

BERNARD PROST



# XML POUR L'ÉDITION

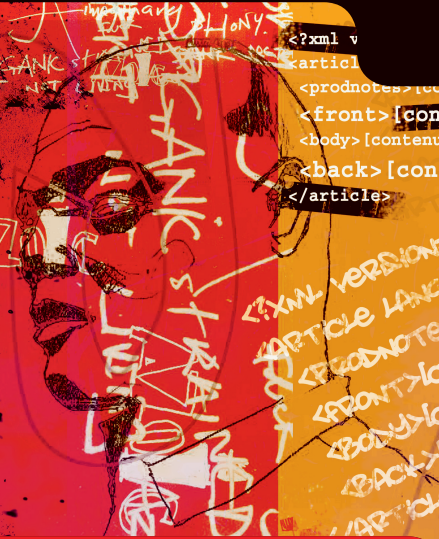
STRUCTURER • SAISIR • PUBLIER

EYROLLES

a'sfored

# XML POUR L'ÉDITION

STRUCTURER • SAISIR • PUBLIER



## La collection a'sfored

Cette collection résulte d'un partenariat avec l'Asfored, le centre de formation de l'édition. Experts dans leur domaine et pédagogues reconnus, les auteurs y livrent leurs méthodes de travail et conseils «métier». Un contenu précieux pour tous les professionnels qui conçoivent, produisent ou diffusent des supports imprimés ou multimédias.

Pour en savoir plus sur le centre de formation et d'expertise de l'édition : [www.asfored.org](http://www.asfored.org).

L'édition vit une mutation dont la portée est considérable : le support papier voit arriver la concurrence de supports dématérialisés chaque jour plus nombreux (Web, *netbooks*, tablettes, téléphones, *readers*...). Pour l'éditeur, il s'agit désormais de publier pour ces supports en tenant compte de leurs spécificités, tout en minimisant les coûts et les délais de production.

Cet ouvrage, qui s'adresse aux éditeurs, aux lecteurs/correcteurs et aux fabricants, mais également aux managers qui souhaitent mieux comprendre les techniques sous-jacentes et l'influence du support sur la conception et la forme de l'objet de publication numérique, vise à donner les clés d'utilisation de XML dans la chaîne éditoriale. Il s'appuie sur un exemple concret d'article encyclopédique repris au fil des chapitres et qui répond à un cahier des charges éditorial simple :

- bibliographies, filmographies et discographies interactives ;
- publication sur papier, sur le Web, le téléphone, les tablettes...

## L'auteur

*Bernard Prost, diplômé de Centrale et de l'Institut d'études politiques, est le fondateur de Ligarant, société spécialisée dans la production de publications numériques sur une base XML, grâce à laquelle il accompagne les éditeurs (édition, presse) dans la mutation numérique de leur activité. Il anime à l'Asfored les stages XML pour l'édition et Bâtir sa stratégie d'édition multisupport.*

# **XML POUR L'ÉDITION**

## Dans la même collection

- M. Dournes, *L'image et le droit – Créer, protéger, reproduire, diffuser*, 2010, 224 pages.  
É. Bacquet, *Préparer des images numériques*, 2009, 200 pages.

