris pour le segment milieu de gamme où se situent les outils Noheto.

### un contrôle éditorial plus fin

Parmi les nombreux exemples qui voient le jour actuellement, un système de gestion de contenu sophistiqué a été développé rapidement pour le Système d'information territorial du Bas-Rhin (coordonné au niveau du préfet), dans le but de mettre en place un intranet avec des ressources documentaires de nombreux acteurs du département. En s'appuyant sur deux logiciels libres (Zope et Nuxeo-CPS, disponibles aussi bien sous Linux que sous Windows), la société Nuxeo a développé





- • • cette application qui dispose notamment d'un workflow avancé, d'un annuaire et de fonctionnalités de gestion documentaire. Plusieurs milliers d'utilisateurs sont prévus à terme, appartenant à toutes sortes de services et d'administrations.

Cet exemple montre, d'une part, que des applications sophistiquées sont désormais complètement crédibles dans un contexte complètement open source (le très gros site de CBS en production aux Etats-Unis est basé sur une gestion de contenus Zope, ainsi qu'en France le site du ministère de l'Intérieur, ou celui du journal *Le Monde informatique*). D'autre part, on sait maintenant implanter facilement des règles de gestion et d'authentification complexes, différenciées entre de nombreux types de rôles, tout en réutilisant des normes d'annuaires (LDAP, Lightweight directory access protocol), d'agenda partagé (c-Calendar) pour l'aide à l'organisation de réunions à la volée, de webmail (Imap). Les outils les plus avancés permettent d'affiner les droits des utilisateurs et des sous-groupes en fonction

de rôles : typiquement, les droits d'un membre peuvent être affinés en fonction des rôles de responsable, de membre, de visiteur, d'abonné privilège, de sous-ensembles de compétences, de zone locale d'un intranet. Les utilisateurs peuvent ainsi être alertés quand est mis en ligne un document lié à leurs rôles, leurs droits et leurs thématiques d'intérêt. « *Les moteurs de recherche se montrent désormais capables de s'y retrouver dans des groupes complexes, constate Arnaud Lefevre, de Nuxeo. La recherche multicritère est possible en qualifiant les documents depuis une variété de systèmes de métadonnées, et le contenu peut être qualifié par les documentalistes pour décentraliser la gestion et la mise à jour des sites. Avec le progrès dans la gestion fine des accès et dans le contrôle éditorial, la gestion de contenus commence à répondre efficacement au contrat inscrit dans ses principes fondateurs, qui est de permettre la création de contenu par des auteurs non qualifiés techniquement, tout en respectant la complexité des procédures organisationnelles de l'entreprise.* »

## au journal Sciences Humaines

### « une véritable capitalisation de connaissances »

Emmanuelle Garcia, responsable de la documentation au mensuel Sciences Humaines, est aussi chef du projet gestion de contenus et au cœur du dispositif, actuellement bien avancé dans sa réalisation, qui devrait voir le jour en mai 2003 sur le site *Scienceshumaines.fr*. Le projet la passionne : « C'est vraiment un projet de toute l'entreprise, qui amène à revoir d'une tout autre façon notre métier, puisque ce que nous allons proposer ce n'est plus seulement un journal, mais toute une capitalisation de connaissances, avec en plus la dimension de l'actualité. »

L'objectif est de mettre en ligne tous les articles écrits depuis 5 ans, pour que l'internaute puisse rechercher des articles à la carte selon une vision transversale et thématique suivant plusieurs points de vues (notamment ceux des diverses disciplines de sciences humaines), avec plusieurs niveaux d'accès et de tarification. « Cela me permettra par exemple de repropager des articles plus anciens, mais qui restent valables, moyennant un petit travail de réactualisation. » Toutes ces fonctionnalités demandent une grande finesse pour prendre en compte une indexation complexe et divers types de rôles. En particulier, l'accès est différent suivant les types de rubrique, les tarifs dépendant par exemple du fait que l'acheteur est abonné ou non (de façon à toujours encourager le support papier qui est vendu en kiosque), au désir des services supplémentaires comme les Echos des recherches, les chroniques de livres (le web permettra de doubler le nombre d'ouvrages traités dans cette rubrique très prisée – jusqu'à cinquante par mois), l'envoi par réseau d'un numéro pour les lecteurs très éloignés, etc. L'agenda sera produit directement en XML.

Ce projet est caractéristique des applications ambitieuses en gestion de contenus web que de petites entreprises n'hésitent plus à monter aujourd'hui. Sciences Humaines est en effet une petite société, basée à Auxerre et regroupant une trentaine de collaborateurs, mais qui s'est construit progressivement une solide réputation au sein d'un lectorat ciblé, tout en réunissant douze ans d'un très important fonds d'articles scientifiques, dont seules les cinq dernières années sont numérisées. L'outil utilisé est RatioMediaT de la société Rosebud (associé avec la base TextML Server de IxiaSoft en mode XML natif). Rosebud intervient également dans cette opération pour le conseil et les services de conversion XML (avec son outil e-Gate) et d'indexation des articles sur les critères structurels (le système peut ainsi traiter de façon très nuancée les encadrés, les fiches sur les auteurs, les fiches de définition des termes). « Cela fait quatre ans, conclut Emmanuelle Garcia, que nous étions tentés par ce projet, et que nous hésitions ; aujourd'hui, nous avons sauté le pas. Pour nous, la mise en place sera vraiment un grand moment, car cela représente l'aboutissement du projet encyclopédique du journal, imaginé par ses fondateurs il y a près de vingt ans. Et nous pouvons enfin le réussir. »

