

Concevoir une encyclopédie multimédia

« Le but d'une Encyclopédie est de rassembler les connaissances éparses sur la surface de la terre, d'en exposer le système général aux hommes avec qui nous vivons, et de le transmettre aux hommes qui viendront après nous ; afin que les travaux des siècles passés n'aient pas été des travaux inutiles pour les siècles qui succéderont ; que nos neveux, devenant plus instruits, deviennent en même temps plus vertueux et plus heureux ».

Denis Diderot (article « Encyclopédie », *L'Encyclopédie*)

Aujourd'hui, les encyclopédies deviennent multimédias. D'abord sur cédérom, puis sur DVD-Rom, actualisées *via* Internet voire intégralement transférées sur la Toile, elles s'enrichissent d'images animées, de sons. Surtout, elles développent des possibilités de recherche sans commune mesure avec les index et tables des encyclopédies imprimées.

Comment se créent ces nouveaux objets, et quelles perspectives font-ils apparaître ?

Le *BBF* a enquêté auprès des réalisateurs d'un de ces produits, *l'Encyclopédie Hachette Multimédia 2000 (EHM 2000)*.

Neil Minkley

Hachette-multimédia
neil.minkley@hachette-multimedia

Jerôme Bessac

Hachette-multimédia
jerome-bessac@hachette-multimedia.fr

Philippe Alcouffe

Plurimédia
philippe.alcouffe@hachette-multimedia.fr

BBF : *Neil Minkley, vous êtes directeur du département Référence de Hachette Multimédia, et, à ce titre, responsable de l'organisation conceptrice et productrice de l'EHM 2000. Comment votre département était-il conçu pour l'élaboration de cette encyclopédie ?*

Neil Minkley : Pour la réalisation de l'EHM 2000, le département est structuré en trois pôles correspondant aux trois étapes logiques de la réalisation : une équipe éditoriale, une équipe d'informatique éditoriale, une équipe de développement logiciel. L'équipe éditoriale, constituée de per-

sonnes originaires de l'édition, des lettres et sciences humaines, de la documentation et de l'audiovisuel, constitue les contenus, élément central qui pourra être décliné ensuite sur de multiples supports. Faisant appel à des auteurs extérieurs, des rédacteurs-correcteurs missionnés, des documentalistes iconographes extérieurs, et à des entreprises d'infographie, cette équipe entretient beaucoup de relations avec d'autres acteurs.

L'équipe d'informatique éditoriale, qu'on devrait plutôt appeler l'ingénierie des données, traite la mise en forme des contenus dans l'ensemble des bases de données, à partir des indexations thématiques de l'équipe éditoriale, mais aussi sur des développements spécifiques d'indexation

* Cet article est le fruit d'entretiens réalisés par Bertrand Calenge entre février et juin 2000.

CONCEVOIR UNE ENCYCLOPÉDIE MULTIMÉDIA

L'Encyclopédie Hachette Multimédia 2000

Extrait de la notice de présentation
(<http://www.encyclopedie.hachette-multimedia.fr/>)

Grâce au nouveau moteur de recherche intelligent, l'*Encyclopédie Hachette Multimédia 2000* comprend toutes vos questions et tolère même vos erreurs. Posez la question comme elle vous vient à l'esprit, la liste des réponses pertinentes apparaît, et vous choisissez celle qui vous intéresse.

À chacune de vos questions, l'*Encyclopédie* vous donne également la liste des sites Internet pertinents. Une fois sur le Web, vérifiez vos informations : un clic sur un mot suffit pour ouvrir l'article de l'*Encyclopédie* correspondant. Une révolution du savoir !

23,3 millions de mots pour un fonds 100 % renouvelé.

- 17 000 médias spectaculaires, dont :
- plus de 60 panoramiques 2D et 3D,
- plus de 200 vidéos et animations
- plus de 11 500 photos, dessins, documents et tableaux de données,
- plus de 1 800 cartes géopolitiques et thématiques interactives, diagrammes, fiches pays,
- plus de 3 000 documents sonores.

Une chronologie interactive avec plus de 1 100 événements importants de l'histoire de l'humanité.

Un atlas mondial complet (30 000 faits et chiffres actualisés, 800 cartes, 200 fiches pays...).

