XediX Tera Solution ™
Une implémentation industrielle du concept NXD

Didier Courtaud

CEA DAM Ile-de-France

Courriel : Didier.Courtaud@cea.fr

Notion de document

- Ensemble d'informations qui représente une unité que l'on peut raisonnablement considérer comme indivisible et complète
- Véhicule principal de transfert des connaissances
- Peut être multimédia
- Exemples
  - Mémo
  - Livre
  - Film
  - Cédérom
  - site Web3

Le document électronique

- Représentation d'un document sous forme d'une structure de données informatiques entreposable dans la mémoire d'un ordinateur et transmissible d'un ordinateur à un autre

- Peut ou non correspondre à un document existant sur support traditionnel

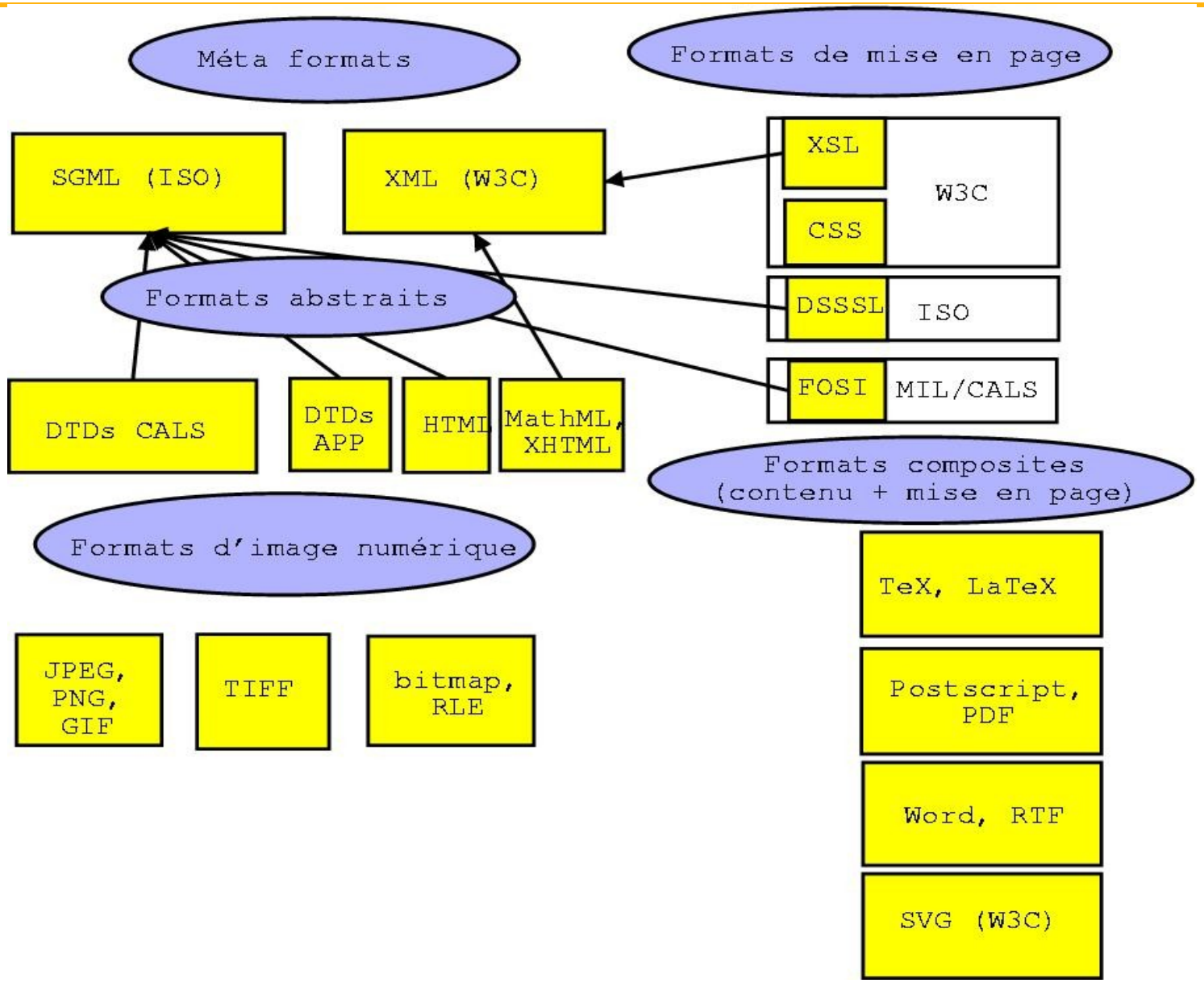
Formats de documents électroniques

- **Ensemble de règles ou de conventions régissant l'interprétation des documents électroniques**

- **Syntaxe**
  - Ensemble des règles auxquelles une séquence de caractères doit se conformer pour être reconnue comme un document valide

- **Sémantique**
  - Ensemble des règles permettant de transformer un document électronique valide en document " réel "

Classification des formats



Le document structuré

- Le format est dit structuré s'il
  - permet le balisage descriptif ou abstrait des documents
  - donne une définition de la grammaire liant ces balises
- Un document est structuré selon un certain type de document (jeu de balises et grammaire)
- Comme une fiche de base de données relationnelle est structuré selon la structure de sa table
- On parle alors de structure logique

Structure physique

- La structure physique d'un document est sa mise en page.
- Un document structuré peut avoir plusieurs mises en page distinctes
- Une mise en page peut être générée par un outil de formatage automatique
- On peut régler le degré d'intervention de l'utilisateur

Le document structuré : impact

- **Factorisation du travail**
  - Séparation du contenu et des traitements
  - Séparation de la définition des structures logiques et de leur utilisation
- **Repérage de l'information**
  - Information structurelle est une valeur ajoutée exploitable pour le repérage
  - Spécification des applications indépendante des contenus
- **Pérennité, réutilisation de l'information**

- Méta-format
  - Définit un format standard pour placer des balises de type descriptif dans un document
  - Permet la description de différents types de structures logiques
- Définit des catégories de documents ayant la même structure logique (ensemble des balises et grammaire liant ces balises)
- Structure logique définie dans la DTD ou dans un schéma XML
- Un document est valide s'il respecte son type
  - Précise aussi les règles syntaxiques d'utilisation des balises

XML (II)

- Langage descriptif
- Sémantique & syntaxe
- Format non dirigiste
  - Ne s'occupe pas de la mise en page
  - La mise en page se fait via une application de restitution

Avantages de XML

- Pérennité
- Normalisation
- Robustesse du balisage généralisé
  - validation syntaxique à la source
- Spécification des applications indépendante des contenus
- Indexation automatique, recherche d'information améliorées
- Format d'échange

- XML peut-être vu comme langage d'import-export de fiches instance de tables dans une base de données.
  - On ne sort pas du modèle relationnel.
  - Format neutre, standard d'échange : XML
  - On parle de base "XML-enabled"

- Le document XML est l'entité élémentaire du système.
  - L'utilisateur utilise XML comme modèle pour les informations qu'il manipule.
  - Le problème du stockage vient après.
  - On peut employer le terme de "base documentaire".

XML & SGBD (II)

- Comment stocker des documents XML ?
  - Fichiers à plat
    - Petits volumes
    - Requêtage faible (*grep*)
  - SGBDR
    - Modèle relationnel mal adapté
    - Un fichier XML décrit un arbre qui se décrit difficilement sous forme de tables
    - Nécessité d 'une multitudes de tables et de jointures
      - Temps de traitement prohibitif
    - Créations de BLOB (Binary Large OBject) ou de CLOB (Character Large OBject) pour stocker tout un arbre XML dans une case de tableau
      - Solution bâtarde qui sauvegarde le modèle relationnel mais n 'autorise aucune granularité

- **Bases de données étendues**

  SGBD (objet-) relationnel étendu avec des outils de traitements des documents XML

  - Définition d'un schéma relationnel pour stocker des documents XML
  - Nouveau type d'attributs XML
  - Interrogation par SQL

Avantages :

- On peut traiter en même temps des données XML et des tables relationnelles classiques
- Passage doux du relationnel vers XML

Stockage de documents XML (III)

- SGBDOO
  - Modèle objet possible mais pas satisfaisant
  - L 'arbre XML n 'introduit pas de notion d 'héritage
  - La réutilisation de l 'objet ne sert à rien
  - Il faut une approche couplée modèle-XML pour pouvoir profiter pleinement du couplage entre les deux mondes (cf approche MDA de l 'OMG)
  - Dans le cas général, les données ne proviennent pas de modèles
    - Couplage faible entre l 'objet et XML
    - Performances correctes mais pas exceptionnelles

Une nouvelle approche

- XML nous oblige à repenser le problème du stockage
- Avant
  - Les données avaient des formats hétérogènes
  - Elles étaient essentiellement numériques
  - Nécessité d 'un modèle de données fédérateur pour le stockage
    - Relationnel : tables 2D
    - Objet : objets *héritants*
- Après
  - Un seul format de données : XML
  - Des données diverses (textes, nombres, multimédia) décrites chacune par une structure XML particulière
  - Plus besoin d 'un modèle de données fédérateur
    - XML est le modèle de données fédérateur

Une nouvelle approche (II)

- **Un nouveau type de base de données est né :**

  les bases de données XML natives (NXD)

  Bases de données spécifiquement conçues pour XML

  - **Modèle conçu pour le stockage et l'accès à des arbres ordonnés**
  - **Le document XML est l'entité centrale de la base**

  Avantages

  - **Chargement efficace de gros documents**
  - **Mise à jour efficaces**

Base de données XML natives

- Tout est décrit en XML sur des structures (DTD ou Schéma) appropriées
  - Les nombres
  - Les documents
  - Les utilisateurs
  - Les droits d 'accès
- Unicité et généricité des traitements
  - *Quincaillerie* XML : XSLT, CSS, …
- Simplification de la gestion de base
  - Tout support physique peut convenir
    - SGBDR, SGBDOO
    - Bases hiérarchiques
    - Système de fichiers (ReiserFS sous Linux)
- Précision et granularité de la recherche

