

1

Le W3C et les standards du Web

Cet ouvrage recourra souvent aux termes « standard » (ou norme), « compatibilité » (ou interopérabilité), « accessibilité » et « sémantique ». En effet, ces notions constituent la clé de voûte du bon fonctionnement du Web : elles revêtent donc une importance particulière dans un traité exposant la bonne manière d'écrire des feuilles de styles CSS et de concevoir des pages web.

Concevoir une mise en page en CSS ne trahit pas qu'une volonté de suivre l'évolution du Web et de ses techniques. C'est aussi se conformer à des normes allégeant le code, facilitant les mises à jour et employant les divers éléments d'une manière plus « propre » (on évitera ainsi toute mise en page fondée sur des tableaux).

Distinguer très clairement le contenu de son apparence dans des styles CSS présente des avantages qui vous convaincront au fur et à mesure de votre avancée dans ce livre. La maintenance du code et son adaptation aux différents navigateurs et terminaux clients (notamment les outils pour handicapés) s'en trouveront facilitées. HTML revient à son objectif premier : structurer logiquement les documents.

Pour toutes ces raisons, nous aborderons ces questions de standards, compatibilité, accessibilité et sémantique avant d'évoquer à proprement parler la conception de pages web et les feuilles de styles qui rendent son expression possible.

Ce qui fait tourner le Web en coulisses : les standards

Les langages et syntaxes officiels du Web (HTML, XHTML, XML, CSS...) reposent sur des standards publiés par des organisations de normalisation telles que le W3C (World Wide Web Consortium). Des experts y débattent en mettant l'accent sur les aspects techniques pour aboutir à des consensus et des principes cohérents dans les architectures ainsi mises en place (voir figure 1-1).

Figure 1-1
Le site web du W3C

The screenshot shows the W3C website homepage. At the top, the W3C logo is displayed with the tagline "Leading the Web to Its Full Potential..". Below the logo are navigation links: [Activities](#) | [Technical Reports](#) | [Site Index](#) | [New Visitors](#) | [About W3C](#) | [Join W3C](#) | [Contact W3C](#). A paragraph of text describes the W3C's mission: "The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential. W3C is a forum for information, commerce, communication, and collective understanding. On this page, you'll find [W3C news](#), links to [W3C technologies](#), and ways to [get involved](#). New visitors can find help in [Finding Your Way at W3C](#). We encourage you to read the [Prospectus](#) and learn [more about W3C](#)." Below this is a table with three columns: "W3C A to Z" (listing various technologies like Accessibility, Amaya, Annotate, Binary XML, CC/PP, etc.), "News" (featuring two articles: "'Architecture of the World Wide Web, Volume One' is a W3C Recommendation" and "Working Draft: WSDL 2.0 Primer"), and "Search" (with a Google search bar and "Search W3C" button). To the right of the search bar is a "Members" section for the "International Webmasters Association / HTML Writers Guild (IWA-HWG)", which includes a logo and text encouraging members to support W3C standards.

Respecter ces normes transfère aux navigateurs la responsabilité d'interpréter correctement le code. L'auteur de pages web bénéficie ainsi des dernières innovations et s'assure de la pérennité des documents dans le futur.

VOCABULAIRE **Internet n'est pas (que) le Web**

Les médias et le grand public font souvent l'amalgame Web/Internet et utilisent les expressions « site Internet », « adresse Internet », etc. Dans cet ouvrage, nous serons plus précis : le Web n'est qu'un des réseaux enchevêtrés dans le cadre d'Internet, où l'on trouve bien d'autres services indépendants (FTP, courrier électronique, serveurs de temps, etc.). Le Web est sans doute la composante la plus connue et la plus populaire d'Internet, mais il lui est largement postérieur... Il a vu le jour au début des années 1990, plus de vingt ans après les premiers pas d'Arpanet, ancêtre alors confidentiel de l'Internet actuel.

HTML, langage de description de la plupart des documents accessibles sur le Web, a connu plusieurs évolutions successives. Sa version 2 date de 1994 ; la version 3.2 a vu le jour en 1996. À cette époque, c'étaient surtout les éditeurs de navigateurs (Netscape puis Microsoft) qui imposaient leur volonté. L'année 1996 a aussi vu éclater les hostilités commerciales entre ces deux sociétés, à l'occasion de la sortie de Netscape Navigator 3.

Pour éviter au Web de se transformer en une tour de Babel contraire aux principes historiques de l'Internet (libre accès aux données, interopérabilité, échange), le W3C a tenté, par ses normes sur le langage HTML (2.0, 3.2, 4.0...), de suivre le mouvement et de retranscrire les fonctionnalités des principaux navigateurs. Le rapport de forces est aujourd'hui inversé, et les normes actuelles devançant désormais la plupart des navigateurs.

Le standard s'appelle maintenant XHTML (1.0 et 1.1 pour ses versions courantes ; XHTML 2 est en cours de conception). Les feuilles de styles CSS 2 le complètent. C'est une syntaxe plus rigide de HTML, cette rigueur favorisant la maintenance et la relecture du code.