## Chez le même éditeur

- H. Plaziat, D. Patte, *Manuel du secrétariat de rédaction* (à paraître).  
A. Brillant, *XML – Cours et exercices*, 2010, 336 pages.  
D. Tardiveau, *L'ActionScript 3 expliqué aux graphistes*, 2010, 224 pages.  
J. Bracquemond, J.-L. Dusong, *Typographie, la lettre, le mot, la page*, 2010, 384 pages.  
E. Lupton (sous la dir. de), *Books et fanzines, do it yourself!* 2009, 176 pages.  
G. Lewis, *2000 accords de couleurs – Pour les graphistes et les designers*, 2009, 336 pages.  
C. Breton-Schreiner, *L'essentiel de la PAO – Mise en pages, logiciels, polices, images*, 2009, 216 pages.  
K. Johansson et al., *La chaîne graphique*, 2<sup>e</sup> édition, 2009, 460 pages.  
G. Davis, *500 grilles et feuilles de style pour l'imprimé et le Web*, 2008, 162 pages + CD-Rom.  
E. Canivet, *InDesign, de la créa à l'exé*, 2008, 232 pages.  
M. Lavant, *XPress en production*, 2007, 244 pages.  
C. W. De Jong et al., *Créations typographiques*, 2007, 400 pages.  
P. Evans, *PLV, Publicité, Packaging*, 2007, 268 pages.  
L. Pasquet, Y. Delpuech, *Réaliser un magazine avec InDesign et Photoshop*, 2007, 164 pages.  
P. Prévôt, *L'informatique de la chaîne graphique*, 2007, 116 pages.  
P. Prévôt, K. Izoulet, *De la prépa de copie à la mise en pages*, 2006, 64 pages.  
P. Prévôt, F. Rocher, *Techniques d'impression*, 2006, 100 pages.  
J. Pieters, *Fabrication du document imprimé*, 2006, 280 pages.  
K. Cheng, *Design typo, concevoir ses polices de caractères*, 2006, 236 pages.  
W. Hill, *Le langage de la typographie*, 2006, 192 pages.  
A. Lonjon et al., *Modélisation XML*, 2006, 498 pages.

**BERNARD PROST**



# **XML POUR L'ÉDITION**

**STRUCTURER • SAISIR • PUBLIER**

**EYROLLES**



a'sfired

Éditions Eyrolles  
61, boulevard Saint-Germain  
75240 Paris cedex 05  
[www.editions-eyrolles.com](http://www.editions-eyrolles.com)

© Groupe Eyrolles 2011  
ISBN : 978-2-212-12657-0

Le code de la propriété intellectuelle du 1<sup>er</sup> juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée notamment dans les établissements d'enseignement, provoquant une baisse brutale des achats de livres, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.

En application de la loi du 11 mars 1957, il est interdit de reproduire intégralement ou partiellement le présent ouvrage, sur quelque support que ce soit, sans l'autorisation de l'Éditeur ou du Centre Français d'exploitation du droit de copie 20, rue des Grands Augustins, 75006 Paris.

# Remerciements

Synthétiser les rapports entre XML et l'édition en un petit manuel est un exercice difficile qu'il est illusoire de penser mener à bien tout seul. Mes remerciements vont donc tout d'abord à Stéphanie Poisson, éditrice de cet ouvrage, à son équipe et à Véronique Dürr, qui l'a assistée dans la relecture : leur aptitude pour rendre lisible une pensée parfois tentée par l'excès de technologie est sans pareille.

Je remercie ensuite tous ceux qui travaillent ou ont travaillé avec moi autour de XML :

- les actionnaires de Ligarane : Alain Pierrot, remarquable concepteur de taxonomies avancées, connaisseur de la suite Open Office, écrivain XSLT, expert de la numérisation de livres ; Xavier Maurin, pilote informatique et graphique du site MyBookForge, dont la vision numérique grand public est brillante ; Olivier Desnoux, ingénieur développeur à la méthodologie sans faille, au code élégant (et lisible !), co-concepteur du moteur de transformation au cœur de MyBookForge ; Adrien Vieilleribière, talentueux chercheur, grand artiste XSLT capable de tout mettre en ligne et rendant la transformation XML vers n'importe quel format cible à la portée de tous, également co-concepteur du moteur de transformation MyBookForge ; Patrick Pierre, ingénieur de talent et l'un des esprits les plus avancés en technologie de publication – sa maîtrise d'IDML (dont je parle très peu dans le livre) est remarquable ; Hugues Cochard, *serial-créateur* d'entreprises high-tech, pour l'heure à Tahiti mais très présent via Internet ;
- tous ceux qui m'ont fait confiance pour leurs projets professionnels ou scientifiques, en particulier Mai NGuyen et Lionel Ridoux pour qui information sur le médicament et XML n'ont plus de secret ;
- les amis de route : Christian Brugeron, pour ses scripts ingénieux destinés à contourner toutes les limitations d'à peu près n'importe quel logiciel de composition, InDesign en premier lieu, et Benoît Leprince, pour avoir bien voulu me livrer différents exemples de maquettes InDesign destinés à illustrer cet ouvrage.