Un exemple au CEA DAM Ile-de-France

- Besoin de conservation de connaissances
  - Expériences passées
  - Fond hétérogène
    - Papier
    - Textes électroniques sous différents formats (Word 6, Word 97, PDF)
    - Vidéos
  - Volume important : de l 'ordre de 400 Go
- Besoin de lier document source - document électronique
  - Gérer des copies d 'écran issues du papier de manière synchrone avec leurs copies électroniques
  - Gérer des documents multimédia (vidéo, images)
- Assurer la pérennité du fond documentaire
- Besoin d'accès à la granularité la plus fine

XediX : Démarche de développement

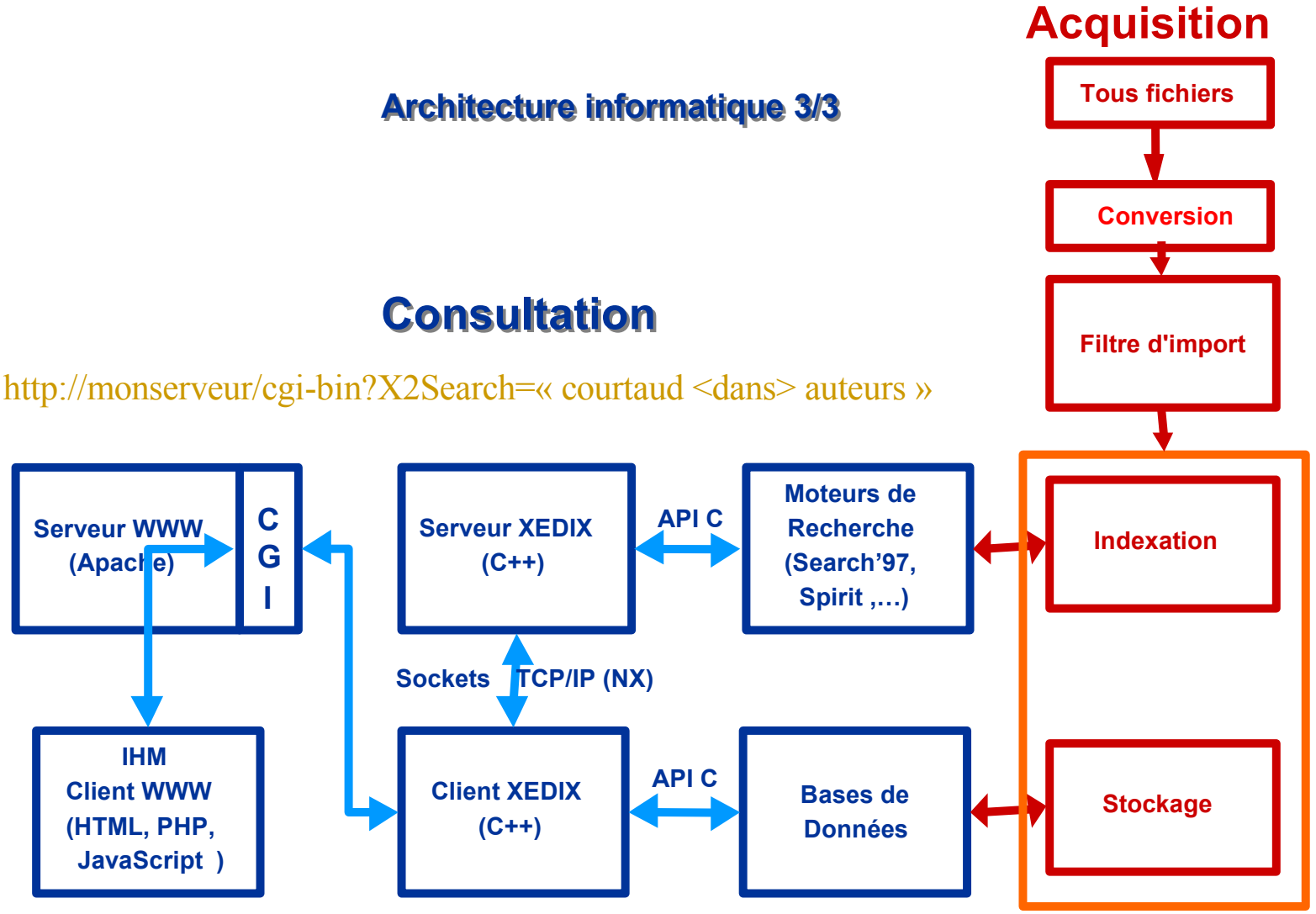
- En 1996, la pérennité nous impose de choisir un format non propriétaire, reconnu comme norme au niveau international
  - Choix de SGML, père de XML
- Comme aucune solution commerciale n'existait, nous avons choisi de développer une solution de type NXD
  - 1996 - 1998 : Première version sur spécification CEA réalisée par Euroclid sur base de données O2
  - 1998 - 2000 : Mise à niveau de la V1
    - Adoption de XML
    - Mise en place d 'un moteur de recherche linguistique
  - 2001 - 2003 : Ré-écriture au CEA DAM Ile-de-France d 'une nouvelle version qui implémente complètement le concept de NXD sur plusieurs bases et plusieurs moteurs de recherche
- Aujourd'hui XediX Tera Solution est passé en phase de production au CEA DAM Ile-de-France

- Pour cela, Xedix associe :
  - Un filtre d'import & un filtre d'export
  - Un gestionnaire de base
  - Des moteurs de recherche
  - Un gestionnaire des droits d'accès
  - Une interface Web de consultation
  - Gestion des rétroconversions (*legacy*)
  - Gestion fine du multimédia
  - Personnalisation de l'interface par CSS
  - Un paquetage de procédures d 'import et de transformation en XML
  - Une API d'accès par HTTP

Architecture informatique 3/3

Consultation

http://monserveur/cgi-bin?X2Search=« courtaud <dans> auteurs »

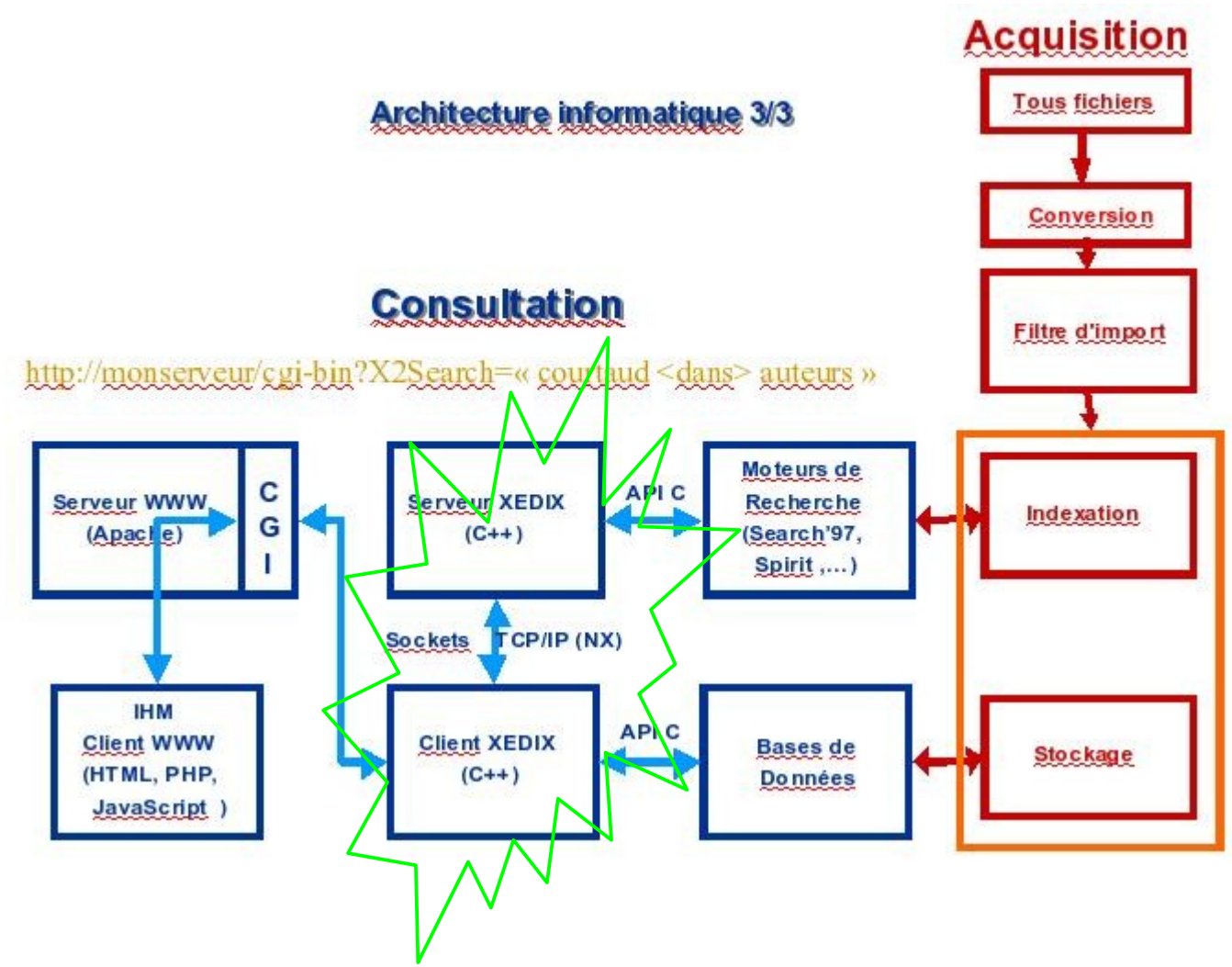


Configuration de XediX

- XediX est régi par deux fichiers de configuration qui explicitent
  - Les répertoires d'installation de Xedix et de ses différents composants
  - Les options choisies pour certains de ces composants

- Ce sont
  - `config.ini` définit les emplacements des différents répertoires de XediX et les paramètrages généraux
  - `dts.ini` définit les paramètres particuliers à chaque DTD qui serviront lors de l'export HTML

