

Présentation

MathDoc

UMS UJF-CNRS

Crée en 1995

Rendre accessible
la documentation
Mathématique

Développer des
bibliothèques
Numériques

NUMDAM !!

- Elizabeth CHERHAL-CLEVERLY
- Elizabeth.Cherhal@ujf-grenoble.fr
- Ingénieure Informaticienne/documentaliste
- à MathDoc (Cellule de Coordination Documentaire Nationale pour les Mathématiques) depuis 1995
- Responsable de :
 - CEDRAM (portail pour les revues académiques de math)
 - Portail Documentaire mathématique
 - Catalogue des périodiques
 - Gallica-Math
 - ...
- Avant: mise en place des premiers applications réseau à Grenoble. (premier serveur web en 1993).

La Documentation électronique

- Plan ->
- **Définitions**
- Production
 - Formats
 - Spécificités
- La trouver
 - Sources
 - Moteurs

- Plan des cours : = 3 séances
 - **1) Documentation électronique**
 - Définitions
 - Production
 - Documents scannés, documents nativement numériques
 - Spécificités mathématiques: LaTeX, mathML
 - La trouver ?
 - Sources
 - Moteurs de recherche
 - **2) Bibliothèque numérique**
 - Définitions
 - Métadonnées
 - Technique
 - **3) L'archivage pérenne** (ou conservation numérique)
 - Problématique
 - Risques et solutions
 - Idées et modèles
 - Projets

La Documentation électronique

- Définitions - >
- Production
 - Formats
 - Spécificités
- La trouver
 - Sources
 - Moteurs

Définitions

- Un vocable: des réalités diverses :
 - BU : ressources (généralement payantes) en ligne accessibles grâce à des abonnements
 - Recouvre : bases de données bibliographiques, journaux en ligne, "bouquets".
 - Dans ce cours : **tout document** disponible sous forme numérique.

La documentation électronique

- Définitions - >
- Production
 - Formats
 - Spécificités
- La trouver
 - Sources
 - Moteurs

Définitions (suite)

Physique

- Toute donnée informatique (autre qu'un programme informatique ou fichier système) qui doit être utilisée sous forme informatique. À l'inverse du document matériel, le document électronique peut permettre de séparer l'aspect présentation (mise en forme, mise en page...) et l'aspect information (contenu, données...), offrant alors la possibilité d'une exploitation séparée
- Ensemble cohérent d'objets numériques (textes, graphiques, photos, images animées et sons) stockés sur des machines informatiques interconnectées ou stockés sur des supports informatiques amovibles et transportables

Logique

- -> Documents scientifiques classiques: article, ouvrage, rapport, thèse, "prépublication" ...

La documentation électronique

- Définitions
- **Production ->**
 - Formats
 - Spécificités
- La trouver
 - Sources
 - Moteurs

- Produire un document ?
 1. Scanner un document papier
 2. Produire un document à l'aide d'un logiciel = "nativement numérique"
- **Un document scanné = une image** (plusieurs images)
 - Plusieurs choix: couleur, niveau de gris, noir/blanc
 - Plusieurs résolutions (ppp, ou dpi), la plus grande n'est pas nécessairement la meilleure.
 - Monopages /multipages.
 - Plusieurs formats d'image:
 - Tiff, jpeg, png, pdf...
- documents scannées de NUMDAM : n/b, 600 dpi, tiff **et** pdf
- Exemples :
 - [Document constitué de plusieurs pages "images"](#)
 - [Document constitué d'un fichier "multipages"](#)

La documentation électronique

- Définitions
- **Production ->**
 - Formats
 - Spécificités
- La trouver
 - Sources
 - Moteurs

- **L'OCR** (optical character recognition, reconnaissance optique de caractères)
 - Extrait le texte de l'image
 - Peut créer un fichier à part ou "cacher" le texte sous l'image
 - Spécifier la langue !
 - Aucun OCR du marché ne sait reconnaître les formules mathématiques.
- **Logiciels courants d'OCR**
 - Finereader
 - Acrobat
- Sans OCR pas de recherche dans le texte, ni copier/coller.
- **Exemples :**
 - Un [fichier pdf](#) en mode "image"
 - Un [fichier pdf](#) avec "texte caché"

La documentation électronique

- Définitions
- **Production**
 - Formats ->
 - Spécificités
- La trouver
 - Sources
 - Moteurs

- Documents nativement numériques = produits à l'aide d'un logiciel
- Multiplicité de formats et logiciels :
 - Format/logiciels propriétaires :
 - Suite office, publisher, works...
 - Formats/logiciel ouverts :
 - OpenOffice
 - TeX/LaTeX
 - HTML, XML, ...
 - Format propriétaire mais spécifications ouvertes :
 - PDF (produit par logiciels propriétaires **et** ouverts)
 - Un fichier PDF "nativement numérique" contiendra toujours du "texte caché".