Comme le montre aussi l'exemple du mensuel Sciences Humaines (voir l'encadré), où l'accès et l'indexation se font suivant plusieurs points de vue métier, les outils de gestion de contenus commencent donc à proposer de réelles opportunités pour décentraliser la production de la sémantique de description, en multipliant la finesse du balisage XML, selon de nombreux axes structurels et thématiques, et en modulant l'accès aux contenus selon une variété de rôles. Les outils complémentaires comme les bases XML (IxiaSoft, Tamino...) et bien sûr les moteurs, outils de gestion de connaissances et de réseaux sociaux (Amoweba, Verity K2...) contribuent beaucoup à accélérer ces tendances. Du point de vue organisationnel, cela passe par exemple par la mise à contribution beaucoup plus active, sur le terrain, des documentalistes ou des producteurs des documents eux-mêmes : donc moins d'erreurs, plus de pertinence, des circuits de collaboration ou de traduction pédagogique mettant en correspondance plusieurs points de vue d'expert ou d'usage... « Les acteurs sont beaucoup plus maillés et motivés pour rendre les nuances des contenus qu'ils comprennent bien, souligne Aimée Potter, de Rosebud, avec des modalités permettant de travailler de façon très fine et très contrôlée au niveau des fragments de contenu, ce qui est souvent stratégique pour nos clients. » Toutes ces innovations, en se répandant, donnent au final de meilleures chances, pour les utilisateurs, de trouver ensuite ce qu'ils cherchent dans le maquis croissant des informations sur le web et les intranets.

### **L'esprit du web dans la gestion de contenus**

Les outils de gestion de contenus servent ainsi à industrialiser la construction permanente de ce qu'on peut appeler les structures sémantiques collectives, ce qui pose aussi des problèmes, plus politiques, de pouvoir sur ces structures. Toutes les applications de gestion de contenus web ne s'accompagnent pas en effet de l'esprit de décentralisation et des possibilités coopératives évoquées, qui correspondent originellement à "l'esprit du web". L'industrialisation spectaculaire des contenus sur Internet, à laquelle nous assistons actuellement, comporterait-elle donc aussi des effets pervers au fur et à mesure que l'on rend intelligente et indispensable la couche de standards, de moteurs et d'automatismes informatiques, notamment ceux qui se développent très rapidement pour les DTD (Document type definition) et les métadonnées accompagnant XML-RDF, Dublin Core, URI

(Universal resource indicator) ? Et il peut être tentant, dans le même mouvement, de s'appuyer sur des ontologies de domaine très développées, prenant elles aussi la valeur de standards.

Cette informatisation de la sémantique des contenus pourrait alors ne pas avoir que des effets positifs. C'est notamment le point de vue de Bruno Bachimont, chercheur en ingénierie des connaissances à l'Université de technologie de Compiègne (UTC), philosophe et spécialiste des ontologies de domaine : « Derrière la vogue actuelle de la notion de contenus et les évolutions actuelles de la gestion de contenus, on laisse en fait s'interposer une couche importante de normes, de langages documentaires et d'informatique qui, d'un point de vue sémiotique, aboutit de fait à "remonter" le seuil qui séparait traditionnellement la sémantique et la pragmatique. Il y a alors un risque de réduire la marge et la liberté d'interprétation humaine sur ces contenus et donc notre contrôle sur ces contenus. »

### **le web sémantique**

De fait, face au développement rapide des outils et aux façons d'utiliser cette richesse fonctionnelle, un débat a commencé à germer dans les communautés concernées, entre différentes approches de la gestion de contenus web : plus automatisées et sémantiquement rigides, ou au contraire plus décentralisées et collaboratives, tout va dépendre aussi des types de contenus mis en ligne, de leur sensibilité à l'actualité, de l'objectif des applications...

Au-delà, on voit se profiler plusieurs horizons pour ce que Tim Berners-Lee, l'inventeur du web, a commencé à baptiser depuis 1998 le "web sémantique", c'est-à-dire le moment où, dans quelques années, toutes ces techniques auront encore gagné en maturité et permettront aux utilisateurs de déposer et de trouver des informations de façon très ouverte sur la Toile, avec un niveau d'aide bien supérieur, pour des problèmes encore plus fins, avec des nuances et des associations sémantiques très importantes pour les métiers. A travers les exemples évoqués, c'est bien ce qui est en jeu avec le passage progressif, à partir du statut de fonds documentaires en ligne tels qu'ils sont encore souvent perçus, vers de réelles bases de connaissances sur le web, pouvant rendre de plus grands services. ●

\* *Entreprise application integration*

\*\* *Enquête 2003 de l'observatoire Dauphine-Cegos du e-management, publiée aux éditions Liaisons.*