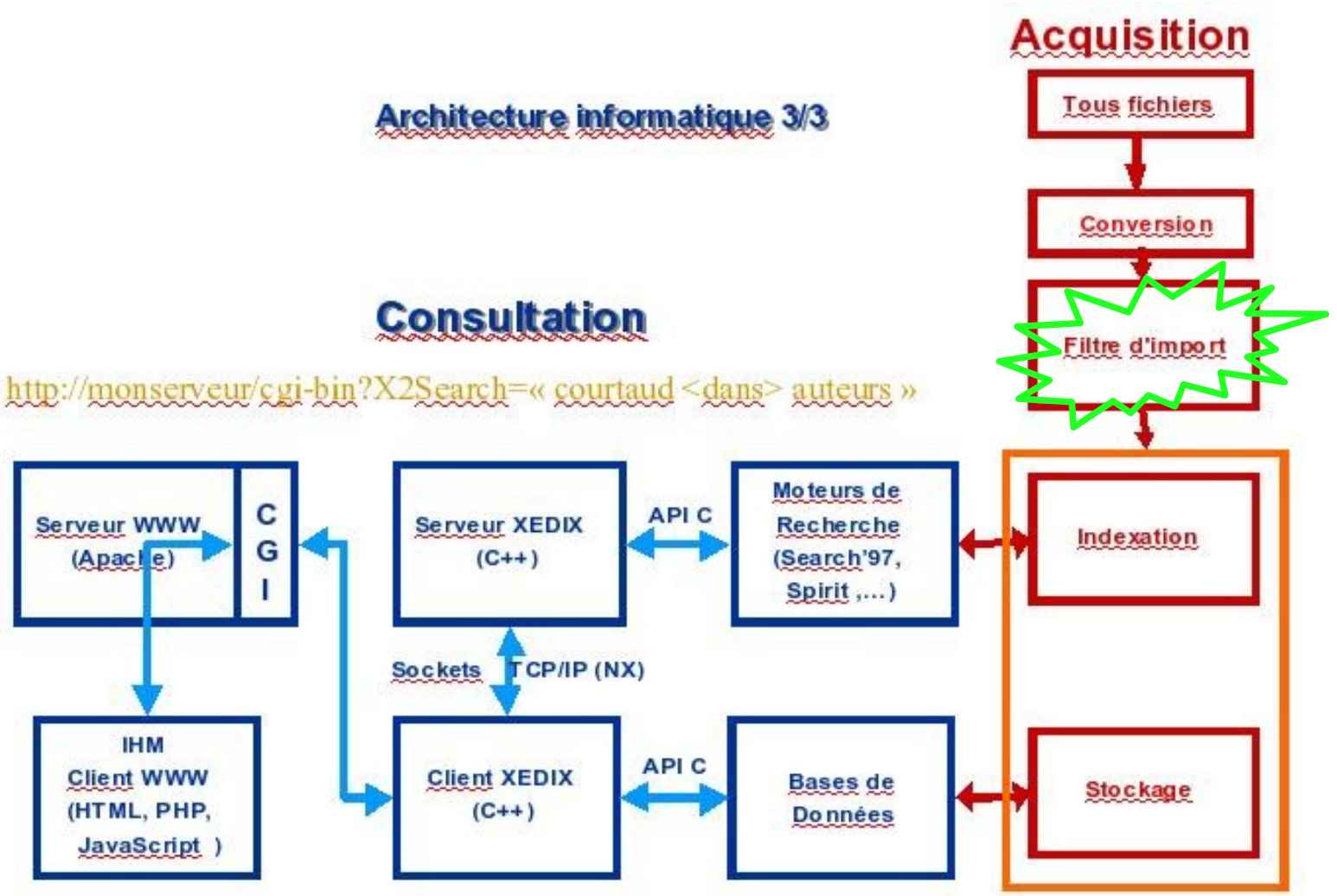
Architecture client-serveur



Client-serveur (II)

- Le client et le serveur de XediX
  - Dialoguent au travers d'un réseau par les sockets IP
  - En utilisant des protocoles propriétaires
- Il peut y avoir
  - Plusieurs clients pour un même serveur
  - Plusieurs serveurs sur la même machine
    - ✦Paramétrés sur des ports sockets différents
- Ils peuvent tourner sur des machines hétérogènes sur un réseau

Filtre d'import



Filtre d'import

- **Son rôle est de :**
  - Parser le fichier XML pour s'assurer de sa validité vis à vis de sa structure déclarée par DOCTYPE
    - Si le fichier est valide, on poursuit le traitement
    - S'il ne l'est pas, on arrête le traitement et on renvoit un message d'erreur : le fichier n'est pas importé dans la base
  - Préparer le stockage physique en éclatant le fichier d'entrée en élément XML séparés
  - Avertir le/les moteurs de recherche des éléments XML à indexer

Filtre d'import (II)

- Il s'appuie sur :
  - OpenSP 1.5 de James Clarck pour le parsing
    - Licence GPL
    - Se sert de SAX pour analyser le fichier
  - SAX
    - Pour analyser le fichier et le décomposer en éléments XML séparés
    - Déclencher les événements pour alerter les moteurs d'indexation

Filtre d'import (III)

- **Arguments du filtre importe**
  - `-remplace ou -ajout ou -reindex`
    - Importe en remplacant, ajoutant ou réindexant des documents
  - `-debug[0,2]`
    - Options de débogage
  - `-enracine`
    - Réalise l'enracinement après l'import (cf plus loin)
  - `-spirit/nospirit, verity/noverity, index/noindex`
    - Commande ou empêche l'indexation par les moteurs de recherche respectivement Spirit, Verity ou ReX
  - `-encoding`
    - Spécifie l'encodage utilisé par les fichiers XML importés

Filtre d'import (IV)

- Le filtre d'import est invoqué dans la procédure de remplissage d'import de la base `rempli_base` qu'il faut appeler à partir de son répertoire d'installation

- ./rempli_base -h
  - Donne la liste des paramètres possibles
  - Appelle les programmes d'indexation en fonction des options passées en paramètes

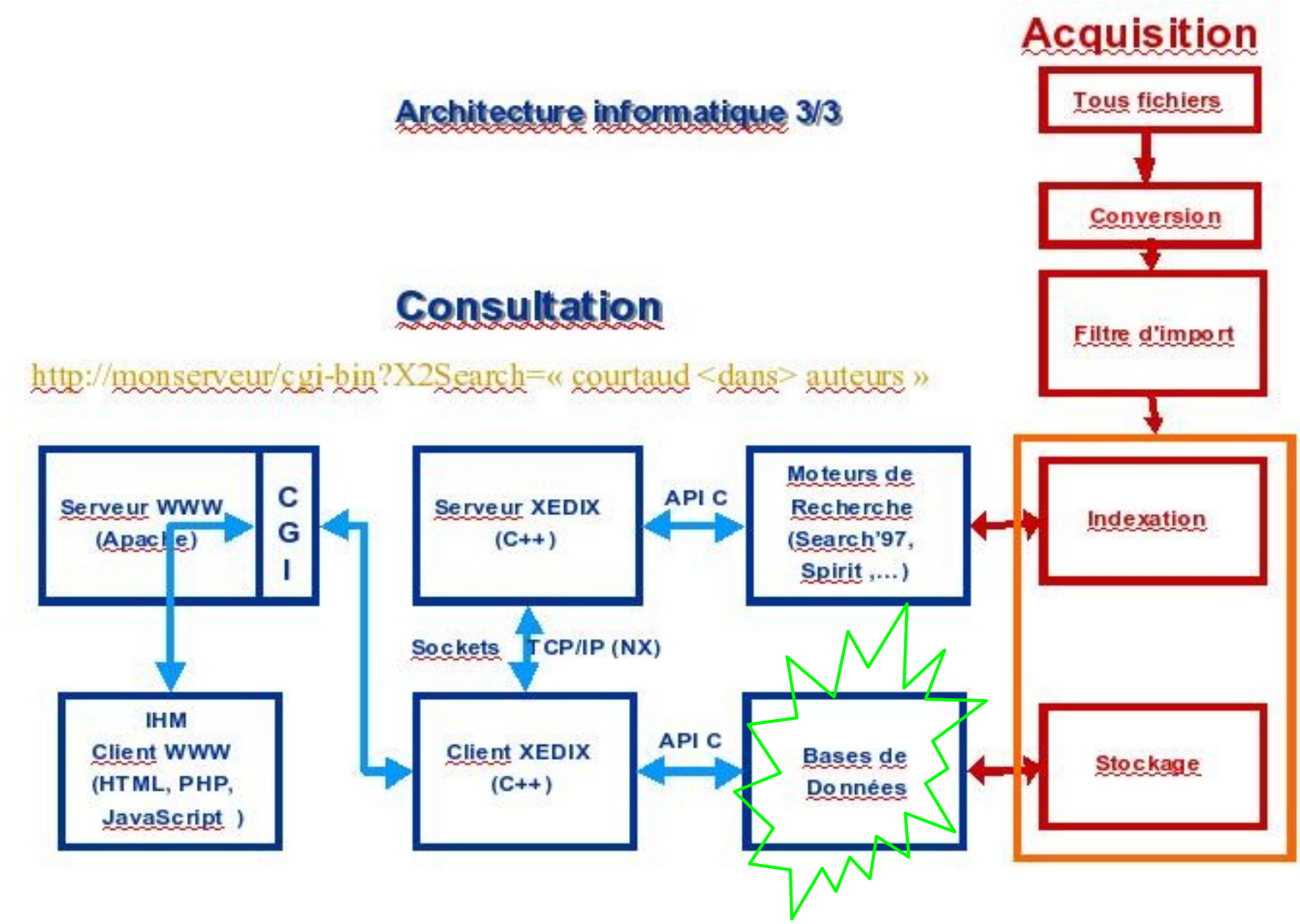
Filtre d'export

- **Rôle**
  - Exporter le/les documents de la base sous forme HTML/XML
  - En particulier, *dumper* la base pour la sauvegarder par exemple
- **XediX**
  - Garantit de restituer un document XML valide
  - Ne garantit pas de restituer un document strictement identique au document importé
- **Utilise l'architecture client-serveur de XediX**
  - Requête *CGI* : `X2Documents+nufonc+userdata+param`
    - `nufonc : type d'affichage`
    - `param : paramètres`
    - `userdata : paramètres d'authentification`

Filtre d'export (II)

- L'export peut se faire
  - Sous forme HTML pour visualisation
    - Le mapping XML -> HTML est défini
      - Soit par un paramétrage dans `dtds.ini`
      - Soit par une feuille de style XSLT attaché au fichier XML
  - Sous forme XML pour traitement
    - On précise dans l'export les balises XML que l'on veut exporter
    - Peut s'utiliser avec le sélécteur pour un export selectif des résultats de recherche

