L'édition structurée et le World–Wide Web

Vincent QUINT (INRIA Rhône–Alpes)
Irène VATTON (CNRS – IMAG)

Outils d'auteur pour WWW

Deux méthodes pour produire des documents destinés au Web :

Les filtres

Un filtre convertit un format de document spécifique (LaTeX, RTF, FrameMaker, Grif, etc.), en HTML

- L'éditeur n'a pas de connaissance du Web
- Des retouches manuelles sont souvent nécessaires

L'écriture en HTML

L'auteur saisit les documents en HTML avec un éditeur de texte

- L'auteur doit bien connaître HTML
- Rien ne garantit que le document soit syntaxiquement correct

Problèmes posés

La très grande majorité des documents disponibles ne sont pas conformes à la DTD HTML 2.0

- Documents ambigus : des browsers différents peuvent donner des résultats différents

Aucune aide aux auteurs pour manipuler les liens hypertexte

- Il faut connaître les URL des documents référencés
- Il est difficile d'utiliser les liens typés (attributs REV, REL)

Une approche naïve

HTML est une DTD SGML

⇒ Utiliser un éditeur SGML avec la DTD HTML 2.0

Problèmes

- La plupart des documents existants sont rejetés par les parsers SGML
- La DTD HTML 2.0 est très plate ⇒ dialogue complexe
- La question des liens et des URL reste posée

L'approche structurée

Les principaux concepts des documents structurés se retrouvent dans HTML, et plus encore dans HTML 3.0

⇒ Un éditeur structuré est une bonne base pour construire un environnement auteur pour le World–Wide Web

Mais il doit être adapté aux conditions particulières du Web.

Adaptation réalisée sur Grif :

- Définition de la structure des documents
- Parser tolérant
- Transformation de structure
- Gestion des liens hypertexte

Structure des documents

HTML est décrit sous une forme plus structurée mais compatible :

- dialogue plus simple lors de la construction de la structure,
- traitements plus efficaces par l'éditeur.

Un parser spécifique à HTML

- accepte les structures invalides et tente de les corriger,
- produit une structure plus riche.

En sortie

- la structure interne est exprimée dans la syntaxe HTML.

Présentation des documents

Plusieurs modèles de présentation sont définis

- pour les manipulations interactives à l'écran
- pour l'impression

Chaque modèle de présentation propose **plusieurs vues** :

- vue intégrale, vues des ancres, vue des titres, vue de la structure HTML

Modification dynamique de l'affichage

- Zoom, changement de modèle de présentation, holophraste, etc.

D'autres modèles peuvent être proposés.

Transformations de structure

Les transformations de structure sont nécessaires pour

- corriger la structure d'un document invalide,
- offrir de la souplesse à l'utilisateur

Commandes de transformation de structure :

- **Diviser** un élément en deux, **Fusionner** deux éléments successifs
- **Englober** des éléments existants dans un nouvel élément
- **Changer** le type d'un élément sur place
- **Coller** un élément structuré, qui a été coupé ou copié ailleurs

Gestion des liens et accès au Web

L'éditeur intègre la bibliothèque libwww :

- fonction de « browsing », accès à tous les documents du Web
- couper, copier, coller entre documents, y compris les documents distants
- écriture à distance (méthode POST)

Manipulation des liens en tant que tels

- établissement des liens par simple clic sur la cible
- commandes spécifiques de manipulation des ancres :
  Créer, Éditer, Copier, Coller, Détruire

Perspectives (1)

Évolution vers HTML 3

Introduction des tableaux et des formules mathématiques

Formatage plus complexe et paramétré par attributs

Gros documents

Écrire une documentation sous la forme d 'un seul document

L'éclater en pages HTML pour la publier sur le Web

Réunir les pages HTML pour lire ou imprimer le document

Perspectives (2)

Édition coopérative

Permettre à plusieurs utilisateurs d'éditer en même temps les mêmes documents

Contrôler dynamiquement le découpage en pages HTML

Évolution vers SGML

Grif peut éditer simultanément des documents HTML et SGML

Copier–Coller entre documents avec restructuration

Échange en HTML des documents SGML

Échange direct de documents SGML entre certains utilisateurs

Conclusion

Caractéristiques souhaitables d'un environnement auteur pour WWW :

- souplesse et tolérance du parser, pour accepter les documents existants
- traitements spécifiques au Web : liens et structure HTML
- accès au Web pendant l'écriture
- manipulation directe sur des documents formatés en cachant HTML
- contrôle de la présentation : mise en page, vues
- production de documents syntaxiquement valides
- évolutivité pour suivre rapidement les extensions de HTML