

## Les outils de capture et de gestion de l'information électronique

# Vous ne pourrez plus vous en passer !

Ce ne sont ni des gestionnaires de signets (même si cette tâche peut rentrer dans leurs attributions), ni des gestionnaires de fichiers classiques, ni des aspirateurs de sites web (même si certains commencent à capturer des sites entiers d'un simple clic), ni des moteurs d'indexation et de recherche (même s'ils en intègrent un pour la plupart), ni des logiciels documentaires au rabais ...

Par Armelle Thomas

**Ils?** Ce sont les assistants personnels de management de l'information web, et plus largement, les outils de capture et de gestion de l'information électronique. En effet, si faire un usage professionnel de l'internet conduit bien sûr à isoler du contenu pertinent, la recherche n'est qu'une étape : reste à optimiser le classement pour un usage ultérieur, phase d'intervention de ces outils. Le concept n'est pas nouveau, puisque dès 1999, I-harvest (disparu depuis) permettait de stocker et d'organiser des ressources web, bientôt suivi par Cogitum ou Surfsaver ; mais depuis deux ans, ces utilitaires se perfectionnent et élargissent leur rayon d'action : ils s'imposent de plus en plus aux "travailleurs du savoir" qui collectent des informations. Voyons donc leurs fonctions clés par l'étude d'un échantillon international de logiciels : américains (Surfsaver et Furl), allemand (ContentSaver), israélien (NetSnippets), anglais (Onfolio), russe (MyFavorites)... et français (Egems et Syfacil). Dans cette série, notons que Furl reste malheureusement le seul à fonctionner non seulement sous Windows, mais aussi sous Mac.



### Capter, organiser, commenter exporter ...

L' internaute rencontrant une ressource intéressante en cours de navigation choisit généralement de l'imprimer, et/ou d'enregistrer la page web au format html, ou d'effectuer un copier-coller de la donnée utile dans un traitement de texte. En effet, le simple ajout de l'URL dans les favoris ne suffit pas à garantir la disponibilité future de la page. Mais ces solutions présentent des inconvénients de taille, notamment l'absence d'"informations sur l'information" (méta-données), la mauvaise copie de certains éléments comme les tableaux, et le repérage ultérieur aléatoire. Nos outils de capture simplifient toutes ces procédures.

## En s'intégrant aux navigateurs

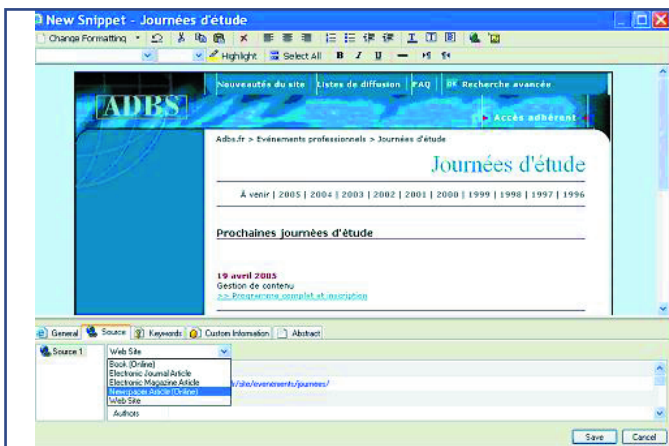
La plupart, tels Surfsaver ou MyFavorites ne supportent malheureusement qu'Internet Explorer, d'autres sont plus ouverts comme NetSnippets, Onfolio ou Content Saver. Une fois le logiciel installé, une nouvelle icône apparaît : un clic suffit alors à ouvrir un cadre à gauche de l'écran, pour gérer et visualiser ses dossiers, le principe général restant un classement arborescent à base de catégories et de sous-catégories. Cette fenêtre dispose d'un menu intégré plus ou moins développé, mais une barre d'outils spécifique est souvent disponible pour garder l'alternative plein écran.