Gestionnaire de base physique



Gestionnaire de base physique (II)

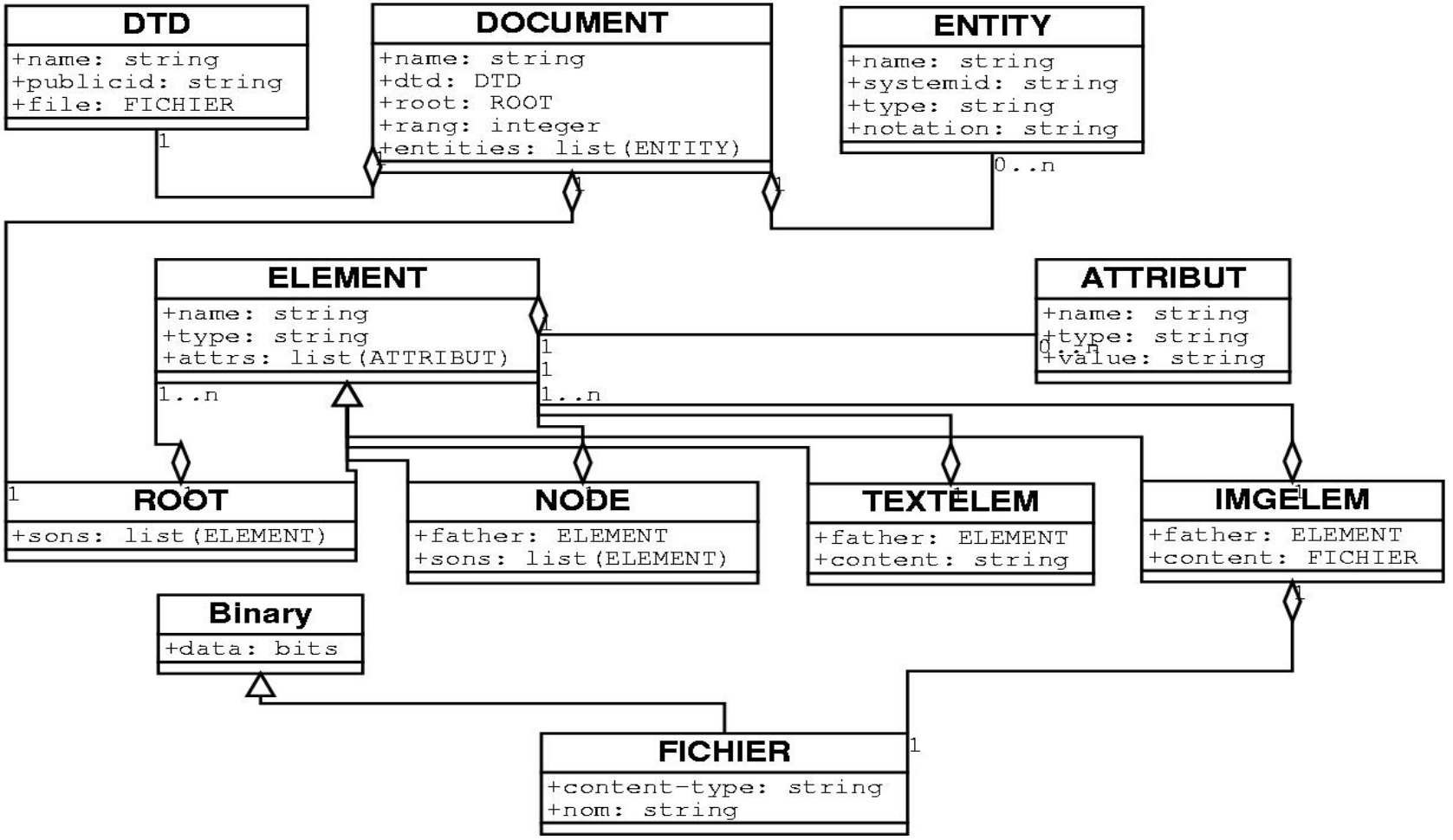
- C'est le composant qui va assurer le stockage physique des éléments XML résultant du parsing et de l'éclatement
- Composant essentiel
  - Pour les performances
    - De stockage
    - De requêtage
  - Différentiel des différentes solutions de NXDs

Gestionnaire de base physique (III)

- Quel schéma de base prendre ?
- Schéma dépendant de la DTD
  - Avantages
    - Améliore la preformance de requêtage
    - Facilite le stockage
  - Inconvénients
    - Nécessite de refaire le schéma pour chaque DTD
- Schéma générique
  - Avantages
    - Universel pour toutes les DTDs
  - Inconvénients
    - Moins optimisé

- XediX utilise un schéma générique

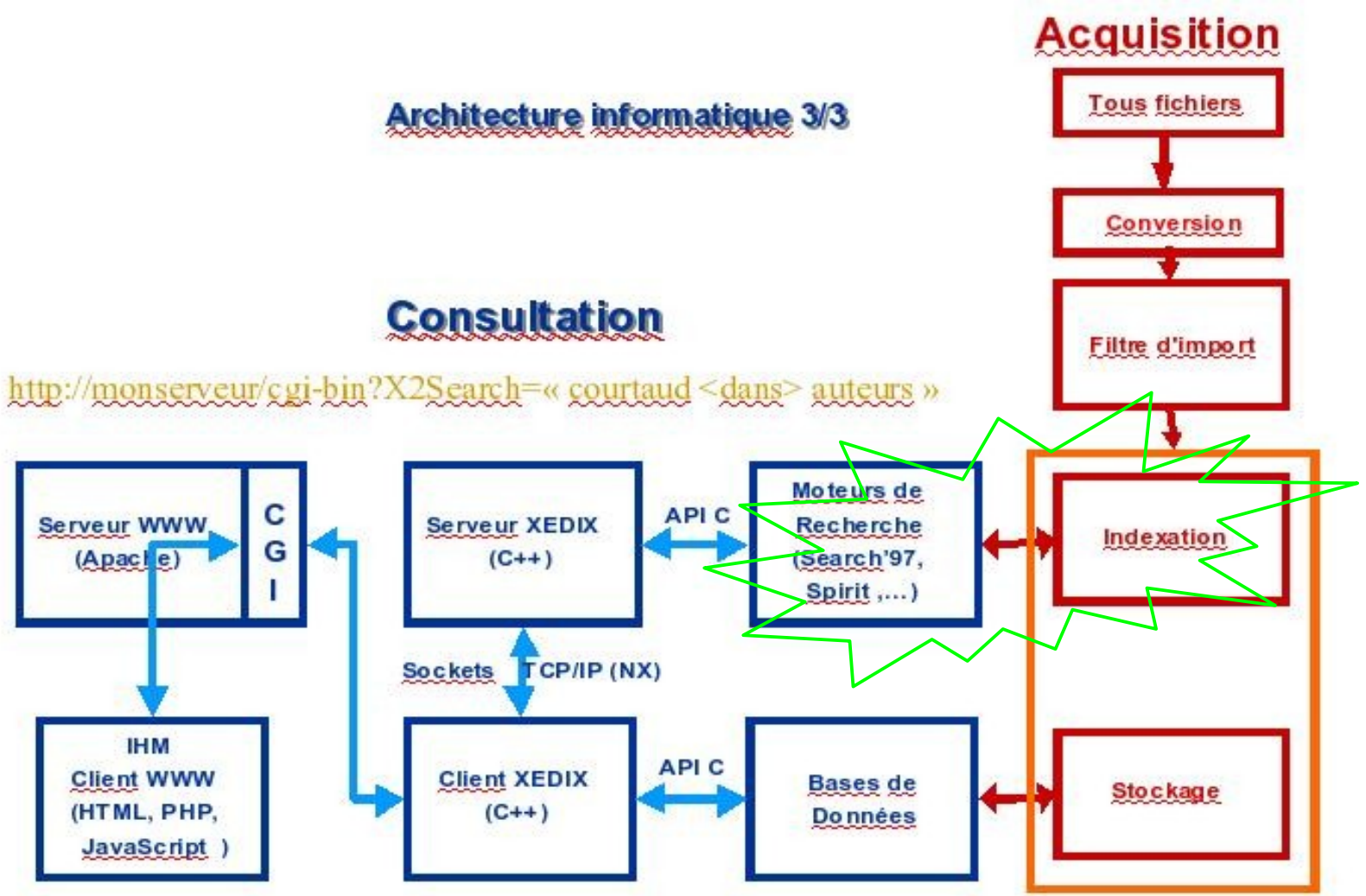
Diagramme de classe UML de la description générique d'un document dans Xedix



Gestionnaire de base physique (V)

- DTD ou schéma XML ?
- Intérêt des schémas XML
  - Ajoute une information de typage à l'information de structure
  - Mais en contrepartie fige le type d'un champ (élément XML)
- Intérêt des DTDs
  - Ne décrit que l'information de structure
  - La seule utile dans le cas documentaire
  - Permet de laisser l'application fixer le type du champ
  - Exemple : date
- XediX travaille avec des DTDs XML ou SGML

- XediX peut fonctionner avec quatre types de stockage (driver virtuel surchargeable)
  - Systèmes de fichiers journalisés
    - Exemple : ReiserFS ou ext3 sous Linux
  - Base hiérachique CLIO propriété du CEA
    - La seule permettant les performances en termes de volumétrie
  - Bases de données objets
    - Tourne au dessus de O2
  - SGBDR
    - Par le protocole ODBC
    - Tourne au dessus de MySQL


Architecture informatique 3/3
Acquisition
Tous fichiers
Conversion
Filtre d'import
Consultation
http://monserveur/cgi-bin?X2Search=« courtaud <dans> auteurs »
Serveur WWW (Apache)
CGI
Serveur XEDIX (C++)
API C
Moteurs de Recherche (Search'97, Spirit ,...)
Indexation
Sockets
TCP/IP (NX)
IHM Client WWW (HTML, PHP, JavaScript)
Client XEDIX (C++)
API C
Bases de Données
Stockage

Moteurs de recherche (II)