Les navigateurs parlent la même langue

La normalisation amorcée par le W3C recherchait principalement l'interopérabilité des documents web. Il s'agit de les faire comprendre à tous les navigateurs, sur toutes les plates-formes (MS Windows, Unix, Macintosh) et outils (téléphone mobile, assistant personnel, lecteur braille, etc.).

DIVERSITÉ Pourquoi s'embarrasser des autres navigateurs ?

En observant des statistiques indépendantes aussi écrasantes que le Google Zeitgeist de juillet 2004 (<http://www.google.com/press/zeitgeist/zeitgeist-jun04.html>), donnant MSIE 6.0 largement en tête des navigateurs utilisés sur le Web, on pourrait être tenté de se focaliser sur ce produit, pensant se simplifier la vie.

Ce serait une erreur, et ce choix n'allégerait pas vraiment le travail, tout en apportant son lot d'inconvénients : utilisateurs exclus et frustrés, aucune garantie sur l'évolution des futures versions du navigateur ni de la politique commerciale (ou la survie) de son éditeur, faible visibilité sur l'évolution des statistiques et profils d'utilisation du Web (et notamment sur le rôle des clients légers et nomades). IE sera corrigé là où il pêche encore dans son respect des standards ; d'éventuelles solutions de contournement ne seront donc que temporaires. Mettre en place une structure de site logique et solide évitera de sans cesse sur le métier remettre son ouvrage, comme c'était le lot régulier des designers web des années 1990.

Conscients de cette nécessité, les éditeurs de ces programmes les rendent peu à peu conformes aux standards du Web. C'est pourquoi la grande majorité des navigateurs interprètent désormais assez bien, voire parfaitement, ces différents langages. On peut rarement en dire autant de leurs anciennes versions...

LIMITES **Compatible ou presque compatible ?**

Comme annoncé en introduction, nous viserons l'honnêteté intellectuelle. Cela nous oblige à signaler la piètre compatibilité avec les standards du Web du navigateur de Microsoft le plus employé à l'heure actuelle : Internet Explorer 6 (ou IE 6). Celui-ci présentant plusieurs carences notables sur ce plan, il ne suffit pas de développer selon les normes pour assurer l'interopérabilité d'un site web avec IE 6, comme nous en verrons quelques exemples pratiques plus loin. Nous découvrirons également que la dernière version de ce navigateur (IE7) fait de gros progrès en termes de conformité et comble plusieurs lacunes de son prédécesseur. Voir aussi l'annexe F, « Compatibilité des navigateurs ».

Le respect de ces recommandations assure non seulement l'accessibilité d'un site sur la majorité des navigateurs actuels, mais garantit surtout sa pérennité. En effet, il est risqué d'utiliser du code approximatif et des balises obsolètes ou propriétaires, car ces techniques ne sont plus reprises dans les normes actuelles du langage HTML telles que définies par le W3C. Ces habitudes datent du temps des conflits entre les deux grandes puissances Microsoft et Netscape, mais ont désormais dépassé leur période de prise en charge et ne sont plus « sous garantie ».

NORME **Balises obsolètes et propriétaires**

Les normes évoluent, car leurs concepteurs n'ont pas toujours le recul nécessaire lors de leur rédaction pour apprécier l'avenir du domaine. Parfois, et malgré le soin apporté à leur réalisation, elles sont « boguées » et la version suivante corrige le tir. Même s'il est nécessaire de prévoir des périodes de transition (les anciens documents devant rester valides suffisamment longtemps pour laisser à leurs auteurs le temps de s'adapter), c'est une impasse que de s'entêter à garantir à vie tout choix du passé qui s'avère désormais erroné ou inutile. C'est pourquoi le W3C déclare parfois certaines balises *deprecated*, ou « obsolètes ».

Une balise « propriétaire » est un élément absent de la norme qu'un éditeur de navigateur web introduit dans l'espoir de devancer la concurrence. Si seul son produit est capable d'interpréter correctement tel effet spécial et si suffisamment de sites web en font usage (encouragés en cela par le produit de rédaction de site web de la même gamme), les concurrents perdront du terrain et seront forcés d'introduire cette fonctionnalité dans la prochaine version de leur programme. Cette tactique d'entreprise, classique dans les années 1990, est en voie de disparition. Les éditeurs ont maintenant compris l'intérêt de respecter des grammaires communes.

En d'autres termes, leur éventuel bon fonctionnement actuel est un héritage du passé, qui ne préjuge en rien de leur prise en charge future, dans les nouvelles versions ou les nouveaux navigateurs.