## En aspirant le contenu souhaité

Il peut s'agir d'un élément d'information (phrase, paragraphe, image, lien, fichier pdf, etc.) ou d'une page entière. La capture s'effectue soit par un clic droit, soit via le menu ou la barre d'outils, soit par simple glisser-déposer. Notons que certains outils se rapprochent d'un "aspirateur" web en proposant de capturer non seulement la page web à l'écran, mais également les pages liées, voire l'ensemble du site (Content Saver, Onfolio, Surfsaver). D'autres, tels NetSnippets et Content Saver, ont un petit éditeur intégré pour modifier la page, l'annoter, surligner des passages, et permettent d'enregistrer des notes et des idées en renseignant une fiche vierge.

## En permettant de renseigner des champs d'indexation

Le titre, l'URL, la date voire l'heure de la capture sont pré-remplis. Des méta-données plus ou moins complètes sont proposées en fonction des outils et des versions des logiciels (au moins pour les mots-clés et le commentaire). Net Snippets ou Onfolio vont jusqu'à proposer la création de champs supplémentaires pour mieux qualifier l'information.



## En permettant une navigation et une recherche aisées dans les contenus stockés.

Il suffit de cliquer sur un dossier ou un élément d'information pour voir son contenu s'afficher, comme dans Surfsaver. Notons que, malgré de nettes améliorations, cet agent, certes peu onéreux (19,95 \$), garde une ergonomie moyenne et des fonctionnalités mal abouties. Certains logiciels permettent un repérage visuel supplémentaire pour les dossiers et les éléments d'information à base de drapeaux de couleurs, ou d'icônes spécifiques.



Les fonctions de recherche ne sont pas oubliées, avec des moteurs qui indexent le contenu des dossiers, et permettent des recherches plein texte et sur champs avec des fonctionnalités d'un niveau souvent correct.

## En générant des rapports

Si l'export est globalement géré par tous les logiciels de notre échantillon (les modules et formats proposés faisant alors la différence), Net Snippets apparaît l'un des outils les plus performants pour la diversité des captures possibles et l'ergonomie générale, mais aussi pour la sophistication extrême des rapports générés. Cette dernière fonction n'est accessible qu'avec la version professionnelle à 129,95 \$, mais signalons que Net Snippets est le seul proposant une version gratuite (usage personnel et professionnel) non limitée dans le temps et malgré tout très fonctionnelle.

## Des outils qui surfent sur la vague RSS

RSS, vous connaissez ? Ce format de fichier dérivé de XML, qui permet de récupérer le nouveau contenu d'un site (articles, actualités) pour peu que l'on dispose d'un lecteur capable d'interpréter ces fichiers. Depuis peu, ces "agrégateurs" d'un nouveau genre, se téléchargent certes en tant qu'applications indépendantes, mais jouent de plus en plus la carte de l'intégration : avec les navigateurs web, les gestionnaires de courrier électronique ou les agents d'alerte pour la veille. Les outils de capture investissent également ce créneau porteur, tels Onfolio et Content Saver. Onfolio automatise même la publication de documents sur un blog ou un serveur web : Si la version personnelle à 29,45 \$ intègre le lecteur RSS et les fonctions les plus simples de capture et

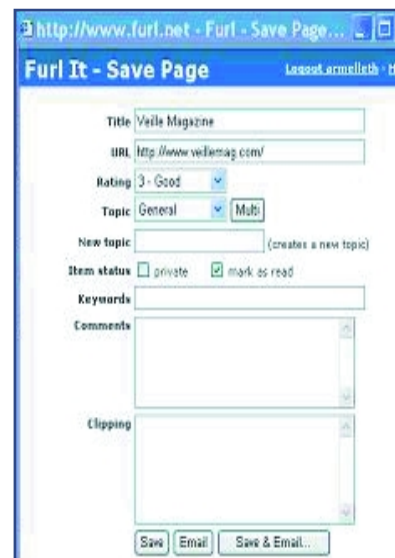


d'organisation de l'information, il faut néanmoins passer à la version professionnelle à 99,95 \$ pour aller plus loin (aspiration d'un site, recherche avancée, partage des collections, publication automatisée, , etc.). Content Saver est proposé en une seule version à 39,90\$.