Merci à toutes celles et ceux du tout jeune écosystème e-book, dont la croissance devrait suivre une pente vertigineuse, dans le monde et peut-être en France, en particulier Houriah Ghebalou (PREMICE, l'incubateur régional de Bourgogne), qui a apporté le financement des études préindustrielles du projet Ligarán/Mybookforge, la région Bourgogne, qui contribue au projet et à son implantation locale, et Nicéphore Cité, qui nous abrite à Chalon-sur-Saône et assure une relation permanente avec les start-up de l'image et du son.

Merci enfin à Ray Charles, qui a compris que le support influençait le contenu : sans la nécessité de retourner un 45 tours pour écouter la suite d'un long morceau de musique, le fameux *break* de *What I'd Say* n'existerait pas !





# SOMMAIRE

Avant-propos .....XII

## Chapitre 1

### **Séparer la forme du fond** ..... 3

Modéliser le document ..... 4

    Identifier les 3 aspects d'un document ..... 4

    Identifier les classes de documents ..... 6

    Identifier les structures à repérer en fonction des médias cibles  
    et des usages pressentis ..... 9

    Organiser les éléments structurels ..... 12

Utiliser les balises pour marquer la structure ..... 14

    Balises ..... 14

    Éléments, relations et arbres ..... 16

    Attributs ..... 17

Visualiser le document XML ..... 19

    Visualiser le XML dans un éditeur de texte ..... 19

    Visualiser le XML dans un navigateur ..... 19

## Chapitre 2

### **Les principaux composants structurels** ..... 23

Structure hiérarchique..... 24

    < front > ..... 26

    < body > ..... 28

    < back > ..... 29

Les structures blocs ..... 30

    Le paragraphe ..... 30

    La liste ..... 31

    Le tableau ..... 32

Repérage de blocs particuliers.....	33
Conteneurs et contenus mixtes.....	34
Les structures inline.....	35
Enrichissement.....	35
Images et objets multimédias.....	36
Formules mathématiques, code.....	37
Notes.....	38
Liens hypertextes.....	38
Repérage d'éléments inline particuliers.....	39
Contenus mixtes.....	40
Structures navigationnelles.....	41
Table des matières.....	41
Index.....	42
Titres courants.....	42
Liens hypertextes.....	43
Les entités.....	44
Les entités caractères.....	44
Partage de composants XML.....	46
Métadonnées.....	47
Structure des métadonnées.....	47
Dublin Core.....	48
ONIX.....	49
Définir des métadonnées supplémentaires.....	49

## **Chapitre 3**

### **Écrire/concevoir une DTD**..... 51

Les règles de structures.....	52
Qu'est ce qu'une DTD?.....	52
Contrôler la DTD.....	53
Définir les balises et leur comportement.....	55
Décrire les attributs.....	60
Écrire et organiser la DTD.....	64
Tenir compte (ou non) des supports cibles.....	71
Le support papier.....	71

Le support électronique .....	72
Quelle DTD utiliser? .....	75
DTD généralistes publiques .....	75
DTD métiers .....	76
DTD maison et DTD fournisseurs .....	77

## **Chapitre 4**

<b>Saisir le XML</b> .....	79
Utiliser un éditeur XML .....	80
Qui doit utiliser un éditeur XML? .....	80
Comprendre la relation entre DTD, document XML et visualisation .....	81
Choisir et paramétrer son éditeur XML .....	82
Utiliser une DTD .....	90
Visualiser une DTD .....	90
Documenter une DTD .....	92
Utiliser un document modèle vide .....	93
Référencer la DTD dans le document XML .....	95
Saisir le texte et la structure .....	96
Utiliser l'éditeur XML .....	96