Pour un Web accessible à tous

L'ambition d'universalité du Web ne concerne pas que les machines et les logiciels : tous les êtres humains devraient pouvoir y accéder, dans la mesure de leurs moyens, sans barrières artificielles. Cela concerne aussi les personnes souffrant de handicaps, quelle que soit leur nature (physique, auditif, visuel ou moteur). Il existe à cet effet des standards et des normes d'accessibilité du Web, dites WCAG :

▶ <http://www.w3.org/WAI/>

On peut classer les handicaps affectant l'accès au Web en quatre catégories :

- Les déficiences visuelles rassemblent les aveugles, malvoyants, daltoniens et porteurs de lunettes. Les illustrations dépourvues d'intitulé ou de texte de remplacement, les polices de caractères trop petites ou les couleurs peu contrastées limitent alors l'accès à l'information.
- Les déficiences auditives des sourds ou malentendants les empêcheront de profiter de certaines informations si le créateur du site n'a pas prévu à leur effet des dispositions d'accessibilité.
- Des handicaps physiques gênent certains utilisateurs dans le maniement du clavier ou de la souris. Ceux-ci seront donc exclus de tout site exigeant de pointer avec précision de petits éléments ou imposant la souris pour interagir avec leurs scripts et menus.
- Les déficiences mentales ou neurologiques ralentiront certains en l'absence de repères clairs et précis. Un système de navigation non intuitif pourra lui aussi troubler de nombreux utilisateurs. L'abus d'effets visuels comme les clignotements ou les animations de fréquence élevée peut quant à lui avoir de graves conséquences sur des sujets sensibles ou épileptiques.

Les personnes souffrant d'une gêne ou d'un handicap représentent 7 à 40 % de la population selon le type de gêne considéré. Ces proportions augmentent chez les personnes âgées.

On estime que 10 à 20 % des individus présentent telle ou telle déficience dans la plupart des pays dits « développés ». Certains handicaps ne gênent pas l'accès au Web, mais la plupart nous concernent tous potentiellement.

Tout lieu public se doit d'être accessible à tous, handicapés ou non. Les cahiers des charges concernés comportent désormais ces obligations. Pourquoi en irait-il autrement du Web ?

Pour un Web accessible à tous, le W3C a publié en 1997 une initiative (dite WAI) visant à garantir la prise en charge des questions d'accessibilité par les nouvelles technologies.

TEXTES Loi sur l'accessibilité

L'agence fédérale américaine pour le handicap considère que l'obligation d'accessibilité aux lieux publics s'applique aussi à tous les sites web. La « Section 508 » (<http://section508.gov/>) impose déjà l'accessibilité des handicapés aux sites gouvernementaux de plusieurs pays (et à ceux qui sont financés en partie par le gouvernement, notamment ses fournisseurs).

En France, la loi n° 2005-102 du 11 février 2005, pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées, publiée au J.O. n° 36 du 12 février 2005 page 2353, prévoit des règles semblables dans son article 47 :

« Les services de communication publique en ligne des services de l'État, des collectivités territoriales et des établissements publics qui en dépendent doivent être accessibles aux personnes handicapées.

L'accessibilité des services de communication publique en ligne concerne l'accès à tout type d'information sous forme numérique quels que soient le moyen d'accès, les contenus et le mode de consultation. Les recommandations internationales pour l'accessibilité de l'Internet doivent être appliquées pour les services de communication publique en ligne. »

Son texte est disponible en ligne à l'adresse :

- ▶ <http://www.legifrance.gouv.fr/WAspad/UnTexteDeJorf?numjo=SANX0300217L>

C'est dans cette optique qu'ont été rédigées les 14 directives du WCAG (Web Content Authoring Guidelines). Elles décrivent les principes généraux d'accessibilité et détaillent les points à respecter, selon divers niveaux de priorité :

- Niveau 1 : obligation. La consigne doit être satisfaite. C'est le niveau minimal requis pour assurer un seuil d'accessibilité au plus grand nombre.
- Niveau 2 : recommandation. La remarque devrait être suivie.
- Niveau 3 : suggestion. Détail qui pourrait être pris en compte. C'est le plus haut niveau d'accessibilité : il traite de tous les handicaps.

La conformité à ces trois niveaux est codée comme suit :

- un site « A » répond à toutes les exigences d'un point de contrôle ;
- un site « AA » satisfait deux points de contrôle ;
- un site « AAA » se conforme à tous les points de contrôle.

OUTILS Valider l'accessibilité d'un document

Il existe sur le Web plusieurs outils de validation permettant de tester le niveau d'accessibilité d'un document :

- WebXact : <http://webxact.watchfire.com/>
- Wave 3 : <http://www.wave.webaim.org/>
- Cinthia's Report : <http://www.contentquality.com/Default.asp>
- Accès-pour-tous : <http://www.acces-pour-tous.net/>

Attention, ces outils ne permettant qu'une validation partielle, ils ne constituent qu'une première étape dans la démarche d'accessibilité. Tous les niveaux ne sont pas contrôlables par des outils automatiques.