## L'alternative "services web"

Les "services web" permettent de travailler en ligne sans télécharger de logiciel, contrairement aux outils vus précédemment. Furl, conçu par l'américain Michael Giles et racheté depuis par Looksmart, permet ainsi depuis début 2004 de se constituer gratuitement en ligne une base d'archives de pages web, organisée en dossiers, et exportable. Celle-ci est alimentée au fil des navigations d'un simple clic sur un bouton inséré sur le navigateur ("bookmarket"), ou via une barre d'outils spécifique à télécharger. Une fiche descriptive pré-renseignée est alors proposée, le champs "clipping" n'étant rempli que si on a préalablement sélectionné une portion de texte sur la page.

L'alternative "services web" facilite les notions de partage et de "navigation sociale", que Furl exploite à fond : A moins que la page n'ait un statut privé, elle apparaîtra sur la page d'accueil de Furl dans les "latest entries" avec les entrées les plus populaires du jour. Par ailleurs, l'utilisateur peut aussi renvoyer les internautes vers sa page web d'archives "publiques", constituer un fil RSS des derniers ajouts, prévoir un envoi quotidien de mail. Le système de notation permet de consulter les



entrées les plus populaires et de sophistication le tri de pertinence lors des recherches sur les bases privées et publiques.

Egalement gratuit jusqu'à 3 Mo de stockage, le récent service français Syfacil permet d'organiser en ligne du contenu web et intègre un lecteur RSS. Il joue aussi la carte de la gestion de fichiers de tous types (prévoir alors du haut débit pour des délais corrects !).

## De la capture web au knowledge management...

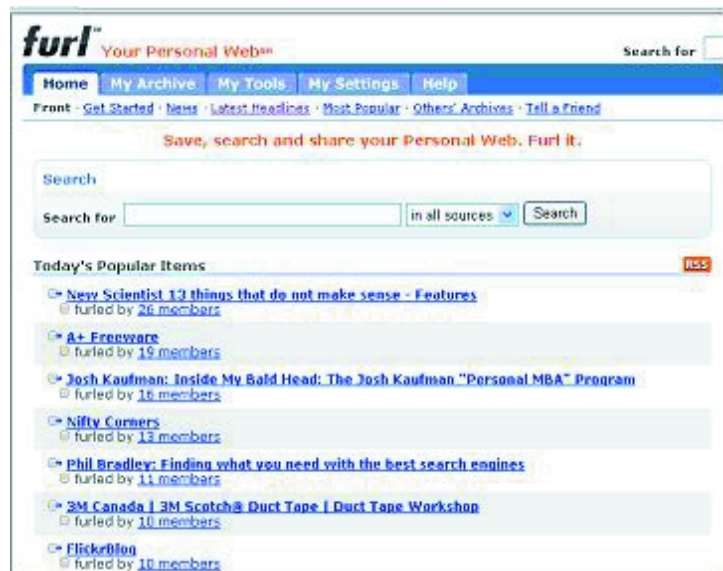
Issus de l'internet, ces outils affichent aujourd'hui de nouvelles ambitions : il ne s'agit plus seulement de capturer et d'organiser du contenu web, mais de gérer tous types de données électroniques présentes sur le poste de travail d'un usager. Dans le cadre d'un projet particulier, on peut ainsi rassembler des documents dans différents formats dispersés sur son ordinateur (pages web, notes, fichiers pdf ou word, mails, graphiques, tableaux ou dessins, photos, textes, images, sons, clips vidéos...) et les agréger dans son logiciel.

Des outils comme Net Snippets ou Egems permettent pour les formats non reconnus directement de capturer l'écran (en le redimensionnant éventuellement) et le contenu du presse-papier, et de les organiser dans le même système arborescent. L'astuce : une barre d'outils toujours présente dans la barre des tâches de windows (Net Snippets) ou une fenêtre flottante qui se réduit à un logo (Egems).

### Quant au partage d'informations...

...il est souvent prévu grâce à plusieurs formats d'exports, soit propriétaires pour s'échanger des documents, des dossiers ou des bases de données complètes avec des collègues disposant aussi du logiciel, soit plus communs ou plus ouverts.