La documentation électronique

- Définitions
- **Production**
 - Formats - >
 - Spécificités
- La trouver
 - Sources
 - Moteurs

- Format de **création**, format de **diffusion**, format d'**archivage**
 - Création : bureautique ou TeX/LaTeX
 - Diffusion : PDF et/ou HTML
 - Archivage : format source + autre format (XML)

La documentation électronique

Un mot sur HTML et XML (cf cours sur les sites web)

- Définitions
- **Production**
 - Formats->
 - Spécificités
- La trouver
 - Sources
 - Moteurs

- **HTML**
 - Langage du Web
 - 1989 ->
 - Maintenu par W3C
 - Balisage
 - Contenu et présentation mélangés
 - Interprété par un navigateur
 - Nombreux éditeurs graphiques
- **XML**
 - "Méta langage"
 - Nombreux vocabulaires
 - Émanation du W3C
 - balisage
 - Contenu et présentation dissociés
 - De multiples applications:
 - Site web dynamique
 - Échange de données
 - Fichiers de configuration des logiciels
 - Traité par un programme ou une feuille de style

La documentation électronique

- Définitions
- Production
 - Formats
 - Spécificités
 - >
- La trouver
 - Sources
 - Moteurs

- Spécificité des documents mathématiques -1
- Les formules ...
- Les éditeurs de formules des logiciels de bureautique ne sont pas très évolués (suite oo ou éditeur d'équation de word).
- TeX/LaTeX reste le moyen de **créer** un texte mathématique.
 - Logiciel libre et maintenu
 - Longévité remarquable
 - Insertion d'images de différents formats
 - Création de PDF

La documentation électronique

- Définitions
- **Production**
 - Formats
 - Spécificités
 - >
- La trouver
 - Sources
 - Moteurs

- Spécificité des documents mathématiques – 2
- Maths sur le web...
- L'HTML ne permet pas l'expression de formules mathématiques
 - Conversion des maths en image (plusieurs logiciels: TeX4ht, latex2html...)
 - MathML....
 - **Un** vocabulaire XML pour exprimer les mathématiques
 - Verbeux, destiné au traitement **par un programme**
 - En progrès: de plus en plus d'applications utilisent MathML Convertisseurs LaTeX->MathML commencent à fonctionner.
 - Plus de possibilités "dynamiques" que PDF
 - Exemples
 - [Une page avec des formules sous forme d'image](#)
 - [MathML sur CEDRAM](#)

La documentation électronique

- Définitions
- **Production**
 - Formats
 - Spécificités ->
- La trouver
 - Sources
 - Moteurs

- En ligne actuellement vous trouverez :
 - Articles courants:
 - écrits (par l'auteur) avec LaTeX
 - Généralement repris par l'éditeur de la revue (mise aux normes, mise en forme...)
 - Disponibles sur le web en PDF
 - (Très rarement : documents en MathML)
 - Articles anciens: (date variable)
 - Résultat de scan: généralement PDF (avec ou sans texte caché)

La documentation électronique

- Définitions
- Production
 - Formats
 - Spécificités
- **La trouver**
 - Sources->
 - Moteurs

Où se trouve-t-elle ?

- Serveurs d'éditeurs commerciaux (masse)
- Serveurs de "petits éditeurs"
- Archives ouvertes (cours du mercredi)
- Autres Bibliothèques numériques
- Serveurs de départements et instituts
- Serveurs de bibliothèques
- Pages personnelles et blogs

La documentation électronique

- Définitions
- Production
 - Formats
 - Spécificités
- **La trouver**
 - Sources ->
 - Moteurs

Serveurs des grands éditeurs

- La masse des documents en ligne (articles actuels et passés) se trouve aux mains des éditeurs commerciaux.
- ([Springer](#), [Elsevier](#), Wiley, grandes presses universitaires)
- Environ 80% de la production courante
- Objet d'accords de consortium.
- Difficulté (impossibilité?) de négocier des accords par discipline.
- En France, Couperin, INIST
- (au Maroc IMIST ?)

La documentation électronique

- Définitions
- Production
 - Formats
 - Spécificités
- **La trouver**
 - Sources ->
 - Moteurs

Serveurs de petits éditeurs

- Sociétés savantes : (AMS, SMF,...)
- Sociétés d'édition avec clientèle société savante (EDP Sciences)
- Petites presses universitaires
- Organismes tels qu'[Euclid](#), [CEDRAM](#) ...
 - N'entrent jamais dans les "bouquets" des consortium.