La liste des points à respecter occupe trop d'espace pour être reprise ici, mais nous pouvons en résumer les principales directives :

- Ne jamais poser de limites à la navigation : un menu exclusivement graphique, utilisant des outils propriétaires, reposant sur des scripts ou nécessitant des plug-ins (Java, Flash, JavaScript) est à éviter, à moins de proposer une autre solution, fonctionnelle sans ces outils.
- Éviter les structures de pages utilisant des cadres (frames, iframes) ou tableaux, véritables écueils pour les visiteurs non voyants.
- Toujours proposer des solutions équivalentes au contenu visuel et sonore : texte de remplacement pour les images (alt), navigation sans souris ou sans clavier (accesskey, tabindex), liens hypertextes explicites (title), etc.
- Ne pas s'en remettre exclusivement aux couleurs et permettre d'augmenter la taille du texte (c'est utile aux malvoyants).
- Utiliser un balisage sémantique pour offrir une structure cohérente, même si l'aspect visuel est absent ou dégradé. Respecter les normes de langage en vigueur proposées par le W3C et séparer le contenu (HTML) de la mise en forme (CSS).

Sens et sémantique sur le Web

Les documents du Web seront d'autant plus accessibles et faciles à maintenir qu'ils seront balisés « sémantiquement », c'est-à-dire en fonction de leur structure logique et non pas en se focalisant sur l'aspect recherché.

Un document web bien écrit respecte d'abord un certain nombre de règles de syntaxe, souvent rappelées, mais son aspect sémantique est moins connu. Tout validateur de syntaxe acceptera sans coup férir un tableau bien écrit, même si celui-ci a pour seul but d'obtenir une certaine mise en page... Cet élément est pourtant théoriquement réservé à la présentation de données en relation les unes avec les autres (tableaux à une ou deux entrées), et n'a en aucun cas un rôle d'organisation de la maquette du document.

La sémantique d'un document concerne donc ce qui relève plutôt de la qualification fonctionnelle de son contenu, par opposition à sa forme organique. Elle s'intéresse au type de données d'un objet, à son utilité, à ce qu'il contient, etc.

Sur le Web et en HTML aussi, chaque élément est porteur de sens. Il convient d'employer chaque balise à bon escient et non selon son rendu visuel par défaut.

L'erreur la plus fréquente consiste à utiliser une balise de paragraphe <p> pour le titre du document (<h1> en temps normal). Pour arriver à ses fins, l'auteur applique ensuite à ce « paragraphe » l'apparence d'un titre : taille de texte, graisse, centrage, marges, etc. Si le résultat visuel final est celui d'un titre, la sémantique de cet élément reste malgré tout celle du paragraphe !

Sans être une hérésie, cette technique renie les principes mêmes du langage HTML : décrire le contenu du document et sa structuration logique.

Une sémantique correcte et indépendante de la mise en forme structurera mieux le document, qui sera ainsi plus facile à interpréter par les différents navigateurs, moteurs de recherche et lecteurs braille (lesquels ne s'embarrassent guère de la présentation réelle).

Une sémantique scrupuleuse permettra aussi d'automatiser l'exploitation du contenu de la page par des programmes tels que les « agents » autonomes parcourant le Web à la collecte d'informations – c'est le cas des robots d'indexation des moteurs de recherche.

SCIENCE-FICTION **L'intelligence artificielle n'existe pas**

La notion de validateur de sémantique est difficile à imaginer : aucun logiciel ne pourra deviner qu'un paragraphe est censé être un titre.

C'est donc à l'auteur de bien formaliser la structure de ses pages et d'attribuer à chaque élément la signification adéquate. Par exemple : `<h1>` pour le titre principal, `` et `` pour les listes de liens (menus), `<blockquote>` et `<q>` pour les citations, etc. L'ensemble des balises HTML sera décrit dans le chapitre 2, « Séparer le fond et la forme avec HTML et CSS ».

Le balisage sémantique implique une bonne connaissance des éléments disponibles et de leur signification, culture dont sont dépourvus la plupart des webmasters amateurs ou débutants. Ceux-ci recourent alors à un balisage pauvre et se limitent à trois ou quatre balises génériques et inadaptées.

On rencontre ainsi des documents composés de multiples blocs `<div>` inutiles, ou dont tous les textes (titres, sous-titres, paragraphes, notes, citations, etc.) sont balisés par l'élément de paragraphe `<p>`.

Malheureusement, ces cas abondent sur la Toile, où de nombreux designers font appel à des logiciels WYSIWYG (*What You See Is What You Get*) tels que Dreamweaver ou Golive. Le code automatique ainsi généré employant massivement des éléments neutres plutôt que des balises pourvues d'un sens, il est loin d'être sémantique.

En résumé : avantages des standards et d'une sémantique sur le Web

- Les standards favorisent la création de sites originaux, jolis, dynamiques, développés dans les règles de l'art et aux contenus clairement dissociés de la présentation.
- Développer dans le respect des normes, c'est ouvrir son site à toute plate-forme, tout système d'exploitation, tout navigateur ou agent utilisateur actuel ou à venir, pour peu que ces outils comprennent et appliquent les standards du Web. Peu à peu, cela couvrira une large gamme de produits (assistants personnels, téléphonie

mobile, télévision interactive, domotique, etc.). Un document HTML conforme sera donc exploitable par de multiples outils en dehors d'un ordinateur de bureau.