Un atlas historique.
Un quiz de 1 200 questions.

Hypertexte total pour naviguer d'une définition à une autre, sur les 23,3 millions de mots de l'*Encyclopédie*, mais aussi sur les sites Internet. Accès direct à Internet, et à plus de 2 500 sites sélectionnés en fonction du sujet traité, avec possibilité de mise à jour mensuelle des articles et chronologie de chaque mois (le rappel de tous les événements du mois écoulé).

Une extension d'aide à la rédaction en français (dictionnaire de langue française : définitions, synonymes, conjugaisons) et en anglais (dictionnaire Hachette-Oxford), copiable sur le disque dur.

Une boîte à outils pour effectuer toutes sortes de manœuvres très simplement : marquer les pages, surligner, rechercher un mot dans un texte, prendre des notes, imprimer, exporter, enrichir le contenu de votre encyclopédie par des notes...

Une organisation optimale des recherches avec des dossiers personnels.

3 cédéroms Mac & PC ou 1 DVD-Rom.

automatique. À un moment t de la procédure d'édition, on déclenche un « *gel éditorial* » : les contenus sont figés, et l'équipe d'informatique éditoriale (des ingénieurs et des spécialistes d'informatique linguistique) fait tourner des procédures qui créent tables et index, avant de transférer l'ensemble à l'équipe du développement. Cette dernière, constituée d'ingénieurs en informatique, développe le logiciel d'application et traite les données pour exploitation par ce logiciel. Cette organisation, effective pour l'*EHM 2000*, tend à évoluer : si les équipes éditoriales sont spécifiques pour chaque produit, l'informatique éditoriale et le développement logiciel deviennent transversaux : par exemple, en 1999, l'équipe du développement a traité l'*Encyclopédie des vins*, pour le compte du département Vie pratique.

Dans le triptyque que j'évoque, les activités de chaque équipe ne sont

pas étanches, particulièrement entre les éditeurs et les informaticiens éditoriaux : les outils de recherche élaborés doivent répondre à une pertinence éditoriale, l'indexation des contenus par l'équipe éditoriale doit être cohérente avec les procédures de l'informatique éditoriale. Enfin, pour achever le tour d'horizon des compétences humaines, ajoutons le recours à des spécialistes extérieurs en édition ou documentation, sur contrats à durée déterminée, pour réaliser les opérations d'indexation : par exemple, pour passer de l'encyclopédie 1999 à l'encyclopédie 2000, il a fallu indexer 7 000 nouveaux médias en trois mois.

BBF : Cette organisation permet de produire une encyclopédie sous deux versions (cédérom et DVD-Rom), à laquelle s'ajoute un lien avec un site Web. Avant d'aborder la question de l'élaboration de cette

encyclopédie, sur quelles fonctionnalités souhaitez-vous mettre l'accent ?

N.M. : Outre la masse d'informations contenues (40 000 articles indexés en texte intégral, par exemple), je crois que les principales avancées tiennent pour l'utilisateur dans le travail d'interprétation de ses requêtes, et dans l'hybridation mise en œuvre entre le cédérom - et ses fonctionnalités - et le Web. Sur le premier point, le très gros travail fourni sur l'indexation, humaine et automatique, permet par exemple d'accéder à un article sur le Lusitania en réponse à une requête évoquant un « paquebot torpillé durant la première guerre mondiale », de découvrir qui était roi de France en 1645, d'avoir la liste des chefs d'État en exercice pendant la guerre du Golfe, ou d'obtenir sans ambiguïté les notices de Vadim et de Bardot par la recherche de *Et Dieu créa la femme...* Notre souci n'est pas tant d'atteindre dans la réponse une performance brute (délai d'obtention d'une liste de résultats au sein desquelles se trouve la bonne réponse), qu'une performance nette (délai d'obtention de cette bonne réponse). Sur le second point, l'*EHM 2000* permet bien sûr d'accéder à un répertoire de sites Internet mis à jour sur le site Hachette, mais aussi d'utiliser les fonctionnalités du cédérom (accès aux définitions de chaque terme par exemple, aux cartes géographiques, aux articles encyclopédiques) sur toutes les pages Web, y compris celles de sites découverts par l'utilisateur. L'*EHM 2000* devient non seulement une ressource de données, mais un outil de recherche.

