

Introduction au contexte documentaire moderne

Plan

- I - Problématiques liées aux documents papiers
- II – Problématiques liées aux logiciels de traitement de texte
- III - Problématiques liées aux techniques de rédaction
- IV - Techniques de publication documentaire pour le Web
- V - Techniques de présentation des documents électroniques

I - Problématiques liées aux documents papiers

Numérisation

- Le traitement du papier implique des coûts importants :
 - Main d'oeuvre
 - Manutention et espace d'archivage
 - Matériel (numériseur)
- Produit simplement une image du document

Reconnaissance de caractères (OCR)

- Indispensable pour la diffusion juridique
 - Édition du contenu du document
 - Fonctionnalités de recherche
 - Intégration graphique
- Nécessite un logiciel spécialisé (Omnipage, Clara OCR)
- Imparfait : vérification humaine subséquente essentielle

II - Problématiques liées aux logiciels de traitement de texte

Informations cachées

- Type d'informations cachées par les logiciels de traitement de texte
 - Attributs de format du texte (police, version)
 - Propriétés du document (auteurs, dates)
 - Données d'édition (textes antérieurs)
- Méthode de visualisation
 - Éditeur de texte générique (Bloc Note)
 - Fonctions d'un logiciel de traitement de texte (Word, WordPerfect)

Informations cachées (suite)

- Procédures de minimisation
 - Configuration des postes de travail
 - Utilisation de modèles de documents
 - Nettoyage des fichiers avant distribution
 - Suppression des informations personnelles
 - Afficher le texte masqué
 - Suppression des marques de révision
 - Convertir les objets incorporés ou liés

Styles

- Fonction permettant d'uniformiser la présentation d'un type d'éléments à l'intérieur d'un document
- Avantages
 - Assure l'uniformité
 - Accélère la mise en forme
 - Facilite la conversion des documents
- Leur mise en forme manuelle entraîne des problèmes de conversion

Paragraphes non standards

- Les fonctions appropriées du logiciel doivent être utilisées pour
 - Tableaux
 - Colonnes
 - Tables des matières
 - Etc
- Leur mise en forme manuelle entraîne des problèmes de conversion

Images et autres objets

- Doivent être collés, et non liés ou incorporés
 - Information cachée
 - Multiplie les fichiers pour un document
- La conversion des images n'est pas toujours assurée
- Certains objets ne peuvent être affichés en HTML

Date

- Doit être insérée manuellement, et non à l'intérieur d'un champ
 - Sinon, il y a risque de mise à jour automatique lors de l'ouverture, modification, enregistrement du fichier
 - Situation entraînant une perte de l'information originale
- Exception au principe de l'utilisation des fonctions appropriées du logiciel

III - Problématiques liées aux techniques de rédaction

Organisation des documents

- Chaque document devrait être contenu dans un fichier unique
- Ce qu'il faut éviter
 - Séparer un document en plusieurs fichiers
 - Rassembler plusieurs documents dans un seul fichier
 - Modifier la séquence de présentation logique d'un document

Numérotation

- Numérotation des paragraphes devrait être privilégiée à la numérotation des pages
 - Problème de présentation visuelle
 - Problème de citation
- Lorsque la numérotation des pages est essentielle
 - Récupération dans l'entête
 - Insertion de balises dans le corps du texte

Mode de citation

- Uniformisation des citations est essentielle pour la reconnaissance automatique
 - Référence neutre
 - Mention du numéro de paragraphe
- Possibilité d'insertion d'hyperliens
 - À la source
 - Au moment de la diffusion

Note de bas de page

- La fonction appropriée du logiciel devrait être utilisée
 - Doit y avoir un lien entre la note et l'appel de note
 - Doit y avoir une suite ininterrompue entre les différentes notes
 - Les références internes doivent être dynamiques
- Converties en note de fin de document

Entête et bas de page

- Ne devraient contenir aucune
 - Information unique
 - Information relative à l'organisation du contenu
- Lors de la conversion
 - Parfois supprimées
 - Parfois incluses, mais uniquement au début et à la fin du document

IV - Techniques de publication documentaire pour le Web

Langage de balisage (HTML, XML, XHTML)

- HTML est un langage de publication
 - Universel sur le Web
 - Édition aisée
- XML est un langage de structuration
 - Permet de regrouper le document avec ses métadonnées
- XHTML combine les deux

Feuilles de style (CSS)

- Permet de contrôler la présentation graphique d'un ensemble documentaire
 - Éléments : titres, paragraphes, listes, etc
 - Attributs : police, case, couleur, etc
- Permet d'uniformiser l'apparence
- Facilite les modifications

PDF

- Format propriétaire d'Adobe dont les spécifications sont disponibles
- Assure l'uniformité multi-plateforme de la présentation d'un document
 - Idéal pour donner accès au document original
 - Idéal comme format de publication secondaire
- Trois types possibles (image, texte, combiné texte et image)

Sites statiques / dynamiques

- Les pages d'un site statique sont générées avant leur publication
 - URL simples et fixes
 - La génération peut être longue et complexe
- Les pages d'un site dynamique sont reconstruites à chaque requête
 - URL complexes et changeants
 - Permet l'affichage de contenus personnalisés

V - Techniques de présentation des documents électroniques

Minimalisme des pages d'affichage de documents

- Seule l'information relative au document devrait être affichée avec celui-ci
 - Accélère le téléchargement
 - Répond aux attentes des usagers
 - Favorise un usage professionnel
- Solutions
 - Retirer les menus, boîtes de recherche, commentaires
 - Créer une entête et un bas de page adaptés

Faciliter l'impression

- Les usagers s'attendent à pouvoir imprimer correctement les documents
- Solutions
 - Pages HTML sans table
 - Pages d'impression
 - Publication d'une version PDF

Uniformisation des styles

- Les styles devraient être cohérents à l'intérieur d'un document
- Les différents documents d'une même collection devraient partager les mêmes styles
- Solutions
 - Feuilles de styles (CSS)
 - Suppression des styles originaux

Métadonnées

- L'information permettant de juger de la fiabilité des documents devrait être mise en évidence
 - Date
 - Auteur / source
 - Référence
 - Taille
 - Etc

Étiquettes

- Les métadonnées ajoutées devraient être identifiées par une étiquette
 - Permet aux nouveaux usagers de s'y retrouver rapidement
 - Accélère le repérage de l'information
- La présentation des étiquettes devrait être uniforme sur tout le site