La documentation électronique

- Définitions
- Production
 - Formats
 - Spécificités
- **La trouver**
 - Sources ->
 - Moteurs

Bibliothèques Numériques :

- Fonction de "bibliothèque" : (on "archive" les collections, on les rend accessibles à un public), indépendant des éditeurs)
- Archives. (la production courante étant souvent accessible sous abonnement seulement)
- Exemples :
 - JSTOR, Euclid (accès pas toujours libre)
 - [Digizeitschriften](#) (Göttingen)
 - [NUMDAM](#)
 - [Gallica](#)
 - (autres initiatives : [Pologne](#), Portugal...)

La documentation électronique

- Définitions
- Production
 - Formats
 - Spécificités
- **La trouver**
 - Sources ->
 - Moteurs

- **Serveurs de départements et Instituts**
 - = Archives Ouvertes "institutionnelles"
 - Documents en général non publiés
 - Problème de pérennité
 - Problème de masse critique
 - Problème de visibilité
 - [Liste](#) sur le portail MathDoc
- **Serveurs de bibliothèques**
 - Cela commence...
 - [SCD Strasbourg](#), projet Paris 6 / Paris 7
 - ...

La documentation électronique

- **Définitions**
- Production
 - Formats
 - Spécificités
- **La trouver**
 - Sources ->
 - Moteurs

- Pages Personnelles
 - Beaucoup de documentation... peu de moyens de la trouver
 - Parfois dupliquée sur des serveurs d'AO
- Blogs
 - Rare en ce qui concerne les mathématiques
 - Pléthorique en sciences de l'information

La documentation électronique

- Définitions
- Production
 - Formats
 - Spécificités
- **La trouver**
 - Sources
 - Moteurs ->

Comment s'y retrouver ?

- Impossible d'aller sur tous les sites susceptibles de contenir des documents ☹
- Les grandes bases de données :
- MathSciNet / Zentralblatt
 - Excellents outils pour trouver une référence (si plus d'un an d'age !)
 - Moins bons pour un document numérique: liens automatiques sur les sources connues de documents en ligne. (pas de liens sur les "petits" serveurs).
 - Exemple : [démonstration mathscinet](#) (bulletin smf)
 - : [démonstration mathscinet](#) (invent.math.)

La documentation électronique

- Définitions
- Production
 - Formats
 - Spécificités
- **La trouver**
 - Sources
 - Moteurs ->

Comment s'y retrouver ?

- Impossible d'aller sur tous les sites susceptibles de contenir des documents ☹
- Moteur de recherche :
 - Généraliste : parcourt des sites web et en indexe le contenu, -> google, [google scholar](#).
 - Particularité de google scholar: permission des grands éditeurs d'indexer leur contenu (texte caché des fichiers PDF).
 - Spécialisé : "moissonne" des méta données provenant de plusieurs sources et les rassemble en une base: -> [miniDML](#), [Base Search](#), etc. (mais pas d'accès au grand éditeurs)

La documentation électronique

- Définitions
- Production
 - Formats
 - Spécificités
- **La trouver**
 - Sources
 - Moteurs ->

Projets fédérateurs internationaux spécifiques aux mathématiques

- Sous l'égide de l'UMI:
 - WDML : pas de contenu
 - DML Un projet entre les universités de Cornell et Göttingen
 - EMANI
- Sous l'égide de la SME
 - EU-DML (projet soumis à l'Union Européenne)
- Seul moteur spécialisé en Mathématiques: miniDML géré par MathDoc.

Projets réalisables grâce au protocole OAI-PMH (voir cours mercredi).

La documentation électronique

- Définitions
- Production
 - Formats
 - Spécificités
- La trouver
 - Sources
 - Moteurs

- **Références: (liens)**
- Springer Link : <http://www.springerlink.com/>
- Elsevier (science direct) : <http://www.sciencedirect.com>
- Accès via l'inist (tous labos CNRS) <http://www.sciencedirect.com.gate4.inist.fr/>
- Euclid : <http://projecteuclid.org/>
- Digizeitschriften: <http://www.digizeitschriften.de/index.php?id=64&L=2>
- NUMDAM: <http://www.numdam.org>
- CEDRAM : <http://www.cedram.org>
- SMF : <http://smf.emath.fr/Publications/>
- SME: <http://www.ems-ph.org/>
- Pologne (Kolekcja Matematyczna): <http://matwbn.icm.edu.pl/>
- MiniDML : <http://minidml.mathdoc.fr>
- BaseSearch (Bielefeld): <http://www.base-search.net>
- SCD Strasbourg : <http://num-scd-ulp.u-strasbg.fr:8080/view/subjects/510.html>
- Liste des serveurs de "preprint" : http://portail.mathdoc.fr/Prepub/liste_sites_prep.php