BBF : Venons-en au contenu de cette encyclopédie. Jérôme Bessac, vous êtes directeur de l'équipe Édition multimédia : comment se constituent les contenus de l'encyclopédie, et quels changements l'introduction de la dimension multimédia dans le travail des encyclopédistes a-t-elle apportés ?

Jérôme Bessac : Cela fait bientôt 35 ans que je suis entré dans le milieu de l'édition des dictionnaires et encyclopédies, et l'édition électronique n'est apparue dans ma carrière que depuis 6 ans environ. Dans l'édition imprimée, la chaîne de fabrication s'est normalisée au cours des siècles ; l'introduction du multimédia a bouleversé en quelques années les organisations ainsi longuement mises en place. Chaque éditeur met en œuvre ses méthodes et ses outils par un système d'essais et d'erreurs, et il faut bien garder à l'esprit qu'il n'y a pas de modèle actuellement. En outre, ce que je peux exposer est précaire, car les avancées technologiques se précipitent, et nous allons être obligés de modifier très vite nos méthodes de travail, voire de pensée actuelles. Aujourd'hui, nous restons imprégnés par le modèle de l'édition imprimée, et une partie importante de nos ouvrages de référence électroniques est une transposition d'ouvrages créés pour le papier, ne serait-ce que pour récupérer le très long investissement consenti pour constituer les contenus de ces éditions imprimées.

Dans le cadre d'une encyclopédie papier, on se trouvait contraint par la matérialité de l'imprimé. Les nécessités de la gestion et de la commercialisation limitaient le nombre des volumes. En revanche, le cédérom ouvre des horizons beaucoup plus larges. On s'est aperçu que le simple transfert des contenus éditoriaux d'une encyclopédie imprimée laissait le support vide aux deux tiers. Dans un premier temps, cet espace libre a fourni l'opportunité d'enrichir l'encyclopédie d'éléments que les ouvrages papier ne pouvaient pas apporter soit par manque de place (développements textuels complémentaires, illustrations, photographies), soit par leur nature même de livres imprimés (sons, vidéos, animations, etc.). Puis on a réfléchi à la récupération d'autres données textuelles ou iconographiques dont nous disposons par ailleurs : dictionnaire de langue, dic-

tionnaire bilingue Hachette-Oxford, Atlas, etc. Le problème des droits d'auteur sur la vidéo ou la musique entraîne alors des investissements supplémentaires : sous le même nom d'encyclopédie, on passe d'une conception où un ensemble de droits forfaitaires sont prédéfinis à un produit sur lequel se greffent de multiples objets, non prévus initialement.

BBF : *Ces modifications entraînent-elles une mutation de la notion d'encyclopédie ?*

J. B. : Cessant d'être un ouvrage clos, l'encyclopédie devient une « *œuvre ouverte* », un portail d'accès à d'autres sources de connaissances et d'informations. Prenez l'article « cancer » : écrit il y a 5 ou 6 ans par une sommité, il représente une synthèse importante qui fait date. Mais si vous êtes spécialiste, ou si simplement vous voulez en savoir davantage, c'est tout à fait insuffisant : il convient donc d'être lié à un site ou à une autre base textuelle sur les cancers.

L'encyclopédie conçue comme projet tendant à « *rassembler les connaissances éparses sur la surface de la terre* » et à « *en exposer le système général aux hommes avec qui nous vivons* » me paraît être devenu aujourd'hui - édition électronique ou non - un projet irréaliste. Buffon, en son temps, était sans aucun doute parfaitement capable de comprendre l'ensemble des travaux mathématiques de D'Alembert ; mais je doute que beaucoup de prix Nobel de biologie soient aujourd'hui en mesure de comprendre les travaux d'un médaillé Fields. Quant à « *l'honnête homme* » de l'an 2000, il est probable qu'il ne sera en mesure de comprendre ni l'un ni l'autre. Dans des domaines de plus en plus nombreux, l'enjeu de l'encyclopédie est ainsi devenu de fournir à beaucoup d'interrogations des réponses honnêtes simplement à un certain niveau de connaissance. Ce qui amène à s'interroger sur la nature non pas de « *l'honnête homme* » mais du « *grand*