- Par leur insistance sur la sémantique du balisage des documents, de nombreux standards facilitent l'accès au Web aux personnes souffrant de divers handicaps (visuels, moteurs, auditifs, etc.). C'est d'ailleurs sur ce balisage sémantique que des moteurs de recherche comme Google s'appuient.
- Même les navigateurs anciens ou limités (Netscape 4, lecteurs braille, navigateurs en mode texte comme Lynx) pourront accéder aux sites construits dans le respect des normes, puisque leur structure est différenciée de leur contenu.
- En séparant les aspects éditoriaux et de mise en page, les standards du Web allègent le code et en accélèrent le rendu : HTML strict ne contient en effet aucune indication de style. Les effets de design sont placés à part, dans une « feuille de styles », conservée en mémoire cache de l'ordinateur. Cette technique réduit le volume transféré et réalise donc des économies de bande passante.
- La maintenance et la mise à jour des sites écrits dans le respect des standards s'en trouvent facilitées : la maquette, placée à part, tient sur une seule page (voir à ce sujet le site CSS Zen Garden : <http://www.csszengarden.com/>, qu'un clic de souris permet de rhabiller intégralement, des centaines d'apparences étant disponibles).
- La lecture du contenu est facilitée : chaque balise y est employée à bon escient, sans superflu ni éléments neutres (comme `<div>`) – qui sur-structurent le contenu inutilement. Il est notamment expurgé de toutes les petites « bidouilles » classiques (multiples tableaux imbriqués, `colspan`, `rowspan` et images de type `spacer.gif`).
- Toutes ces raisons (allègement du code, réduction de la bande passante nécessaire, simplification des mises à jour et des refontes graphiques complètes) permettront donc aux entreprises optant pour les standards du Web de réaliser de substantielles économies.
- L'apprentissage de la programmation dans le respect des normes n'est pas plus difficile que celui de HTML classique sans CSS. Par conséquent, pourquoi ne pas gagner du temps et travailler correctement dès le début ?

VOCABULAIRE **Éléments bloc et inline**

On peut classer les balises (X)HTML en deux familles principales. Certaines mettent en place un objet dans la page : elles sont de types « bloc » (exemple : paragraphe, tableau, titre). D'autres ne portent que sur une portion de mot ou de phrase, en agissant par exemple sur sa couleur, sa police, ou d'autres caractéristiques organiques comparables dans des agents utilisateur moins traditionnels (comme l'intonation de la voix dans un lecteur braille). Ces dernières sont dites *inline* en anglais ; on pourrait traduire cela par « au fil du texte » ; dans cet ouvrage nous les qualifierons d'éléments « en ligne ».

PIÈGE Trop de div tue le div !

L'ère des mises en page en tableaux imbriqués est révolue : c'est l'association de balises <div> aux styles CSS qui prend la relève et permet la mise en forme des documents.

Cette nouvelle mode apporte fatalement son lot de fanatismes, incompréhensions et mauvaises utilisations. La paresse simplificatrice naturelle de l'homme fait conclure à de nombreux designers : « bien, la nouvelle lubie consiste donc à remplacer les tableaux par des div ». Cette analyse bâclée produit mécaniquement des documents comptant autant de div que leurs prédécesseurs comptaient de cellules de tableaux, ce qui n'a évidemment aucun intérêt. Ces cancre ont alors beau jeu de se plaindre et de dénoncer l'inutilité de CSS.

Ce phénomène est malheureusement souvent observé : ceux qui n'en comprennent pas la logique utilisent des éléments div à toutes les sauces, en les imbriquant et en créant une multitude de sous-blocs... ce qui n'est pas une fin en soi.

Les inconvénients sont nombreux : mises à jour difficiles, code alourdi inutilement et rendu incompréhensible à autrui, ainsi qu'à son propre auteur après quelque temps. Tel est le prix à payer pour la sur-structuration en HTML.

Rien de tel qu'un exemple parlant pour illustrer ce type de débordements :

```
<div class="gauche">
<div class="menu">
<ul>
<li><a href="#">Lien 1</a></li>
<li><a href="#">Lien 2</a></li>
<li><a href="#">Lien 3</a></li>
<li><a href="#">Lien 4</a></li>
</ul>
</div>
</div>
```

Le menu (bloc) est imbriqué dans un div "menu", lui-même placé dans un div "gauche"... tous deux sont pourtant inutiles ! Il suffisait ici de doter la balise importante () des attributs concernés : c'est en effet possible pour tout élément de niveau bloc.

```
<ul class="menu gauche">
<li><a href="#">Lien 1</a></li>
<li><a href="#">Lien 2</a></li>
<li><a href="#">Lien 3</a></li>
<li><a href="#">Lien 4</a></li>
</ul>
```

La morale ? N'oubliez pas que de nombreuses balises mettent en place des blocs : <p>, , , <h1>...<h6>, <blockquote>...

Le recours à l'élément `<div>` ne doit pas être un réflexe conditionné. Vous ne réaliserez des économies et n'écrirez un code propre et compréhensible qu'en employant chaque balise à bon escient.