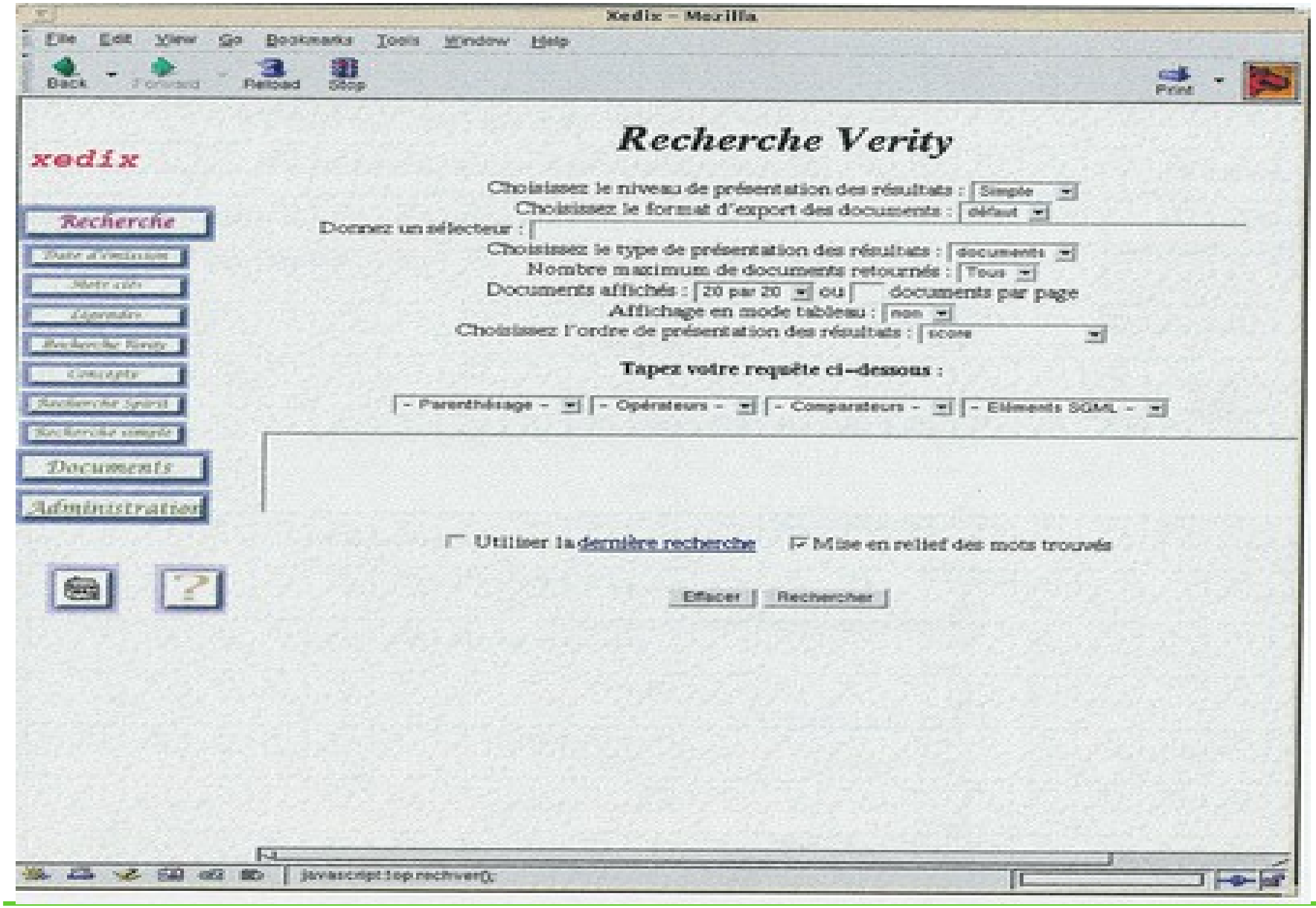
- **Rôle**
  - Alertés par le filtre d'import, ils indexent les documents importés
    - En retenant leur position absolue
    - En retenant leur position dans la structure XML
- Il existe dans Xedix une interface générique pour les moteurs de recherche surchargeable pour instancier facilement un nouveau moteur
- Trois moteurs sont actuellement interfacés avec XediX
  - Verity K2
  - Spirit
  - ReX

Moteurs de recherche (III)

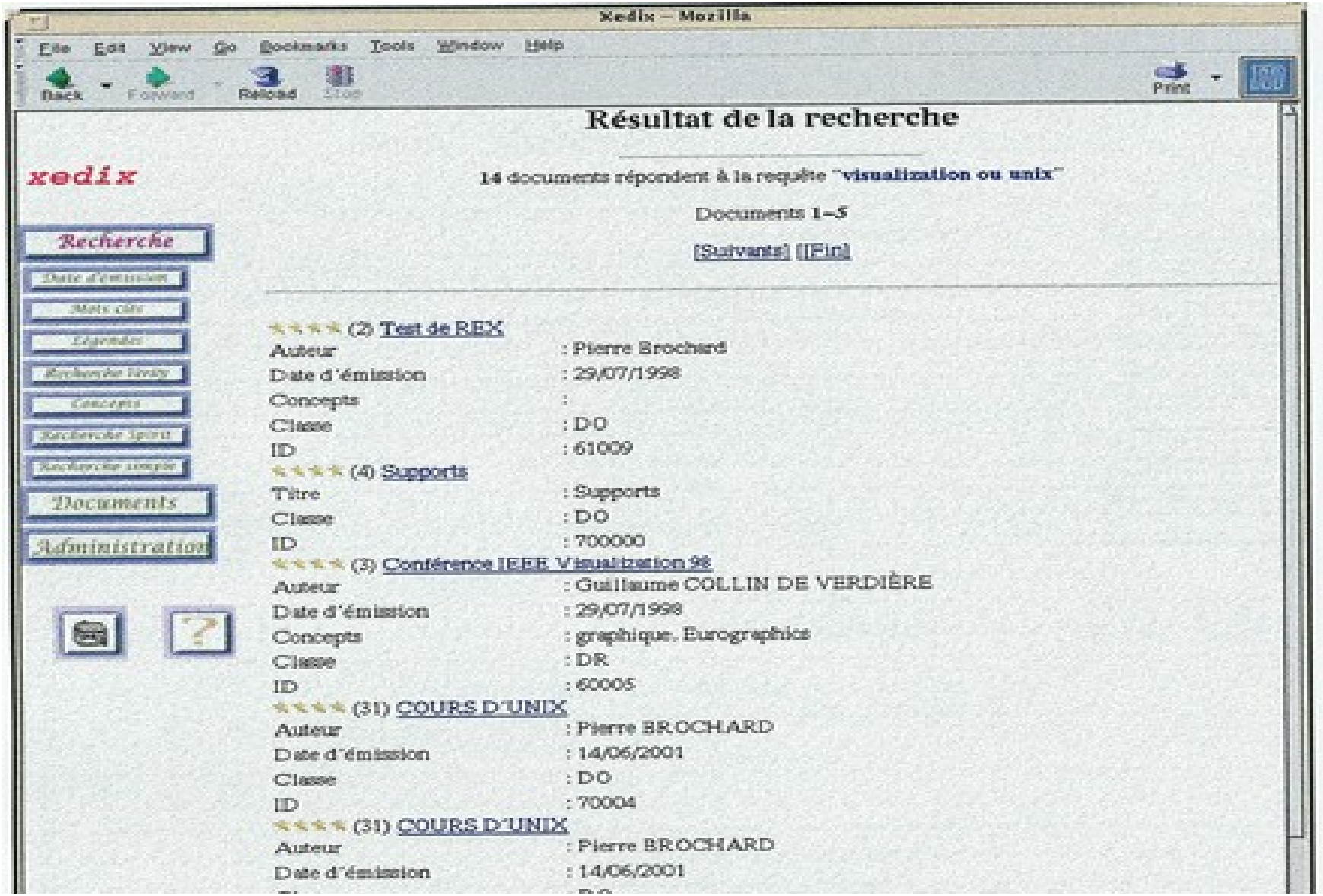
- **Verity K2**
  - Moteur statisque de Verity
  - Requêtage booléen
  - Permet d'indexer un nombre limité de champs XML
- **Spirit**
  - Moteur linguistique de Technologies SA sur brevet CEA
  - Permet un requêtage en langage naturel
  - Pas d'indexation de la structure
- **ReX (Regular Expressions for XML)**
  - Moteur créé par le CEA et fourni avec XediX
  - Requêtage booléen + Expressions régulières POSIX
  - Sélecteur associé (fonctionnalités proches de XPath)
  - Tous les éléments XML sont indexés

- **Requêtage "XML"**
  - Pour les moteurs le permettant, il sera possible de requêter en utilisant la structure XML
    - Un sélecteur permettra de sélectionner une liste de documents susceptibles de répondre à la requête (fonctionnalités identiques à Xpath)
    - La requête booléenne ou par expression régulière sera alors évaluée sur cette liste
  - Exemple
    - Cours [Unix,XediX] <DANS> Cours <DANS>Evry

