

# Systèmes d'information documentaires distribués

Yannick Prié  
UFR Informatique  
Université Claude Bernard Lyon 1

2008-2009 – Master SIB  
M1 – UE 3 / Bloc 4 – Cours 6-7

## Systèmes d'information documentaires distribués

- Systèmes d'information
  - Organisation : fonctionnement / membres
  - Soutien à l'activité de l'organisation
- Systèmes d'information documentaire
  - Echange de documents
  - Recherche de documents
  - Stockage de documents
  - Utilisation de documents
- Systèmes distribués
  - Réseau
  - Systèmes et machines
  - Hétérogénéité

Master SIB M1 – SIB 3 : traitement, exploitation de l'information et systèmes d'information  
Bloc 4 : architecture et fonctionnement des systèmes d'information documentaires

CM6-7 : SI documentaires distribués  
Yannick Prié – 2008/2009

2

## CM6-7 : Systèmes d'information documentaires distribués

- Objectifs du cours
  - Présenter plusieurs façons de considérer un SID suivant différentes analyses. Point de vue utilisateur, concepteur, installateur, acheteur, réseau, etc.
  - Etre capable de comprendre la description d'un système d'information documentaire, de tous ces points de vue.

Master SIB M1 – SIB 3 : traitement, exploitation de l'information et systèmes d'information  
Bloc 4 : architecture et fonctionnement des systèmes d'information documentaires

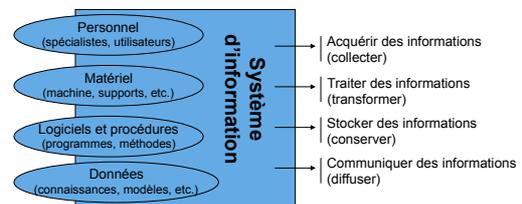
CM6-7 : SI documentaires distribués  
Yannick Prié – 2008/2009

3

(Reix, 2004, pp. 3-4)

## Système d'information

*Un SI est un ensemble organisé de ressources : matériel, logiciel, personnel, données, procédures... permettant d'acquérir, de traiter, de stocker des informations (sous formes de données, textes, images, sons, etc.) dans et entre des organisations.*



Master SIB M1 – SIB 3 : traitement, exploitation de l'information et systèmes d'information  
Bloc 4 : architecture et fonctionnement des systèmes d'information documentaires

CM6-7 : SI documentaires distribués  
Yannick Prié – 2008/2009

4

(Reix, 2004, p. 50)

## C'est quoi une organisation ?

*L'organisation est à la base de l'action collective. Dès qu'une activité dépasse la capacité d'un seul individu, l'organisation constitue la réponse appropriée. Elle se caractérise donc par :*

- un ensemble d'individus : participants, acteurs ;
- un accord, implicite ou explicite, sur un ou plusieurs objectifs partagés par les divers participants ;
- une division du travail, définissant le rôle de chaque participant ;
- une coordination plus ou moins formalisée, qui assure la cohérence des comportements et donc le respect des objectifs communs en dépit de la division du travail.

Master SIB M1 – SIB 3 : traitement, exploitation de l'information et systèmes d'information  
Bloc 4 : architecture et fonctionnement des systèmes d'information documentaires

CM6-7 : SI documentaires distribués  
Yannick Prié – 2008/2009

5

(Morley, 2005, p. 71)

## Que fait-on dans une organisation ?

- Processus principaux
  - résultat = raison d'être de l'organisation
  - Ex. : production de biens ou de services
- Processus secondaires
  - résultats nécessaires pour l'exécution des processus principaux
  - Ex. : comptabilité, paye
- Processus de pilotage
  - contrôle de l'atteinte des objectifs, mise en œuvre de la stratégie
  - Ex. qualité

Master SIB M1 – SIB 3 : traitement, exploitation de l'information et systèmes d'information  
Bloc 4 : architecture et fonctionnement des systèmes d'information documentaires

CM6-7 : SI documentaires distribués  
Yannick Prié – 2008/2009

6

## Objectifs d'un SI

*Le but de tout système d'information est d'apporter un soutien aux processus de travail dans l'organisation selon trois modalités principales (pouvant être combinées) : fournir de l'information, assister le travail humain, automatiser le travail. Les systèmes de travail qu'il assiste peuvent être individuels [...] ou collectifs [...]. L'usage du SI peut être obligatoire, recommandé ou discrétionnaire... mais dans tous les cas, les caractéristiques du SI sont déterminées par ses finalités, c'est à dire par les objectifs de l'organisation auxquels il doit contribuer.*