C'est tout l'esprit de la sémantique des balises : `<p>` structure des paragraphes, `<h1>` à `<h6>` des niveaux de titres, `` des listes et des menus, etc. On peut les voir comme des macros prédéfinies : `<p>` remplacerait `<div role="paragraph">`, etc. Évitez donc de surcharger votre code de `<div>` inutiles et autres balises superflues ; il ne s'en portera que mieux. Ainsi allégé, il sera bien plus facile à relire.

HTML ou XHTML ?

Dans cet ouvrage nous emploierons indifféremment les termes HTML et XHTML pour désigner le langage de codage des documents du Web – XHTML n'est en effet qu'une reformulation de HTML en XML. XML est un méta-langage, c'est-à-dire un ensemble de règles à respecter pour qu'une grammaire de langage puisse se réclamer de cette norme. Quelques implémentations fameuses de XML : DocBook (documents techniques), MathML (formules mathématiques), RDF (description des contenus par leurs méta-données comme auteur, titre, description), SVG (description de figures), etc.

La différence principale entre HTML et XHTML est syntaxique, ce dernier étant bien plus rigide. C'est pourtant cette rigueur qui lui donne toute sa souplesse, et en multiplie les applications.

Syntaxe générale du XHTML

La grammaire du XHTML répond à un certain nombre de règles, pour la plupart fondées sur l'absence d'implicite, une grande régularité et un « bon parenthésage » du document :

- Les noms des balises et des attributs sont écrits en minuscules.
Exemple : on écrit : `<p>` et non plus `<P>`.
- Les valeurs des attributs sont placées entre apostrophes simples ou doubles.
On écrira : `<p class="center">` au lieu de `<p class=center>`.
- Tout attribut doit impérativement recevoir une valeur.
`<input type="checkbox" checked="checked" />` remplacera ainsi `<input type="checkbox" checked>`.
- Toute balise ouverte doit être refermée.
On écrira donc : `<p>Bonjour.</p>` au lieu de `<p>Bonjour.`

- Les balises vides doivent être explicitées.
Exemples : `
` supplantera `
` et `<hr id="top" />` est le successeur de `<hr id="top">`.
- Les balises seront correctement imbriquées (et le document, « bien parenthésé »).
Il semble plus logique de préférer `<p><i>Bonjour</i></p>` au bancal `<p><i>Bonjour</p></i>`.

Tout document se conformant strictement à ces règles sera dit « bien formé » syntaxiquement et respectera la grammaire XHTML.

Indiquer la grammaire au navigateur

Il existe plusieurs versions du langage HTML, cohabitant sur le Web. On indiquera aux navigateurs la version retenue pour un document donné en préfaçant celui-ci d'un doctype accompagné d'une DTD, ou « définition de type de document ». C'est cette DTD qui définit la grammaire précise du document.

Ce doctype est un code spécifique placé au tout début d'un document HTML. Il précise au navigateur le langage retenu pour l'écriture de la page : HTML ou XHTML, en version stricte ou transitionnelle, etc.

C'est non seulement un moyen de s'assurer que le document sera bien interprété par les navigateurs, mais aussi un élément indispensable à sa validation par l'organisme de référence, le W3C.

XHTML (et HTML) existent en deux versions principales :

- *Transitional* (transitionnelle), version prévue pour réaliser la transition vers la version *strict*, est par conséquent beaucoup plus permissive que cette dernière.
- *Strict* est la version rigoureuse que nous utiliserons. Beaucoup d'éléments ou d'attributs y sont devenus obsolètes (et interdits). Il s'agit notamment des balises et propriétés de mise en forme (`<center>`, ``, `"align"...`). Le but est de pousser les développeurs à utiliser le CSS pour la présentation et de limiter le document HTML au seul contenu.

On trouve encore d'autres doctypes, proposant par exemple les cadres (frames), mais nous les éviterons car ces techniques sont à déconseiller et entravent l'accessibilité des sites web.

ÉVITER À propos des cadres

Rien de tel qu'un exposé argumenté pour convaincre le lecteur sceptique de la nocivité des cadres. Voyez à ce sujet l'excellent article du site de référence OpenWeb à l'adresse :

- ▶ http://openweb.eu.org/articles/finir_cadres/

À l'heure actuelle, il existe sept doctypes, tous aussi valides et conformes selon le W3C : HTML 4.01 frameset, HTML 4.01 transitional, HTML 4.01 strict, XHTML 1.0 frameset, XHTML 1.0 transitional, XHTML 1.0 strict et XHTML 1.1.

Le concepteur web est libre de choisir son doctype selon ses besoins et convictions. Par exemple, les versions frameset autorisent les cadres et les versions transitionnelles permettent d'insérer des éléments de mise en page au sein du code HTML. La différence entre chaque doctype est résumée dans un article en ligne de Laurent Denis : <http://css.alsacreations.com/Bases-et-indispensables/DTD-comment-choisir>

Cet ouvrage emploiera généralement la syntaxe du doctype XHTML strict afin de se référer au document le plus rigoureux possible. Par conséquent, nos documents web débiteront généralement par :

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

Cette formulation est précise et rigoureuse (elle mêle notamment des balises et des parties en majuscules). Pour éviter d'éventuelles erreurs, contentez-vous de la copier-coller dans vos créations.