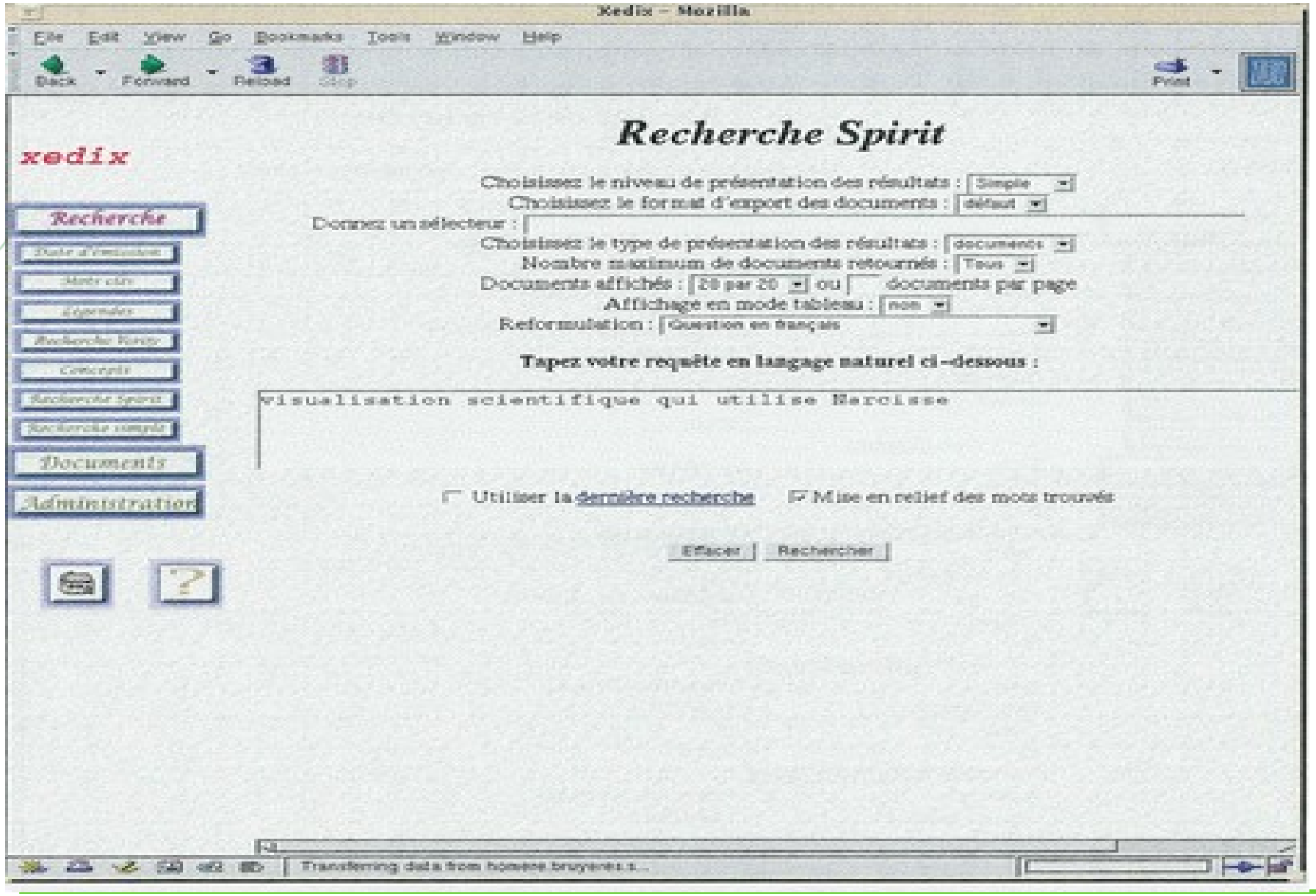
Moteurs de recherche (IV)



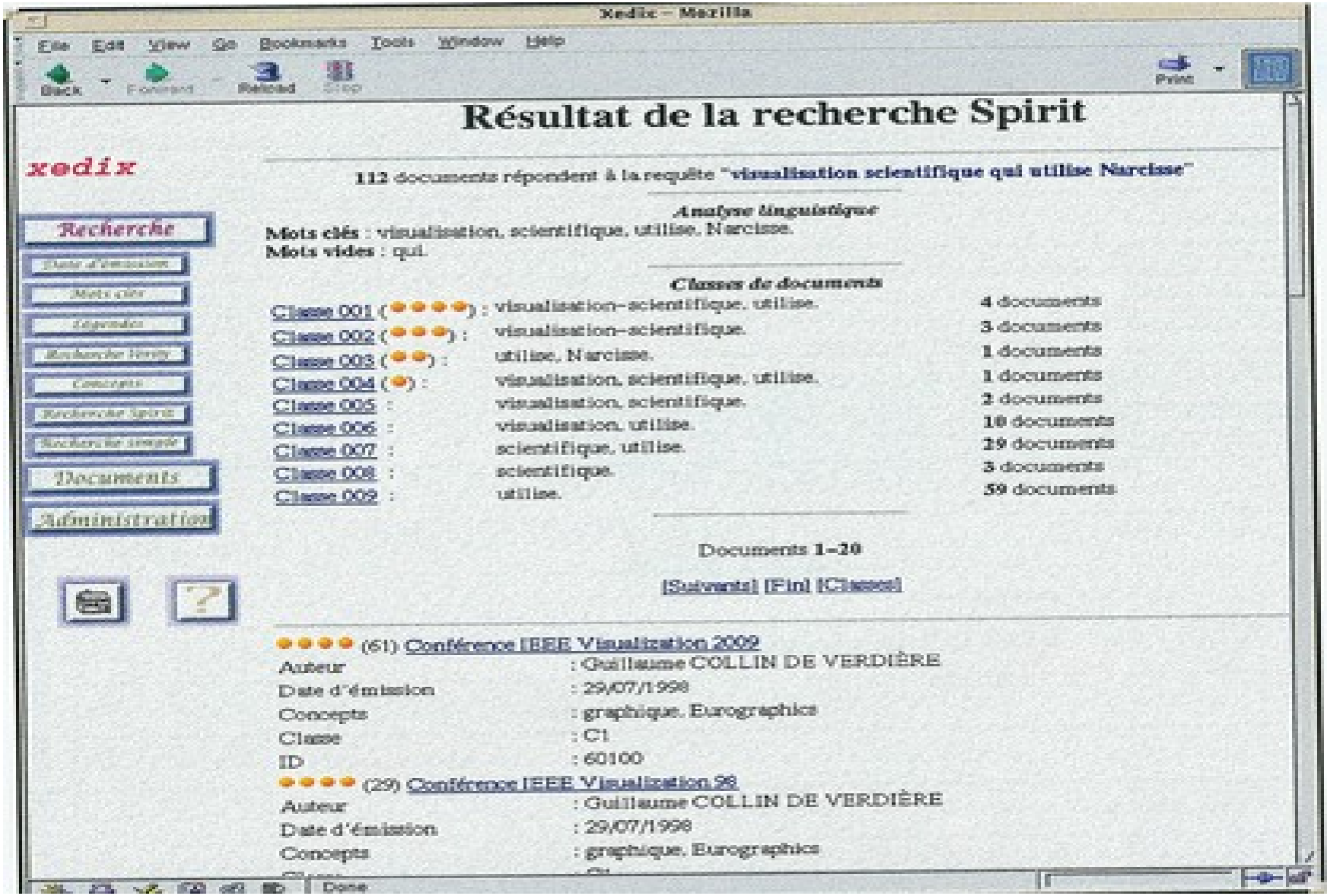
Moteurs de recherche (V)



Moteurs de recherche (VI)



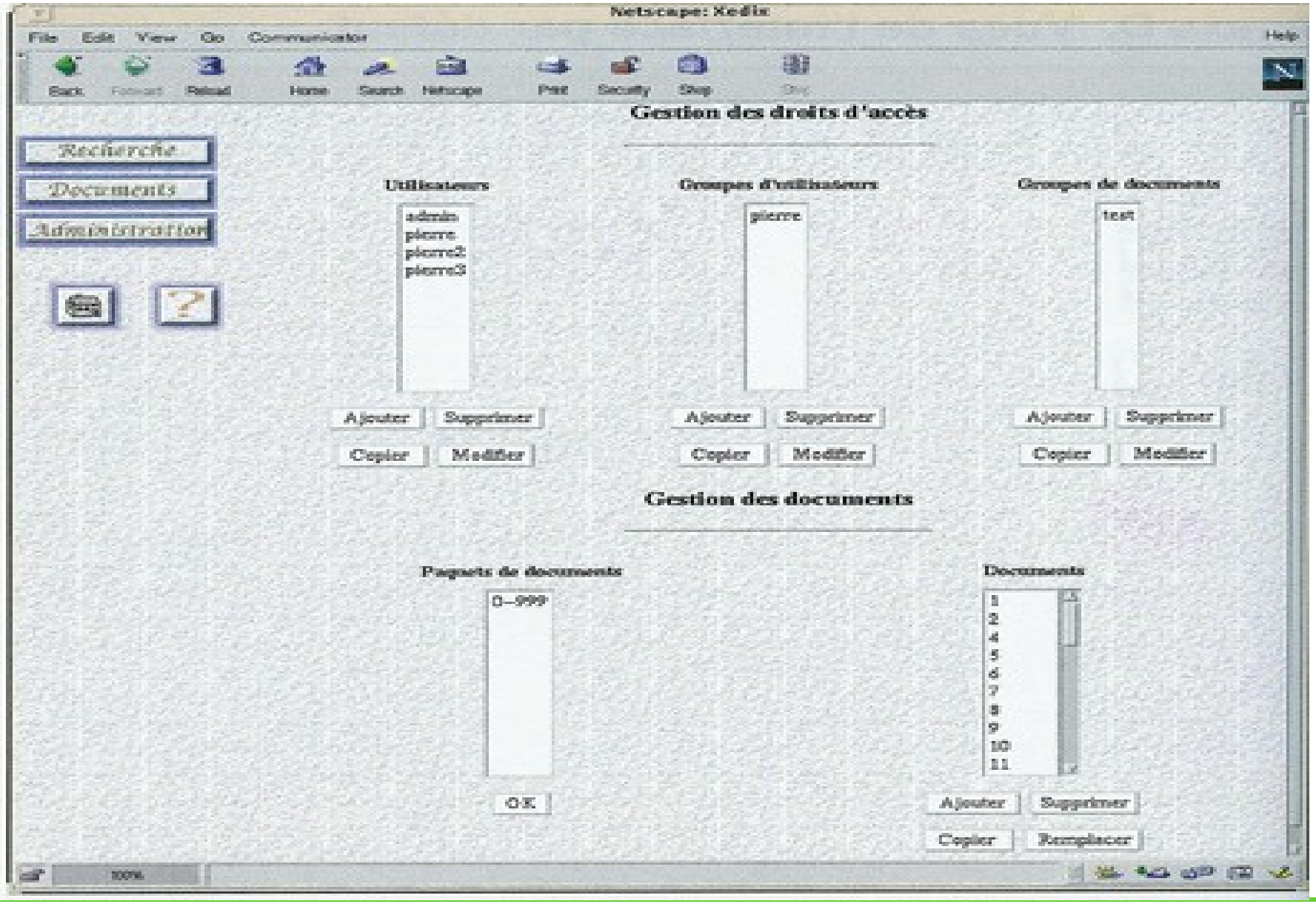
Moteurs de recherche (VII)



Gestion des droits d'accès

- Xedix incorpore un module permettant de gérer les droits d'accès
  - Aux documents
    - Par utilisateur
    - Par groupes d'utilisateurs constitués
  - Aux éléments XML individuels
    - Par utilisateur
    - Par groupes d'utilisateurs constitués
- Gestion par interface Web fenêtrée