## **Chapitre 5**

<b>Préparer et diriger la saisie XML</b> .....	103
Styler le document Word .....	104
Qu'est-ce qu'un style dans Word? .....	104
Objectifs du stylage .....	105
Les limites du stylage à vocation XML .....	106
Déclarer un style dans Word .....	106
Nommer un style .....	107
Lister les styles de Word .....	108
Gérer les feuilles de styles .....	109
Utiliser les styles pour produire du XML .....	110
Identification des zones à styler .....	110

Nommage, choix typographiques et pose des styles.....	111
La liaison style Word/XML .....	113
Écrire les consignes de saisie .....	114
Que contiennent les consignes de saisie? .....	114
Qui doit écrire les consignes?.....	115
Comment rédiger les consignes?.....	115
Organiser la saisie XML.....	117
Désigner un responsable de DTD.....	117
Gérer les exceptions .....	117

## **Chapitre 6**

<b>Éprouver le XML</b> .....	121
À quelle étape faire intervenir le XML? .....	122
Production XML en amont .....	122
Production XML en cogénération .....	124
Production XML en aval .....	126
Éprouver le XML? .....	127
Objectifs et principes de l'éprouvage XML .....	127
Contrôler les caractères .....	129
Vérifier le contenu hiérarchique .....	130
Vérifier les images et les tableaux.....	131
Vérifier le balisage spécifique .....	131
Vérifier les liens.....	132
Vérifier une livraison de mise à jour .....	133
Les outils de comparaison des éditeurs XML .....	133
Les outils bureautiques.....	135

## **Chapitre 7**

<b>Transformer le XML avec XSLT</b> .....	137
La publication multisupport.....	138
Les transformations multisupports .....	138
Les cibles de publication .....	141



Le langage XSLT .....	146
Principe d'une transformation XSLT .....	146
Spécifier une transformation .....	147
Utiliser un processeur XSLT .....	149
Écrire un programme XSLT .....	150
X-Path .....	152

## **Chapitre 8**

### **Publier pour l'électronique** .....

Publier pour le Web .....	156
Quelle unité d'affichage? .....	156
Transformer les données pour le Web .....	160
Publier au format epub .....	170
Le format epub .....	170
Créer un epub .....	171
Convertir l'epub en mobipocket (Kindle) .....	179
Les supports cibles .....	181
Lire un epub : les logiciels de lecture .....	181
Les tablettes .....	185
Les téléphones .....	187

## **Chapitre 9**

### **Publier pour le papier** .....

XSL-FO : un langage XML ouvert de description de page ....	192
Comment produire un fichier XSL-FO? .....	193
Représentation de la mise en pages à l'aide d'XSL-FO .....	193
Mettre en œuvre une composition XSL-FO .....	198
De XSL-FO à PDF : les moteurs de rendu .....	202
Utiliser InDesign avec XML .....	206
La correspondance élément/styles .....	206
La correspondance attributs InDesign/styles .....	211
En guise de conclusion .....	215

# Avant-propos

*Wait a minute, wait a minute, oh hold it! Hold it! Hold it!*

---

*Hey (hey) ho (ho) hey (hey) ho (ho) hey (hey) ho (ho) hey*  
**Ray Charles** (*What I'd Say*)

*If everything is under control, you are going too slow*  
**Mario Andrett**

L'édition vit une mutation dont la portée est considérable. Le support papier voit arriver la concurrence de supports dématérialisés chaque jour plus nombreux : le Web bien sûr, l'ordinateur ultra compact, appelé *netbook* – sans qu'il s'agisse d'un hasard –, mais surtout le téléphone et les outils nomades spécialisés comme le *reader* (liseuse) d'e-books ou la tablette qui mettent à la portée de chacun, *urbi et orbi*, des bibliothèques complètes d'œuvres littéraires ou professionnelles. Pour l'éditeur, il s'agit désormais de publier pour ces supports en tenant compte de leurs spécificités, tout en minimisant les coûts et les délais de production. Aux expériences nécessitant des reprises multiples du même fonds documentaire a succédé une approche plus industrielle – mais aussi plus normative ou contraignante – basée sur XML.