Pour un travail en réseau, Onfolio propose d'enregistrer une "collection" sur un répertoire partagé du ser-



Les bookmarklets sont des programmes contenus dans des liens, c'est à dire des éléments de code java qui se mettent dans les favoris comme des URL classiques, mais qui déclenchent quand on les appelle une action particulière, souvent l'ouverture d'une fenêtre "pop-up". L'intérêt est qu'ils ne nécessitent aucun téléchargement.

veur, plusieurs personnes pouvant alors y contribuer. Un fil RSS peut même y être associé !

Pour se glisser dans la catégorie des outils de knowledge management, il ne reste finalement plus qu'à proposer une version serveur. C'est le pari réalisé par la société française Alogic (lire l'article FOCUS SUR...) qui développe,

rappelons-le "l'atelier logiciel" Aperto Libro. A côté de eGems Collector Pro (65,92 € la licence), on trouve aussi Gems Entreprise, développé pour être une plate-forme indépendante accessible grâce à un navigateur standard. Dotés du client Egems Entreprise Pro, des utilisateurs peuvent simultanément récupérer et partager tous types de documents (indépendamment du type de contenu ou des limites géographiques). NetSnippets dispose également une version serveur, et Syfacil propose sa technologie aux entreprises.

Sans aller jusque là, nous conseillons à nos lecteurs d'opter au moins pour un assistant personnel et de découvrir ainsi un nouveau confort de travail : alimenter simplement une base de connaissances personnelle toujours à portée de clic.

Armelle Thomas

### Produits cités

Cogitum co-citer (Cogitum)	www.cogitum.com
Content Saver (Macropool)	www.macropool.com
Egems (Alogic)	www.alogic.fr
Furl (Looksmart)	www.furl.net ou www.looksmart.com
My Favorites (Arcadia)	www.bookmark-manager.net
Net Snippets (Net Snippets)	www.netsnippets.com
Onfolio (Onfolio)	www.onfolio.com
Surfsaver (AskSam Systems)	www.surfsaver.com
Syfacil (Com'onSoft)	www.syfacil.com

# Alogic

## La conjonction du renseignement et de l'informatique



Alogic est la fin et le début d'une aventure qui a commencé en 1968. Cette business unit de COFRAMI Inforama rassemble tout ce que le groupe sait faire dans le domaine du renseignement. Elle est construite autour de deux lignes de produit, Aperto Libro et E-gems. Aperto Libro a pour objet de produire des info-gemmes. E-gems rend ces info-gemmes, consommables, compréhensibles, diffusables, exploitables pour les besoins des utilisateurs.

**“...nous avons réussi à créer des logiciels accompagnant le mouvement évolutif rapide des traitements de l'information.”**



Alogic est l'aboutissement d'une aventure qui a commencé en 1968, explique Robert Guillaumot, Président de Alogic, Vice-président de COFRAMI. "J'ai alors créé une des premières SSII françaises, en même temps que nous fondions

une chambre syndicale qui s'appelle le S y n t e c Informatique." Cette aventure venait d'une conjonction qui

va se révéler payante ultérieurement entre un savoir-faire qui est celui de la production de renseignements : - Robert Guillaumot a été dans le domaine militaire au 2ème Bureau de l'Armée, puis un des fondateurs de SVP-, avec l'in-

formatique IBM."En 1999, nous décidions de rassembler tout notre savoir-faire dans le domaine du renseignement, mode opératoire, outils logiciels, et ce fut la création d'Alogic, totalement dédiée à l'acquisition, le traitement, l'exploitation du contenu et du sens." Les effectifs de Alogic sont des effectifs de R&D, 11 personnes, mais la société utilise toutes les ressources du groupe COFRAMI, 2000 personnes. Une des spécialités de COFRAMI a toujours été les systèmes de mise à disposition d'information dans le secteur militaro-industriel, l'interprétation de bases de données, l'acquisition de données par les satellites, jusqu'à la production d'informations et de connaissances pour le management des entreprises.