NORME DTD XHTML 1.1

À titre indicatif, signalons l'existence de la DTD XHTML 1.1, modularisation de XHTML 1.0 strict. Elle pose quelques contraintes supplémentaires et nous ne l'utiliserons pas dans le cadre de cet ouvrage d'introduction.

Le passage de XHTML 1.0 à XHTML 1.1 marque une transition nette vers le langage XML. Ces deux versions se distinguent nettement au niveau de la déclaration de contenu. Il ne suffit pas de choisir le bon doctype pour qu'un document XHTML 1.1 soit valide ; les documents XHTML 1.1 servis en "text/html" sont invalides.

Indiquer la langue et l'encodage du document : lang et charset

La petite bureaucratie virtuelle des normes du Web impose d'autres déclarations obligatoires dans l'en-tête de tout document HTML : les spécifications de langue et d'encodage. Survolons-les rapidement.

La langue du document est précisée dans la balise <html> comme suit :

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="fr" lang="fr">
```

L'encodage est spécifié dans l'espace délimité par les balises <head> et </head>, à l'aide de l'attribut charset. Par exemple :

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="fr" lang="fr">
<head>
<title>titre évocateur</title>
<meta http-equiv="Content-Type"
      content="text/html; charset=iso-8859-15" />
</head>
```

L'encodage du document indique aux navigateurs les caractères potentiellement utilisés dans le texte de la page. Pour un document en français, on pourra choisir entre :

- iso-8859-1 qui est l'encodage classique, presque complet pour les langues de l'Europe occidentale ;
- iso-8859-15 qui complète le précédent par quelques caractères supplémentaires, tels que le signe € ou le caractère œ ;
- utf-8 qui permet d'utiliser la plupart des caractères de la majorité des langues du monde : c'est un code de l'Unicode.

CULTURE Unicode

Pour s'échanger des données, notamment textuelles, les ordinateurs doivent s'accorder sur la manière de les représenter (ou encoder). Ils ne connaissent en effet que les nombres : comment associer à chaque caractère son numéro ? La multitude des conventions inventées localement a subi une première unification avec l'avènement de l'Internet dans les pays occidentaux dans les années 1970 : c'est le code ASCII qui numérotait les vingt-six lettres de l'alphabet, les chiffres et quelques signes typographiques ou codes de contrôle supplémentaires. Il fut étendu différemment selon la zone géographique pour prendre en compte les signes diacritiques des diverses langues. L'entrée de cultures aux systèmes d'écriture moins classiques dans le cercle de la communication a bouleversé ce fragile équilibre ; c'est désormais Unicode qui vise à centraliser les dizaines de milliers de caractères existant dans tous les systèmes d'écriture. On peut désormais mêler tout sous-ensemble de langues du monde dans un même document.

Valider son site avec le W3C

Pour réaliser un site conforme aux standards du Web, il faut :

- Préciser la version de HTML ou de XHTML utilisée. On l'indiquera par un doctype en début de document, comme nous venons de le voir.

- Se conformer à la grammaire retenue. Pour l'aspect syntaxique, les validateurs du W3C conviendront, mais aucun programme ne peut garantir ou contrôler la bonne sémantique des balises d'un document.

Comme le validateur (X)HTML (<http://validator.w3.org/>) ou le validateur CSS (<http://jigsaw.w3.org/css-validator/>), des outils en ligne vous aideront à vous conformer aux règles du langage correspondant au doctype choisi pour le document.

Ces programmes analysent les documents web qu'on leur fournit et indiquent diverses erreurs commises par rapport au doctype stipulé ((X)HTML *Transitional*, *Frameset* ou *Strict*). Sauf bogue improbable du validateur, chaque indication sera pertinente et il conviendra de la prendre en compte.

Confronté à un document construit proprement, sans erreur de balises, imbrications hasardeuses ni recours à des propriétés non autorisées ou obsolètes, le validateur affichera une page de félicitations confirmant ainsi la bonne conformité au doctype retenu.

Figure 1-2

Le validateur de code (X)HTML du W3C

Rien n'oblige un développeur à se conformer à une norme, mais ce choix présente plusieurs avantages : assurance de la pérennité du site, affichage correct sur une plus large gamme de navigateurs, meilleure accessibilité, etc.

Un site respectant les normes « strictes » sera plus rigoureux et facile à faire évoluer vers des normes à venir. Malheureusement, il sera plus difficile d'en assurer un rendu satisfaisant tant sur les anciens navigateurs que sur les nouveaux.

CULTURE La mission du W3C

Le W3C ne se limite pas à créer et maintenir des langages du Web. Sa mission, telle qu'il l'a définie, est de mener celui-ci à son potentiel maximal. Cela passe par sept fronts principaux : accès universel, Web sémantique, confiance, interopérabilité, évolutivité, décentralisation, multimédia interactif.