## DÉFINITION **Liseuse**

Liseuse est le terme français proposé pour *reader* qui, chez les Anglo-Saxons, désigne le lecteur nomade de livres électroniques (e-book). La liseuse désigne un matériel dédié à la lecture utilisant la technologie d'affichage dite *e-paper*, terme marketing désignant un écran non rétro-éclairé, donc peu consommateur en énergie et réputé moins fatigant à la lecture. Associé à l'*e-paper*, le marketing a introduit la notion d'*e-ink* pour qualifier un pixel...

La souplesse et l'universalité de XML ont séduit aussi bien les éditeurs – en premier lieu les éditeurs juridiques, habitués il est vrai à SGML – que les informaticiens qui ont pu ainsi échanger des données entre ordinateurs aux fonctionnements hétérogènes.

Normalisé en 1999, XML a atteint la maturité : il existe désormais un écosystème XML où l'on rencontre des logiciels spécialisés (les éditeurs XML), des prestataires *on-shore*, *near-shore* et *off-shore* rompus à ce

langage, des développeurs d'applications sachant utiliser le DOM (*Document Object Model*) pour créer des produits électroniques toujours plus innovants, et des modèles documentaires spécialisés verticalement par secteur ou horizontalement par type de publication.

Pour autant, la pratique n'est pas encore stabilisée et les habitudes diffèrent considérablement d'un éditeur à l'autre. Le propos de cet ouvrage est d'apporter une vision pratique de l'utilisation d'XML dans les maisons d'édition, fondée sur des usages concrets ayant fait leurs preuves mais par nature limités à des cas particuliers. *XML pour l'édition* n'est ni une bible, ni un essai dogmatique sur le sujet, et chacun pourra adapter sa pratique aux exemples fournis.

#### DÉFINITION DOM

DOM (*Document Object Model*) est une modélisation informatique sous forme d'arbre de documents XML ou HTML. DOM est indépendant de toute taxonomie (voir définition plus loin). Cette modélisation permet une manipulation par programme des constituants (éléments) du document.

## Structure de l'ouvrage

L'ouvrage est organisé en trois parties – structurer, saisir, publier – qui recouvrent l'ensemble du cycle XML pour une publication multisupport de type «livre». *XML pour l'édition* s'adresse aux éditeurs, aux lecteurs/correcteurs, aux fabricants, en premier lieu, mais également aux managers qui souhaitent mieux comprendre les techniques sous-jacentes et l'influence du support sur la conception et la forme de l'objet de publication numérique. Enfin, les auteurs curieux des possibilités qu'apporte cette technique y trouveront matière à concevoir différemment leur œuvre.

L'ouvrage s'appuie sur un exemple d'article encyclopédique, de type Wikipedia, qui sera repris au fil des chapitres. Cet exemple fait appel à une structure spécifiquement développée pour le présent ouvrage (`article_v1.2.dtd`). Les balises ont été choisies en langue anglaise afin de se conformer aux usages internationaux. L'exemple répond à un cahier des charges éditorial simple :

- pouvoir publier l'article sur le papier, le Web, le téléphone ;
- en dotant les objets de publication électronique d'interactivité au niveau des auteurs, des bibliographies, des filmographies et des discographies, cette interactivité doit être indépendante des bases de données cibles.

Pour des raisons de simplicité, il n'est pas prévu de tableaux ni de formules mathématiques (en dehors de l'utilisation éventuelle de celles-ci sous forme d'images).

## Structurer le XML

### DÉFINITION **Taxonomie**

La taxonomie désigne le jeu de balises utilisé pour l'encodage XML d'un document. La taxonomie est généralement décrite à l'aide d'un langage spécialisé (DTD, XML Schema, Relax NG).