Gestion des droits d'accès (II)



Interface de consultation

- Toutes les informations provenant de la base ou à destination de la base passent par l'intermédiaire du navigateur Web
  - En HTML
    - Selon le paramétrage de dtds.ini
  - En XML
    - Selon la feuille de style associée au document par un PI XSLT
      - <?xml-stylesheet ....
    - S'il n'y a pas de feuille de style associée, sortie sous forme de raw XML visualisé avec la feuille par défaut du navigateur

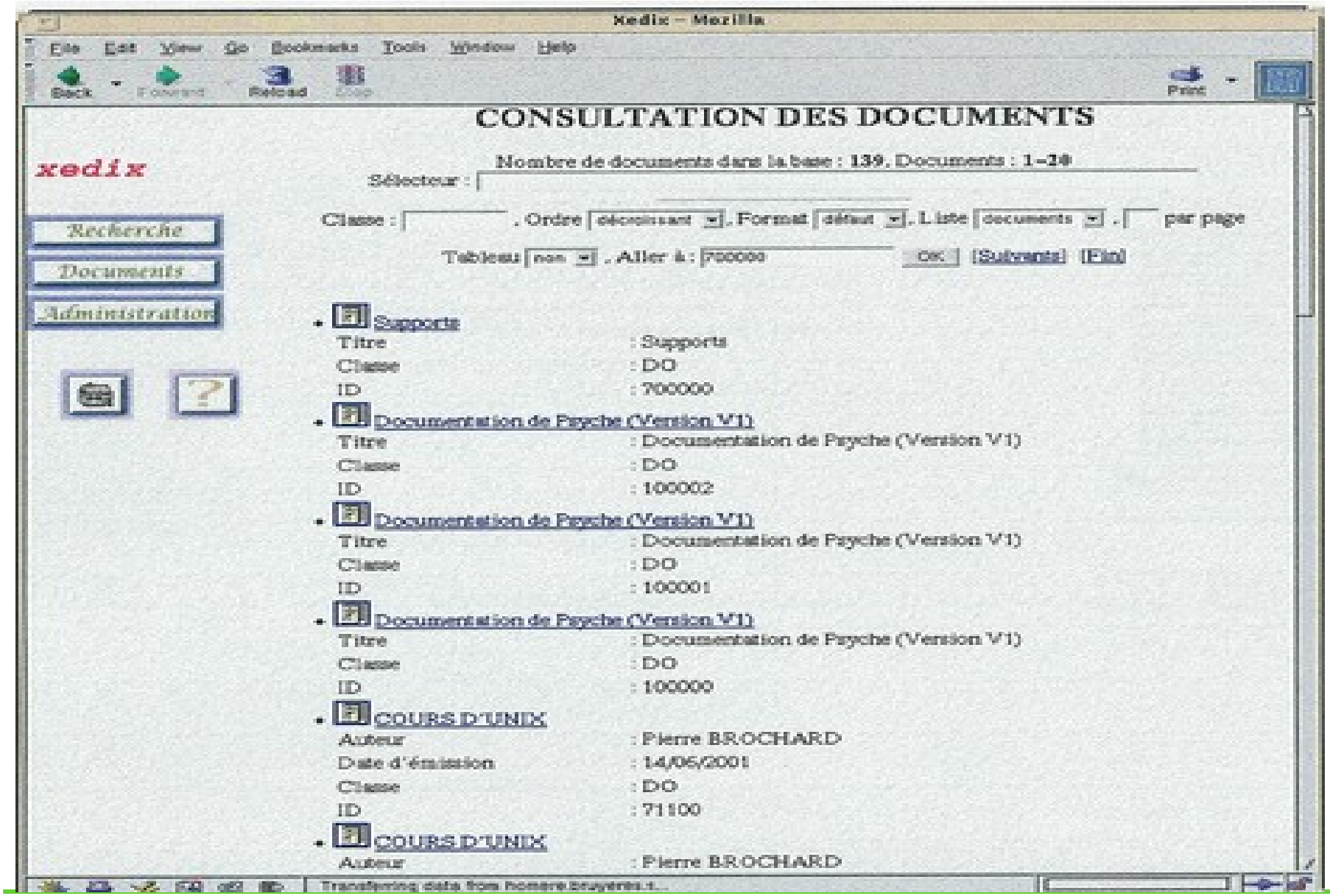
Interface de consultation (II)

- Il existe différentes formes de visualisation des résultats de recherche :
  - Document entier
  - Document avec table des matières clicable (enracinement)
  - Portions de documents (délimitées par une ou des balises XML)
  - Liste de documents
  - Raw XML
  - ...

Interface de consultation (III)

- Toutes les visualisations sont réalisées
  - En HTML 4.01
  - Javascript
- Les exports XML sont conformes à XML 1.0 et valides
- L'interface de XediX est donc indépendant du navigateur Web utilisé
  - Vérifié pour
    - Internet Explorer
    - Mozilla
    - Firefox
    - Opera

Interface de consultation (IV)



Interface de consultation (V)

Personnalisation de l'interface

- Du fait de leur conformité aux standards du Web, toutes les pages générées par Xedix sont paramétrables par des feuilles de style XSLT et/ou CSS
  - Relookage de l'interface
  - Patrons de requête
  - Patrons de réponse
- Il existe dans `dtds.ini` un certain nombre de paramètres agissant sur l'interface
  - Liste des paramètres à afficher dans les listes de documents

Gestion des rétroconversions

- Tous les documents ne sont pas forcément au format XML ni même électroniques
- Nécessité d'une chaîne de traitement pour les documents papier
  - Numérisation du document (=> image)
  - Reconnaissance de caractères (=> fichier textuel)
  - Reconnaissance de la structure logique (=> XML)
- Certaines applications nécessitent de conserver le papier pour des raisons légales
  - Nécessité d'avoir deux versions
    - XML pour les recherches
    - Image pour la preuve

Gestion des rétroconversions (II)

- XediX permet de gérer ces deux versions de façon synchrone
  - Les images des pages sont stockées dans la base
  - Les fichiers XML sont indexés normalement
    mais avec un PI XML indiquant les changements de page physique
- A la consultation, lorsqu'on récupère un document XML de ce type
  - Il y a un lien vers la page image correspondante à chaque page physique
  - Il ouvre sur un visualiseur d'image qui se positionne directement à la page indiquée

- XediX est particulièrement bien adapté à la gestion de documents multimédia
- Il peut en effet stocker et indexer des documents multimédia de deux façons
  - Soit en associant aux séquences multimédia des métadonnées XML qui serviront au reprérage des séquences une fois indexées
  - Soit utiliser les métadonnées présentes dans les formats audiovisuels les plus récents et généralement codés en XML
    - MPEG4
    - Et surtout MPEG7

Gestion du multimédia (II)

- Une des particularités de XediX est que les vidéos sont physiquement stockés DANS la base et non pas à l'extérieur
  - Sécurisation des flux vidéos
  - Facilité d'indexation
  - Facilité de déplacement de la base
- Grâce à l'indexation, on peut adresser individuellement un plan particulier de la vidéo et communiquer ce plan à un player externe
  - Utilisation de l'en-tête HTTP " Range : "
  - Permet de sélectionner une plage du flux demandé

Utilitaires XML

- Il existe autour de XediX un ensemble d'utilitaires permettant la conversion ou la transformation de documents vers XML
  - Documents bureautiques => XML
  - Nettoyage de code HTML
  - Nettoyage de code XML
  - Heuristiques d'XMLisation de documents papier
  - Possibilité d 'import off-line par lot de grands volumes
  - ...

API

- Toutes les opérations que nous venons de voir peuvent également être réalisées par une application tierce dialoguant avec XediX au moyen de son API
- Cette API est celle du client XediX avec son serveur
- Elle s'appuie sur HTTP et pourrait être facilement ré écrite en SOAP
- Les arguments peuvent être envoyés
  - Soit par la méthode GET
  - Soit par la méthode POST

- Les arguments se présentent sous la forme générique
  - `racine+nufonc+userdata+param`
  - où
    - `racine` est une chaine de caractères caractérisant la nature de l'opération
    - `nufonc` est un numéro précisant la fonction demandée
    - `userdata` est une chaine de caractères identifiant l'utilisateur
    - `param` est une liste de paramètres dépendant de la fonction choisie

API (III)

- Les fonctions possibles sont :
  - X2Documents
    - Pour l'export HTML / XML
  - X2Admin
    - Pour le contrôle d'accès
  - X2Accueil
    - Pour une liste de documents
  - X2Search
    - Pour appeler les moteurs de recherche
  - ...
- Exemple d'appel
  - Les 20 premiers documents de la liste

Performances de la base

- L'organisation d'une NXD permet de mettre en place des services nouveaux
  - Création de documents virtuels
    - Documents physiques qui encapsulent des appels (via l'API de XediX) aux moteurs de recherche
    - La requête est évaluée lors de la consultation du document et son résultat remplace automatiquement la requête
  - Synthèse automatique
    - En s'appuyant sur un requêtage ciblé
- Cela suppose de vérifier les performances atteintes par la base dans un environnement de production