public » ; pour moi, ce dernier est constitué d'une multitude de « *micro spécialistes* » : tel chercheur du CNRS en physique des particules trouvera satisfaisant pour sa curiosité tel article sur l'architecture précolombienne qu'un spécialiste de la culture aztèque trouvera sans doute incomplet. L'important pour l'éditeur me semble alors : d'une part, que le spécialiste des Aztèques ne trouve aucune erreur ni aucun oubli fondamental dans cet article ; d'autre part, qu'il puisse lui-même trouver dans le même ouvrage un article qui lui permette de comprendre - s'il en a la curiosité - ce qu'on entend par le charme d'un *quark*, sans que le chercheur au CNRS puisse trouver à son tour que l'article concerné contienne des approximations erronées. Enfin, troisièmement, que l'un et l'autre, s'ils en ont le désir, trouvent dans l'encyclopédie le moyen d'accéder, notamment par Internet, à des sources d'informations complémentaires validées. L'éditeur d'encyclopédie aura ainsi finalement un travail de choix et d'orientation du lecteur désireux d'en savoir plus vers des sites analysés, expertisés, validés, que ces sites soient ou non son œuvre. En réalisant cet aiguillage, cette carte routière, il rejoindra le travail de l'éditeur d'une revue scientifique, qui soumet avant publication les articles à un comité d'experts. Bien sûr, dans le cadre de sa propédeutique, l'éditeur d'encyclopédie continuera d'être le « *généraliste en curiosités spécifiques* » et de proposer en première lecture un article qu'il aura rédigé ou fait rédiger, et qui devront être accessibles au plus grand nombre.

BBF : *Malgré cette large ouverture sur les ressources du réseau, le contenu de l'encyclopédie reste encore aujourd'hui essentiellement un produit réalisé par les encyclopédistes ou à leur initiative. Comment les informations sont-elles structurées ?*

J.B. : Historiquement, nous avons été les premiers en France à proposer

une encyclopédie sur support électronique - Axis. À cette époque, nous avons travaillé et adapté un thésaurus utilisé dans l'éducation, Mobis. Progressivement, nous avons évolué vers une approche d'indexation des données autour de trois concepts : des thèmes, des dates, des lieux, avec diverses déclinaisons de ces concepts : par exemple, on signale s'il s'agit d'un nom commun ou d'un nom propre. Ces éléments sont notés pour chaque entrée et, par le travail de l'informatique éditoriale, des tables de concepts sont constituées. Un travail d'analyse lexicale et de liens sémantiques est également réalisé à ce stade : Austerlitz doit renvoyer au thème Bataille, l'article sur Didon et Énée doit comporter un lien vers l'article sur l'opéra baroque, une recherche sur Henri IV doit, selon son contexte, renvoyer à l'empereur germanique ou au roi de France, etc. Ces tâches

sont réalisées au niveau éditorial et informatique pour chaque article, et souvent aussi pour les différentes sous-sections de l'article. Mais, sur ce sujet, je laisse la parole à Philippe Alcouffe, qui a été directeur de l'informatique éditoriale.

Philippe Alcouffe : On imagine volontiers que l'informatique éditoriale est un outil qui permet la simple transcription sur support électronique des informations rédigées par un éditeur. En fait, c'est faux : l'outil informatique traite et transforme les différents produits éditoriaux en objets nouveaux qu'il manipule de différentes manières. L'éditeur rédige et l'informaticien isole, extrait et relie les concepts présents jusque-là sous une simple forme rédactionnelle. Comment cela se passe-t-il ? Au départ, tout texte rédigé est naturellement structuré. Grâce à XML, on peut

baliser l'information indépendamment de sa présentation, selon différents niveaux d'abstraction : texte, typologie, titre... Pour un certain nombre de concepts de base, tels le lieu de naissance, la date de naissance, les noms d'œuvres, les références, l'identification est aisée. Des étiqueteurs syntaxiques construisent des tables mettant certains objets en relation (par exemple l'auteur et l'œuvre). Sur le texte structuré lui-même, des analyses permettent d'aller plus loin : lorsqu'un texte dit que Untel est « né à Paris, mort id. », un algorithme permet de traiter cet *id.* comme étant le lieu Paris. Des expressions comme « accéder au pouvoir » sont traitées pour y associer une date, etc. À ce niveau, le travail informatique permet de déceler des incohérences dans les contenus : décalages de dates pour un même événement entre un article et un autre, ortho-

graphes différentes pour un même nom, succession des chefs d'État d'après les chronologies (on avait ainsi découvert qu'il nous manquait un président des États-Unis !), etc. Ces divers problèmes sont retransmis à l'équipe éditoriale, pour recherches, rectifications, mise en cohérence des contenus.