Le premier chapitre s'intéresse plus particulièrement à la modélisation des documents et à la démarche du balisage XML. Le second chapitre est, lui, consacré à la description des principales structures que l'on peut rencontrer dans un ouvrage,

ou plus généralement dans un document. Le troisième chapitre donne enfin les clés pour écrire une DTD, c'est-à-dire une des représentations (la plus simple en fait) d'une taxonomie.

## Saisir le XML

Le quatrième chapitre concerne la saisie proprement dite. Dans la plupart des cas, ce travail est confié à un prestataire externe, mais l'éditeur doit de plus en plus être capable de modifier un document à l'aide d'un outil de saisie XML utilisé en interne pour rattraper de petites erreurs ou effectuer des corrections de dernière minute. Ce chapitre aborde en particulier la configuration d'un éditeur XML du marché et son utilisation dans le cadre d'une DTD particulière.

Le cinquième chapitre s'intéresse à la relation avec les sous-traitants : comment préparer la copie pour minimiser les risques d'erreurs d'interprétation au niveau de la structure par le prestataire et bâtir des consignes de saisie efficaces ?

Le sixième chapitre est ensuite consacré à une étape rarement décrite dans les processus de production : l'épreuve XML. Comment s'assurer que le XML fourni par le prestataire est conforme aux attentes de l'éditeur ? Ce chapitre aborde également les différents modèles de production XML selon que la saisie XML s'effectue en amont de la mise en pages papier, pendant celle-ci ou a posteriori.

## Publier

Le septième chapitre expose, dans leurs grandes lignes, les techniques de transformation d'un document XML en un format cible qui peut être aussi bien du XML (par exemple en entrée d'InDesign), du XHTML ou tout autre format textuel. Bien que très technique, le langage de transformation XSLT n'a rien de mystérieux et il importe que les acteurs de l'édition en comprennent le fonctionnement pour apprécier l'impact d'une décision éditoriale.



Le huitième chapitre décrit brièvement la publication pour les supports électroniques en se limitant au Web, aux tablettes, aux liseuses et à l'iPhone choisi comme représentant le plus abouti (actuellement) de la lecture sur téléphone.


Le neuvième chapitre aborde enfin la publication sur papier à partir d'un document XML, selon deux approches :

- la transformation directe du XML en PDF en utilisant XSL-FO, langage de mise en pages lui-même exprimé en... XML ;
- l'importation du XML, éventuellement remanié, par un outil de PAO (ex. : InDesign).

Cet ouvrage vise à donner les clés d'utilisation de XML dans la chaîne éditoriale et se limite à l'essentiel de cette approche moderne de l'édition ; pour chacune des techniques abordées, il existe, aussi bien en français qu'en anglais, des ouvrages très complets auxquels le lecteur se référera avec profit.

#### À NOTER

Le vocabulaire du domaine XML est assez opaque. De nombreux termes sont chargés de références à SGML, aux feuilles de styles, etc., qui ont perdu leur sens d'origine et ne sont plus du tout représentatifs de leur fonction. Il faudra donc faire l'effort de les mémoriser indépendamment de leur sens usuel en français (ou en anglais) courant.



```
<?xml version="1.0"?>
<article language="EN">
<prodnotes>[contenu]</prodnotes>
<front>[contenu]</front>
<body> [contenu] </body>
<back> [contenu] </back>
</article>
```


## Chapitre



# Séparer la forme du fond



La question cruciale pour l'éditeur est de bâtir une méthodologie lui permettant, au prix d'une seule saisie initiale ou d'une saisie a posteriori la moins coûteuse possible, de publier pour de nombreux supports connus ou à venir. La première étape consiste à séparer autant que possible la forme du fond, bien au-delà des techniques mises en œuvre par les feuilles de styles des traitements de texte.



# Modéliser le document

Un livre, ou plus généralement un document sous forme XML, nécessite de disposer d'une modélisation suffisamment générale pour s'adapter à tous les cas de figure probables. On modélisera abstraitement un ensemble de documents – une « classe de documents » – pour les soumettre à des traitements informatiques communs.