Quelques sociétés très influentes composant le W3C (Microsoft, Mozilla, Opera, entre autres), on peut raisonnablement penser que le souhait de normalisation n'est pas une utopie.

Vous trouverez des informations (en anglais) à l'adresse :

- ▶ <http://www.w3.org/2002/07/W3Cin7PointsTranslations.html>

Gardez à l'esprit que passer l'étape de validation est une condition nécessaire mais en aucun cas suffisante. C'est un test mécanique qui ne contrôle que la syntaxe. De même qu'on peut imaginer des phrases grammaticalement correctes mais dépourvues de sens en français (exemple : « D'incolores idées vertes dorment furieusement »), il est tout à fait possible qu'un document web organiquement correct emploie les balises à tort et à travers. En particulier, il est impossible au validateur de vérifier si vous avez utilisé les bonnes balises au bon endroit, de savoir si votre <p> est bien un paragraphe et non un titre (<h1>), etc.

Rien ne s'oppose, par exemple, à la définition d'un bloc de 70 000 pixels de haut. C'est propre et valide, mais cela n'a aucun sens. On enchaînera donc sur la prise en compte d'autres paramètres essentiels, comme la sémantique :

http://openweb.eu.org/articles/respecter_semantique/

Les validateurs d'accessibilité compléteront avantagement cet arsenal :

<http://www.acces-pour-tous.net/validateur/validateur.php>

Pour résumer, on peut dire que les normes et leurs validateurs ne régissent que l'aspect organique des documents. Leur interprétation et leur sens restent l'affaire de leur auteur.

VOCABULAIRE « Recommandation » : un terme qui pêche par modestie

Le W3C encadre le travail de groupes d'experts qui élaborent ce qu'ils appellent des « recommandations » lors d'un processus formalisé en plusieurs étapes. Ce mot est mal choisi : il s'agit en réalité des standards du Web, qu'il convient de fait de respecter. Voir par exemple :

- ▶ http://www.w3schools.com/w3c/w3c_intro.asp.

Testez vos connaissances

- ❶ Combien de navigateurs web existe-t-il dans le monde ? Combien en connaissez-vous ?
- ❷ Combien de personnes souffrent de handicap visuel (et notamment de cécité) en France ?
- ❸ Quand fut créé le W3C (World Wide Web Consortium) ?
- ❹ Quel attribut obligatoire rendra une image accessible aux non voyants ?
 - alt
 - title
 - info
 - desc
 - longdesc
- ❺ Qui est considéré comme le fondateur du Web ?
 - Jeffrey Zeldman
 - Bill Gates
 - Timothy Berners-Lee
 - Alan Turing
- ❻ Les balises `` et `<i>` ont souvent la même apparence visuelle. Pourquoi préférer l'une à l'autre, et laquelle ?
- ❼ Quel est le site de référence en français pour les standards du Web ?
 - www.cnil.fr
 - www.w3c.org
 - www.openweb.eu.org

Vous retrouverez certaines de ces questions à l'adresse suivante :
<http://www.tutoweb.com/quiz.php>

Réponses

- ① Il existe près de 100 navigateurs web, dont certains sont déjà obsolètes.
Les navigateurs les plus courants du moment sont : Internet Explorer, Netscape Navigator, Mozilla, Firefox, Opera, Safari, Camino, Konqueror, Galeon, w3m, Lynx, Links (ces trois derniers fonctionnant en mode texte).
Vous trouverez une liste complète des navigateurs sur le site :
<http://browsers.evolt.org>
- ② La France compte plus de 1 500 000 déficients visuels (soit 2,6 % de la population), dont, en 1997, environ 110 000 aveugles (acuité visuelle inférieure à 1/20 de la normale au meilleur œil après correction) et 250 000 malvoyants (acuité visuelle inférieure à 4/10). Source : *Quid*.
- ③ Le W3C a vu le jour en 1994. Vous trouverez des informations complémentaires sur son site <http://www.w3.org/Consortium/>
- ④ L'attribut `alt` est un minimum obligatoire. Il permet aux non voyants de disposer d'un texte alternatif décrivant brièvement l'image.
Attention à ne pas confondre les attributs `alt` et `title`. Ce dernier est une infobulle s'affichant lors du survol des éléments par le curseur de la souris. La confusion vient du fait qu'Internet Explorer affiche aussi une infobulle avec l'attribut `alt`.
- ⑤ Timothy Berners-Lee est considéré comme le père fondateur du Web.
- ⑥ Les balises `` (« emphase ») et `<i>` (« italique ») ont souvent le même effet dans les navigateurs graphiques. Leur interprétation diffère pourtant, et le contenu mis en emphase sera plus facile à interpréter ou représenter dans des contextes moins classiques. Pour résumer, `<i>` a une sémantique de présentation et `` une sémantique plus abstraite.
- ⑦ Le site de référence en français sur les standards du Web est www.openweb.eu.org, que je vous invite à parcourir sans modération !