"Cette renaissance en 1999 marque un pas considérable. Avec la compréhension méthodologique de ce que c'est que partir d'une matière première comme l'information pour aboutir à un produit manufacturé de renseignement et d'information pertinente, nous avons réussi à créer des logiciels accompagnant le mouvement évolutif rapide des traitements de l'information."

• Robert Guillaumot

## L'analogie de l'industrie pétrolière



Pour Robert Guillaumot, l'analogie de l'industrie pétrolière est particulièrement éclairante : identification des possibilités de gisement, premier sondage pour vérifier la richesse du contenu, décision d'exploiter, exploitation de production, stockage. Toute cette partie préliminaire nécessite une ligne d'outils dont dispose Alogic. Ensuite, pour l'information brute, il y a une opération de stockage, d'accession, d'extraction, de raffinage, pour fournir au fur et à mesure des contenus différents qui peuvent être à nouveau exploités. Cela correspond aux colonnes de raffinage et de distillation. "Nous produisons des info-gemmes, des bijoux, des diamants d'information qui apparaissent à la fin de ce processus de raffinage et de distillation."

Dans les deux lignes de produit qui existent chez Alogic, Aperto Libro a pour objet de produire des info-gemmes. C'est tout le traitement en amont. C'est une colonne de distillation qui aboutit avec les tuyaux qu'il faut, à pouvoir donner à quelqu'un la possibilité d'exploiter un contenu auquel il accède. Quand on a ces diamants, il va falloir les traiter de telle manière qu'ils puissent être consommables, compréhensibles, diffusables, exploitables pour les besoins des utilisateurs. C'est l'univers de E-gems. Les deux technologies sont des technologies de pointe. E-gems est brevetée aux Etats Unis. C'est un produit reconnu et apprécié pour son excellence. Principale référence : 750000 élèves de Californie sont équipés de e-gems pour garder, classer, utiliser ultérieurement leurs bijoux. La deuxième référence est Pearson qui a choisi e-gem pour en faire cadeau à toute personne qui achète un ouvrage édité par Pearson.

### Collector, produit grand public

"Je suis tombé amoureux du produit qui n'a pas d'équivalent, pour gérer l'information. Il est plus pratique que google desktop qui, induit un traitement a posteriori. C'est instantané, pas de manipulation contraignante, je l'utilise même pour l'information des e-mails que je reçois," confie Charles Nouÿrit chef produits eGems.

La ligne E-gems est composée de trois produits, Collector et Collector Pro, Communicator et Server.

Collector, produit grand public, s'adresse à toute personne qui a besoin de stocker, archiver, traiter de l'information. E-gem, petit diamant pas intrusif, va collecter exactement ce qui est sélectionné et relever automatiquement les bibliographies. Il a des répertoires et sous-répertoires

comme sous Windows, où peuvent être stockées les pépites d'information. Il fonctionne sous PDF, sous Excel, Internet, n'importe quel browser. A l'intérieur, un moteur de recherche permet d'effectuer une deuxième recherche.

Communicator est la même base que Collector auquel on a rajouté tous les outils nécessaires pour avoir une base sécurisée et privée

Vous entrez dans votre communauté les gems que vous créez, vous pouvez les partager si vous le souhaitez; vous avez tous les outils de communication, instant messaging, téléphonie sur IP, visioconférence; la transmission de l'information se fait en peer to peer.

Vous pouvez partager des documents en plus des gems. Tout fonctionne en drag and drop,

glisser-déposer. Tout est sécurisé, crypté. Un système garantit que le système envoyé est l'équivalent de l'original. Tout fonctionne par système d'invitation-validation, pas d'intrusion. Cela permet d'augmenter la productivité des communautés de travail, et c'est surtout pratique pour ceux qui sont délocalisés. Le système de prise de contrôle de PC à distance est particulièrement utile aux administrateurs réseau. La version Serveur s'adresse à une très large communauté, pour les très gros groupes et accompagne idéalement l'accord récent de Google avec les grandes bibliothèques américaines. Il permet à des centaines de milliers de personnes, de se connecter avec un login. Collector et Collector pro, sont diffusés uniquement sur Internet.

**Mireille Boris**