Revenons aux tables de relations. Ces tables sont établies à partir de listes de concepts ordonnés en trois arbres : thèmes, lieux et dates. Seuls les thèmes sont obligatoires, lieux et dates dépendant évidemment du thème ou du contenu de conceptualisation. Par exemple, Henri IV est *Roi*, né et mort à telles *dates*, dans tels *lieux*. Les thèmes viennent souvent de Mobis, mais on a dû travailler plus loin : certains thèmes sont transversaux (par exemple « l'impressionnisme », qui touche un grand nombre d'acteurs et d'événements), d'autres sont dits ontologiques (l'impressionnisme est un « mouvement pictural »). Au total, l'ensemble des tables de relations liant des concepts forme un réseau sémantique, qui est extrait du texte - et des images, sons, etc. -, mais en est distinct. Il devient une base de données relationnelles, à partir de laquelle on peut imaginer créer de multiples produits éditoriaux : associer la date d'une œuvre picturale avec des événements simultanés, construire la liste des rois de France, partir des *Nymphéas* pour découvrir l'impressionnisme, voire un panorama de la peinture, etc.

Tout ne se fait pas absolument par les analyseurs et les algorithmes : l'équipe éditoriale est amenée à établir des renvois, lorsque le texte structuré ne permet pas de les établir automatiquement. Mais le réseau sémantique produit finalement a bien sa cohérence et son existence propres. Tout ce travail offre des possibilités très intéressantes pour l'utilisateur, d'une part parce qu'il permet des interrogations « presque » en langage naturel, d'autre part parce qu'il ouvre des pistes pour une lecture élargie. Sur le premier point, nous

avons développé un interpréteur des requêtes en langage naturel qui interroge le réseau conceptuel et, en cas de besoin, le texte intégral ; une question comme « *Qui a eu le prix Albert Londres en 1980 ?* » est traitée de la façon suivante : le mot « qui » renvoie à des articles de personnes, « prix Albert Londres » est recherché en texte intégral et « 1980 » renvoie aux concepts de date.

Deuxième point, les tables de relations offrent ensuite des pistes supplémentaires : à la suite d'un article sur un homme d'État par exemple, on peut proposer ses prédécesseurs et successeurs, mais aussi la liste des chefs d'État en fonction au moment de sa mort, etc. Ce souci de l'utilisateur va jusqu'aux fonctionnalités du dictionnaire. Comme toutes les encyclopédies multimédias, l'*EHM 2000* propose la définition du mot en plein texte lorsque l'utilisateur a cliqué dessus ; mais à la différence des autres, elle offre un dictionnaire de flexions : sous le mot « avions », vous aurez bien sûr la définition de l'aéroplane, mais aussi la notice du verbe avoir. Comme on peut le constater, cette organisation en réseau conceptuel révolutionne l'approche encyclopédique. Notez que toutes les possibilités ne sont pas nécessairement exploitées : l'informatique éditoriale propose des relations qu'elle sait mettre en œuvre, et c'est l'équipe éditoriale qui juge de la pertinence de ces relations, en fonction de leur intérêt pour l'utilisateur.

BBF : *Cet utilisateur, justement, comment le voyez-vous réagir ? Comment le sentez-vous évoluer et, partant, comment une encyclopédie va-t-elle évoluer demain ?*

J.B. : Cette question recouvre d'abord deux interrogations distinctes : peut-on définir le public d'une encyclopédie ? Et comment ce public réagit-il aux innovations qu'on lui propose ?

