

Publication des documents

Contexte

- Concerne l'identification de l'ensemble des éléments techniques qui composent les systèmes de diffusion du droit
- Publication
 - Étape du processus de diffusion se situant entre l'acquisition des documents et leur consultation par les usagers

Plan

I – Classement et organisation des données

II – Conversion

III – Interface

IV – Moteur de recherche

V – Archivage

I – Classement et organisation des données

Organisation générale

- Les données devraient être classées de façon logique et mnémonique
- Se réalise par le biais d'une architecture de répertoire hiérarchique
- Niveau hiérarchique commun aux documents juridiques
 - Statut
 - Langue
 - Collection
 - Année

Statut

- Devrait se situer à la racine de l'architecture
- Représente l'état d'avancement du document dans le processus de publication
- Exemples de répertoires à adapter selon les besoins
 - reception/
 - traitement/
 - source/
 - web/
 - archives/

Langues

- Nécessaire lorsque l'interface est disponible en plusieurs langues
- Il faut séparer les contenus car chaque version linguistique doit être autonome
 - Permet de conserver les mêmes chemins d'accès
 - Facilite l'insertion de liens contextuels
- Exemples
 - web/en/
 - web/fr/

Collections

- Les documents provenant d'une même source devraient toujours être conservés ensemble
 - Permet une navigation logique
 - Si un problème avec un document survient, la source est facilement identifiable
- Exemples
 - [web/fr/csc/](#)
 - [web/fr/caq/](#)

Année

- Le nombre de document provenant d'une seule source peut être important
 - L'année de production du document est un dénominateur commun pour diviser le contenu
 - Si un problème survient à un moment dans le temps, permet de circonscrire rapidement les fichiers concernés
- Exemples
 - [web/fr/csc/2003/](#)
 - [web/fr/csc/2004/](#)

Noms de fichiers

- Le même nom de fichier devrait être utilisé tout au long du processus de publication
 - Référence neutre (si disponible)
 - Référence interne
- Il est important de conserver le nom de fichier original afin de faciliter les communications avec sa source

II – Conversion

Numérisation et reconnaissance de caractères

- Trois possibilités
 - Texte
 - Image
 - Combiné texte et image
- La reconnaissance de caractères implique une étape de validation et de saisie manuelle
- Temps de traitement long et coûteux

Outils de conversion automatisés

- Peuvent être intégrés à l'intérieur d'un processus de traitement automatisé
- Permet la conversion de masse de documents
- Nécessite du développement informatique pour adapter l'outil choisi au contexte particulier

Outils de conversion automatisés (suite)

- Polyglotte
 - Logiciel libre développé au LexUM
 - Communique avec Word et Acrobat pour leur faire exécuter des tâches (ouverture du fichier, conversion, sauvegarde)
 - Nécessite une licence d'utilisation payante pour ces logiciels
- Tidy
 - Nettoie le HTML généré par Word pour le rendre compatible avec les standards du W3C

Conversion à l'aide des logiciels de traitement de texte

- La plupart des logiciels de traitement de texte permettent de sauvegarder les documents en HTML, voir en XML
- Ne requiert aucun autre outil informatique
- Ne permet pas de traiter des masses de documents
- Fichiers générés
 - Le code ne respecte pas toujours les standards
 - Contiennent beaucoup de code inutile

III – Interface

Types d'accès

- Plusieurs façons d'accéder aux documents
 - Thèmes / catégories
 - Domaine du droit
 - Collection
 - Géographie
- Possibilité d'en combiner plusieurs
- Doit être déterminé en fonction de
 - Besoins identifiés
 - Nature des documents
 - Métadonnées disponibles

Index

- Composés de liens pointant vers chacun des documents
 - Facilite la navigation
 - Grande accessibilité aux documents
 - Respect des critères d'ouverture de la diffusion libre
- Parfois, les documents sont uniquement accessibles par un moteur de recherche
 - Vise à empêcher le téléchargement des données à grande échelle

Outils de navigation

- Navigation doit être aisée pour qu'un visiteur s'y retrouve dès sa première visite
 - Liens vers les pages accessoires dans l'entête et le bas de page
 - Menu pointant vers les grandes sections du site
 - Lien sur le logo pointant vers la page d'accueil
 - Insertions d'hyperliens dans les documents
 - Existence d'une page générale pour chaque niveau hiérarchique
 - Liens dans les titres permettant de remonter vers un niveau hiérarchique supérieur

Apparence graphique

- L'apparence d'un site à l'écran a une grande influence sur sa facilité de navigation et sa consultation
- Élaboration d'une charte graphique
 - Associer une apparence à chaque élément
 - Trois couleurs maximum
 - Fond d'écran blanc
 - Minimiser les images (temps de chargement)
- Au besoin, peut être adaptée à chaque collection

Accès au document original

- Consiste à ajouter un lien pointant vers le document original, tel que fourni par la source
 - Fichier de traitement de texte ou version PDF
 - Permet aux usagers de constater les transformations survenues lors de la conversion
 - Permet d'imprimer le document sans l'apparence graphique du site
 - Donne plus de crédibilité pour un usage professionnel

IV – Moteur de recherche

Contexte

- Tout site Web publiant un nombre important de documents devrait posséder un moteur de recherche
- Recherche devrait toujours être plein texte
- Exemple de moteur de recherche libre
 - Lucene

Recherche globale / contextuelle

- Sur la page d'accueil
 - Recherche globale sur tout le site
- Sur chacune des pages de navigation
 - Possibilité de rechercher dans la section concernée
 - Choix laissé à l'utilisateur
- Cette façon de faire permet de limiter le champ des recherches sans utiliser la page de recherche avancée

Recherche simple / avancée

- Les boites de recherche des pages de navigation devraient être simples à utiliser (peu d'options)
- Une page doit rassembler toutes les fonctionnalités de recherche (page de recherche avancée)
- Facilement accessible de toutes les pages (entête, bas de page)
- Recherche par champ
 - Varie en fonction des métadonnées disponibles

Recherche booléenne

- Les juristes ont souvent des recherche très pointues à effectuer
 - Ils désirent pouvoir raffiner leur recherche
 - Aidé par les opérateurs booléens
 - et, ou, n/2, " ", etc
- La plupart des moteur de recherche offrent plusieurs modes, dont la recherche booléenne
 - Devrait être mis à disposition des usagers

Autres fonctionnalités pertinentes

- Repérage des occurrences
 - Identification de l'endroit où les termes recherchés se situent dans le texte
- Mécanisme de suivi
 - Identification des documents ayant cité celui affiché (fondé sur les liens des citations)
- Affichage d'extraits
 - Affichage de la ou les phrases où le terme recherché a été repéré

V – Archivage

Objectif

- Conserver une copie réutilisable du contenu pour les situations où
 - Le projet se termine
 - Attaques informatiques
 - Panne de serveur
 - Sinistre physique

Quoi archiver

- Au minimum
 - Document original tel que fourni par la source
 - Document diffusé
- Au mieux
 - Archivage effectué à chacune des étapes du processus de publication
 - Document original, converti, source et diffusé
- Le code informatique nécessaire à la diffusion doit également être archivé

Techniques d'archivage

- Manuellement
 - Création de fichiers ZIP après chaque opération et les sauvegarder sur un autre support
- Automatiquement
 - Sur un LAN, serveur d'archivage sauvegardant régulièrement le contenu des répertoires sur d'autres disques
- Une copie devrait être conservée dans un lieu physique éloigné