## Identifier les 3 aspects d'un document

Après vous être familiarisé avec XML, vous ne verrez plus jamais un document comme par le passé. Au-delà du fond constitué par la juxtaposition des mots sans enrichissement typographique, et de la forme qui met en valeur partiellement la pensée de l'auteur, la (ou plutôt une) structure est une nouvelle composante documentaire apportant des fonctionnalités qui dépendent, en particulier, des usages papier et électronique envisagés.

### Le fond

Le fond est le texte, c'est-à-dire ce qui se lit en dehors de toute considération de mise en forme. Sa version la plus libérée de la forme est l'enregistrement audio : le mot n'a que sa valeur sémantique et n'est pas souligné par une variation typographique, même si quelques variantes sonores peuvent lui donner un peu plus d'épaisseur.

### La forme

La forme enrichit l'information. Elle s'appuie sur une traduction graphique fortement culturelle et linguistique qui rend implicite son décodage.

Dans nos sociétés, graisser un caractère a une signification de mise en valeur aussi bien au niveau d'un titre qu'au sein du texte courant. Le corps du caractère porte le sens du niveau d'importance, généralement couplé à un décalage

visuel vers la gauche : « plus gros et plus à gauche » exprime le plus souvent un titre de niveau hiérarchie plus important.

## La structure

Mettre au singulier le mot « structure » est impropre : il n'existe pas une structure, mais une infinité de structures dépendant de ce que l'on souhaite identifier en vue d'un usage ultérieur.

- Dans le cas d'un roman que l'on veut éditer sur papier et en version électronique, on se limitera à repérer les chapitres, les titres de chapitres, les paragraphes, et à l'intérieur de ceux-ci les mises en valeur au sein du texte.
- Dans le cas d'un article de revue que l'on veut publier sur le Web avec des fonctionnalités de recherche automatique sur Google Scholar ou Google Books (ou n'importe quelle base bibliographique), on repérera les entrées bibliographiques, les noms et éventuellement prénoms des auteurs ainsi que les titres d'ouvrages ou de revues cités.

● Le fond (a) est constitué par le texte brut, celui que l'on peut restituer par la lecture à voix haute (ouvrage audio).

La forme (b) est un apport d'information supplémentaire, largement influencé par la culture et la pratique. Un titre se reconnaît à ce qu'il est écrit en plus gros et en gras. Il exerce une fonction de « repérage » (il s'agit d'un titre) et une fonction « résumé » permettant au lecteur de décider s'il va aller plus loin dans la découverte du texte.

La structure – identifiée ici par des bulles – est une représentation abstraite (souvent guidée par la forme lorsque celle-ci préexiste) destinée à un usage multisupport, sans choix a priori sur le rendu final.

(a)

Miles Davis  
 Miles Dewey Davis III (26 mai 1926, Alton, Illinois - 28 septembre 1991, Santa Monica, Californie) est un compositeur et trompettiste de jazz américain.  
 Une figure centrale du jazz Miles Davis fut à la pointe de beaucoup d'évolutions dans le jazz et s'est particulièrement distingué par sa capacité à découvrir et à s'entourer de nouveaux talents. Son jeu se caractérise par une grande sensibilité musicale et par une fragilité qu'il arrive à donner au son. Il a marqué l'histoire du jazz et de la musique du XXe siècle. Beaucoup de grands noms du jazz des années 1940 à 1980 ont travaillé avec lui.

(b)

Titre (de l'article) → Miles Davis (1926 - 1991)

Illustration (photo) →

Mises en valeur → Miles Dewey Davis III (26 mai 1926, Alton, Illinois - 28 septembre 1991, Santa Monica, Californie) est un compositeur et trompettiste de jazz américain.

Titre (de section) → Une figure centrale du jazz

Paragraphe (de section) → Miles Davis fut à la pointe de beaucoup d'évolutions dans le jazz et s'est particulièrement distingué par sa capacité à découvrir et à s'entourer de nouveaux talents. Son jeu se caractérise par une grande sensibilité musicale et par une fragilité qu'il arrive à donner au son. Il a marqué l'histoire du jazz et de la musique du XXe siècle.