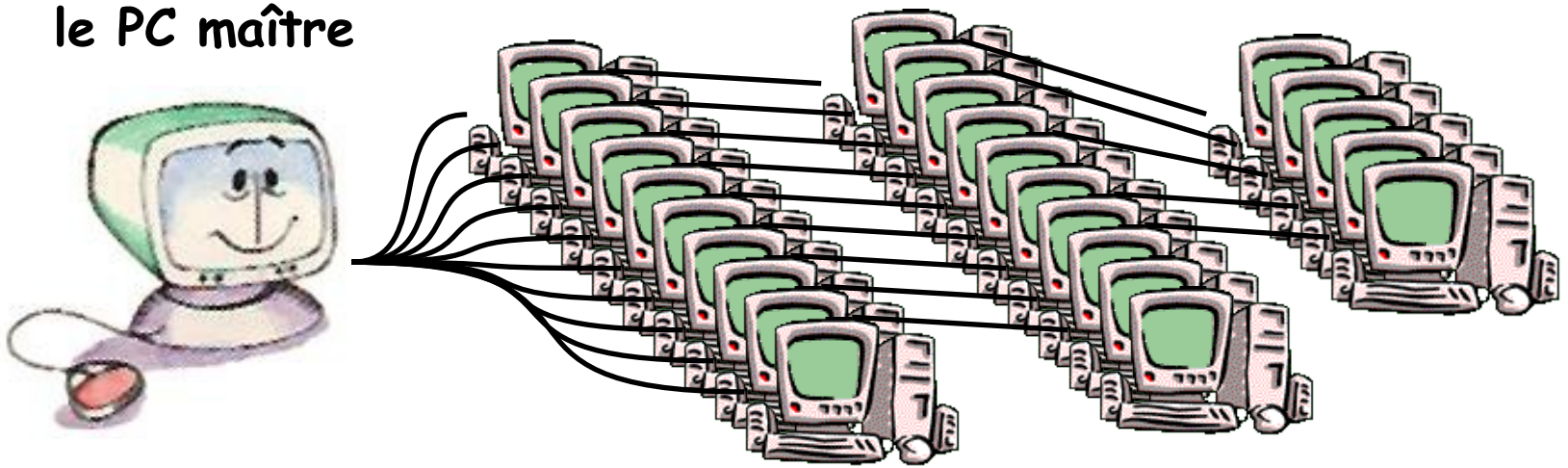
Le projet XTera

- Projet de benchmark pour comparer les performances de XediX à celles de ses principaux concurrents
- Alors que la plupart des implémentations actuelles des NXDs saturent au voisinage de 10 Go, nous voulons démontrer que XediX peut gérer un volume de 1 To de documents XML !
- Mesures
  - des performances d 'import
  - des performances d 'indexation
  - des performances d 'export

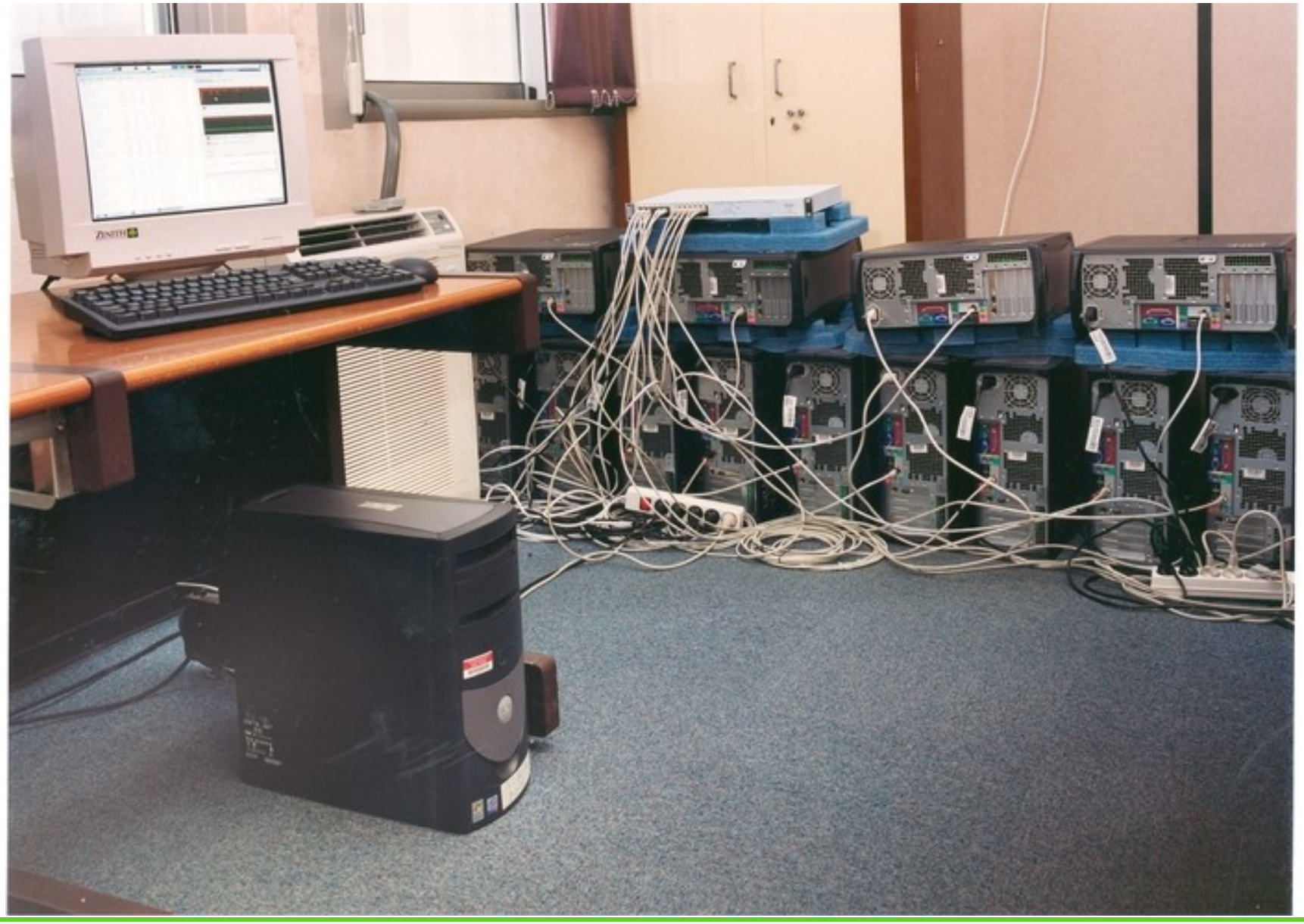
Le projet XTera (II)

- Nécessité de prendre des cas tests reconnus au niveau international
  - Xmach
    - http://dbs.uni-leipzig.de/en/projekte/XML/XmlBenchmarking.html
  - The Michigan Benchmark (Mbench) :
    - http://www.eecs.umich.edu/db/mbench/
  - XBench
    - http://db.uwaterloo.ca/~ddbms/projects/xbench/
  - XMark :
    - http://monetdb.cwi.nl/xml/index.html
  - XOO7 :
    - http://www.comp.nus.edu.sg/~ebh/XOO7.html

- **Architecture du test**
  - 25 PCs de 40 Go + 1 PC maître reliés par un switch
    - Chaque PC esclave tourne un système Linux *minimal*
    - Le PC maître tourne une distribution Linux Slackware standard
  - LVM (Logical Volume Manager) + eNBD (extended Network Block Device) nous permettent de voir tous les disques individuels des PCs esclaves comme un seul disque de 1 To sur le PC maître



Le projet XTera : montage réel



- **Résultats obtenus**
  - La base a importé un peu plus de 1 To sans aucun problème
  - Performances de requêtage
    - <u>Le temps de réponse à une requête est indépendant de la taille de la base et ne dépend que du nombre total de documents retournés</u>
    - Le temps de réponse moyen d'une requête sur la structure est de moins de 10 secondes

Articles dans :

- Le Monde Informatique
- 01 Informatique

XML ● Cinq secondes pour retrouver un document dans une base de 1 Téraoctet : c'est l'exploit réalisé par XediX, la base de données native XML développée par le CEA, dont les performances surpassent largement celles de tous les outils équivalents du marché. AM2 Systems, une start-up de l'INRIA, vient de lancer la commercialisation. Premiers utilisateurs visés : l'archivage, les médiathèques, les projets coopératifs de recherche ; et à terme, tous les gestionnaires de grandes bases de données.

Base de données XediX

Cinq secondes pour retrouver une aiguille dans une botte de foin

POINTS FORTS

- Base de données multimédia sous Unix.
- Cinq secondes pour une requête simple sur 1 Téraoctet de données.
- Base native XML, indépendante des logiciels d'origine des données.
- Détermination et modification libre des structures de données.
- Gestion intégrée et ultra-fine des droits d'accès.

Portage sur architecture parallèle

- XTera a démontré
  - La robustesse et les performances de la base sur un grand volume
  - Que cette performance peut être atteinte avec du matériel standard
- Cible visée : Archivage des PME-PMIs
- Le portage sur architecture parallèle vise à démontrer
  - Que XediX peut tirer parti d'une architecture parallèle distribuée pour augmenter ses performances en
    - Volumétrie
    - Temps d'indexation
- Cible visée : Grands centres de calcul ou d'archivage

Portage : principes

- **Principes**
  - Import et indexation des documents
    - Division et séparation *étanche* du domaine des documents
    - Affectation à chaque processeur d'un sous-ensemble des documents
  - Consultation et requêtage
    - Un serveur XediX (et son serveur Web) est affecté à chaque processeur et contrôle son sous-ensemble de documents
    - L'un de ses serveurs (n'importe lequel) est choisi pour être le serveur de syndication qui dialoguera
      - Avec les autres serveurs pour leur demander les documents ou les fragments de documents
      - Avec l'IHM pour prendre la requête et afficher les résultats

- La deuxième expérimentation Xtera vise à prouver :
  - Que les résultats obtenus sur Xtera 1 sont extrapolables à de plus grands volumes
  - Que le développement du méta serveur de syndication de XediX
    - L'autorise à exploiter les possibilités des architectures parallèles en terme de performance
    - Prouve la scalabilité du produit quant au nombre de noeuds
  - Que la version 2 du moteur de recherche ReX permet un gain important à l'indexation tant au niveau du temps que de la place disque
  - D'atteindre la barre symbolique des 10 To (10 x disque dur grand public)

Le projet Xtera10 (II)

- Expérimentation réalisée en partenariat avec Bull sur le cluster TeraNova de Bull installé en zone Ter@tec près du Centre CEA DAM Île-de-France
  - 300 processeurs Pentium Itanium 2
- La base a importé 10 Téraoctets de documents XML sans aucun problème
  - Tests de requêtage en cours

- Annonce dans la presse dans quelques semaines

Conclusion

- XediX Tera Solution™ est une implémentation mature du concept de NXD pour gérer des données /documents
  - Permettant une gestion aisée de très grands volumes de données
  - Offrant de nombreux services pour une utilisation industrielle de documents ou de données au format XML
  - Pouvant facilement se coupler à d'autres produits
- Elle offre de nouvelles solutions en terme de
  - Volumétrie : jusqu'à 100 To et au dela
  - Sécurisation : de l'élément XML jusqu'aux vidéos
  - Gestion du multimédia : indexation fine des vidéos
- Bientôt "dans les bacs"
  - Création de la société "XediX Tera Solution" en cours