Il faut avouer une chose : on ne sait pas vraiment quel est le désir ini-

tial, le besoin, la motivation qui pousse quelqu'un à acheter une encyclopédie. L'apparition des encyclopédies sur support électronique a eu au moins le mérite de faire taire définitivement les sarcasmes sur « *les mètres de reliures en skivertex destinés à épater les voisins* ». Il est clair que la « *valeur de signe* » attachée à la possession d'un boîtier de cédérom est quasi nulle. Or, puisqu'il se vend aujourd'hui chaque année infiniment plus d'encyclopédies sur cédérom qu'on en a jamais vendu en collections reliées, on est bien amené à en conclure d'une part que les motivations d'achat sont réelles et qu'elles s'attachent davantage à une valeur d'usage qu'à une valeur de signe.

À partir de là, il est difficile de préciser la nature du public que nous touchons. Dans les faits, même la différence traditionnelle entre « enfants 13-17 ans » et « grand public adulte » tend à s'estomper. La plupart des jeunes bacheliers en savent beaucoup plus sur la reproduction des grenouilles, l'intégrale d'une fonction continue ou les grandes zones agricoles de la Chine que bien des adultes ! Par ailleurs, il apparaît évident que ce grand public adulte n'est pas tant constitué de « Français moyens » que d'une multitude de spécialistes experts chacun sur des sujets particuliers. Il suffit de voir l'abondance des courriers - électroniques notamment ; c'est là un changement majeur d'attitude par rapport aux encyclopédies papier, qui provoquaient moins de réactions : tel citoyen d'une ville critique la date qui est donnée pour la construction du transept de la cathédrale de sa ville, tel autre signale que la citation du titre d'un ouvrage de botanique du XVI^e siècle n'est pas complète, etc.

Pour reprendre encore ce que nous disions tout à l'heure, l'éditeur est amené à penser de plus en plus non plus en termes de niveaux de publics, mais de niveaux d'intelligibilité, et son travail consiste dès lors à

ménager des accès à chacun de ces niveaux : à l'intérieur même de l'ouvrage, un premier niveau d'exposition du sujet ménage des liens vers un second niveau de développement, avant d'ouvrir vers l'extérieur, vers ce troisième niveau qu'est le site Internet, spécialisé et validé. En définitive, ce vers quoi nous tendons est de proposer une arborescence complète de liens éditoriaux ou hypertextuels, en établissant des filtres selon les niveaux.

Alors, comment allons-nous évoluer ? C'est très difficile à prédire. L'évolution du marché et des technologies est très rapide, et nous n'en sommes certainement qu'à la phrase initiale de ces nouveaux produits. Bien sûr, tout objet nouveau marche bien au départ, parce qu'il existe toujours quelque part une clientèle à la fois fortunée et férue de modernisme. Mais, dans un deuxième

temps, il faut que le grand public prenne le relais (baladeurs, téléphone portable...), sinon c'est l'échec (CD-vidéo, laser disc...).

Aujourd'hui, nous ne connaissons pas l'avenir du cédérom, ni même du DVD-Rom. Ces supports posent beaucoup de problèmes : ils sont lourds en investissements de programmation, posent des problèmes de compatibilité (Mac/PC), y compris au niveau des systèmes d'exploitation (il y a des différences entre le Windows vendu en France et celui vendu au Québec !).

Certes, le cédérom reste un support qui conservera sans doute la faveur des parents, car les possibilités techniques de bridage de la navigation donnent la garantie que l'enfant restera dans l'espace du cédérom, validé par un éditeur reconnu. Mais l'avenir semble prometteur pour Internet, et l'on exa-

mine avec intérêt et perplexité les nouvelles technologies d'accès à ce réseau (ADSL, Internet satellitaire, etc.). C'est pourquoi le raccordement du cédérom/DVD-Rom à Internet est réellement une volonté éditoriale, qui insiste sur l'encyclopédie conçue à la fois comme un contenu éditorial toujours plus riche et comme un outil de recherche ouvert en direction d'Internet. Car, si même la meilleure des encyclopédies ne peut donner sur tel ou tel sujet ce qu'on pourra trouver sur certain site savant ou institutionnel, seule une encyclopédie peut à la fois fournir sur tant de sujets d'interrogation les informations fondamentales et ménager des pistes, des liens, des itinéraires adaptés à chaque besoin ou à chaque désir « d'en savoir plus ».

Août 2000