## Identifier les classes de documents

Le modèle documentaire générique capable de représenter n'importe quel type de document n'existe pas. S'il existait, il serait d'une complexité telle que son usage deviendrait prohibitif. On s'attache dès lors à définir des « classes » de documents qui correspondent à des types d'organisation de l'information – par exemple un dictionnaire – ou aux regroupements naturels que sont les collections au sein des maisons d'édition.

Le processus de définition des classes de documents, appelé « Analyse documentaire », consiste à extraire d'un ensemble de documents préalablement identifiés comme étant similaires dans leur organisation, les composants structurels en vue d'un usage ultérieur. On part pour cela d'un nombre limité d'exemples d'ouvrages disponibles jugés représentatifs et l'on bâtit progressivement un modèle répondant aux besoins éditoriaux multisupports.

### Ouvrages hiérarchiques

L'ouvrage hiérarchique le plus élémentaire est le roman ou l'essai : c'est le modèle à la fois le plus simple, le plus répandu, le plus intuitif, mais en même temps complexe car il existe une infinité de variantes structurelles qu'il faudra gérer (quitte à ne pas les prendre en compte ou à les simplifier pour les versions électroniques).

Un essai est souvent (mais pas toujours) organisé en parties, elles-mêmes organisées en chapitres. Le chapitre est doté d'un titre optionnel, lui-même précédé (toujours optionnellement) d'un numéro ou d'un label permettant de le replacer dans l'organisation du livre. Lorsqu'il n'y a ni numéro, ni titre de chapitre, se pose dans la version électronique le problème du repérage de ces mêmes chapitres. Bien sûr, des solutions existent...

Le constituant structurel le plus répandu au sein du chapitre est le paragraphe : unité sémantique voulue par l'auteur et identifiable typographiquement par un retrait de première ligne – ce qui permet de le repérer lorsqu'il apparaît fortuitement en début de page – et un retour en fin de paragraphe. Au sein des paragraphes, l'auteur pourra procéder à des mises en valeur sémantiques qu'il exprimera, entre autres moyens, par du gras ou de l'italique.

Enfin, des variantes typographiques portant sur le paragraphe, fer à droite par exemple, permettent d'exprimer des notions différentes, comme une citation, un exergue, des épigraphes, etc., sans qu'il y ait de limite dans le nombre de variantes envisageables.

#### DÉFINITION **Label**

Le terme anglo-saxon Label désigne un indicateur (graphique, textuel ou numérique) de navigation : numérotation d'une liste, indication du numéro de chapitre...

## Dictionnaires

Chaque dictionnaire est une structure en soi et il n'est pas réaliste de parler de LA classe « dictionnaire » ; on trouvera plutôt des structures propres à chaque dictionnaire, l'objectif étant souvent de disposer de versions papier réduites (par exemple un dictionnaire de poche) ou de versions électroniques offrant des fonctionnalités avancées (liens hypertextes en particulier).

Un dictionnaire est plus proche d'une base de données que d'un document de type livre : il est doté d'entrées, le plus souvent classées par ordre alphabétique, organisées en unités sémantiques ordonnées à la manière des champs d'une base de données.

En pratique, on utilise des entrées modélisées en XML, sorte de micro-documents, que l'on intègre au sein d'une base de données. Cette dernière offre des fonctionnalités puissantes de gestion des entrées (par exemple l'assignation de telle ou telle entrée à tel auteur, le verrouillage d'une entrée lorsqu'une correction est en cours, la gestion des versions, etc.).

Dans le cas d'un dictionnaire comme le Littré, par exemple, on organisera l'information de la manière suivante :

Entrée

en-tête

terme

prononciation

nature grammaticale

variante

numéro de variante

citation,

etc.

Au sein de chaque entrée, l'information sera organisée en autant de blocs qu'il y a de variantes de sens.