## Usages des systèmes d'information

- Applications fonctionnelles
  - Applications « classiques »
  - Gestion commerciale, gestion de production, gestion comptable et financière
    - paye, facturation, stocks
- Applications d'aide à la décision
  - Assistance au décideurs (systèmes experts, entrepôts de données, datamining)

## Usages des systèmes d'information

- Applications d'aide à la communication
  - Communication interne
    - travail de groupe (collectif, CSCW)
  - Communication externe
    - SI inter-organisation (B2B, EDI, SCM) : entreprise étendue, entreprise étendue
    - SI de gestion de la relation client (B2C, CRM)
- Applications d'aide à la gestion des connaissances
  - Acquisition, conservation, diffusion des savoirs et des savoir-faire (KM)

## Les trois dimensions d'un SI

- Informationnelle
  - le SI produit des représentations, manipule et produit de l'information
- Technologique
  - le SI est un construit à base d'outils, utilise les technologies de l'information
  - Remarque : système d'information ≠ système informatique
- Organisationnelle
  - le SI est un élément des processus et de la structure de l'organisation

## Niveau organisationnel

- Différents degrés d'intervention du SI
  - SI comme source d'information externe
  - SI comme outil interactif mobilisable dans l'activité
  - SI intégré dans le système de travail
  - SI = système de travail
- Bref
  - le SI informe des processus fonctionnels
    - besoin des processus eux-mêmes
    - besoins de communication entre processus (coordination)
  - le SI structure et intègre des systèmes de travail

## Bilan

- Système d'information
  - élément constitutif de la structure de l'organisation
- Mise en place d'un SI = choix organisationnel et technologique
  - choix de partage de l'information
  - degré d'intégration du SI
  - technologies de communication

## Bilan (suite)

- Résultat technologique issu d'un double processus de construction
  - Délibérée : conception et implantation dans l'organisation
    - génératrice de règles, de contraintes
    - offre des ressources supplémentaires (automatisation, nouvelles présentation d'information)
    - changements prévus
  - Emergente : appropriation de la technologie, assimilation, détournement
    - le résultat est indéterminé (impossible à prévoir)
    - changements imprévus

## Métiers autour d'un SI

- Du point de vue de l'organisation, le SI
  - comme instrument de gestion
    - métiers = utilisateurs
    - membres de l'organisation / autres organisations / public
  - comme objet à gérer
    - métiers = gestionnaires
    - concepteurs, décideurs, informaticiens, ...

## Décomposition des éléments logiciels d'un SI

- Stockage
  - Bases de données
- Calculs
  - Traitements métier
- Interaction utilisateur
  - Interfaces graphiques

## Éléments des SIR : Base de données / SGBD

- Base de données
  - Au sens le plus général
    - Ensemble de données quelconques : mémoire, fichiers, web, etc.
  - Au sens plus strict
    - Ensemble de données
      - fortement structurées
      - persistantes
      - structure définie dans un schéma
      - gérées par système de gestion de bases de données
- Système de Gestion de Bases de Données (SGBD)
  - Logiciel spécialisé pour la gestion de base de données
  - En anglais : DBMS

## SGBD relationnels : tables

- On gère les données dans des tables (ou *relations*)
  - ensemble de champs (attributs) avec un certain domaine
    - oeuvre = [ Titre (chaîne max. 100 caractères), Date 1<sup>ère</sup> parution (entier > 0) ]
- Les tables contiennent des enregistrements
  - ensemble de valeurs pour les attributs d'une table
    - [ « Notre Dame de Paris », 1831 ]
- On identifie les enregistrements avec des clés primaires
  - Valeur unique pour chaque enregistrement
    - oeuvre = [ Identificateur (entier unique), Titre (chaîne max. 100 caractères), Date 1<sup>ère</sup> parution (entier > 0) ]
    - [ 1264, « Notre Dame de Paris », 1831 ]
    - [ 1298, « Ruy Blas », 1838 ]

## SGBD : tables

Table Oeuvre		
Identificateur	Titre	Date 1 <sup>ère</sup> Parution
1	La confession d'un ...	1836
...	...	...
1264	Notre Dame de Paris	1831
...	...	...
1298	Ruy Blas	1838
1299	Faust	1808
...	...	...

## SGBD relationnels : ensembles de tables

- On crée des relations entre enregistrements à l'aide de clés secondaires
  - lien vers un enregistrement
    - Auteur = [ Identificateur (entier unique), Nom (chaîne max. 100 caractères), Prénom (chaîne max. 100 caractères), ]
      - [ 12, « Hugo », « Victor » ]
      - [ 13, « de Musset », « Alfred » ]
    - Oeuvre = [ Identificateur (entier unique), Titre (chaîne max. 100 caractères), Date 1<sup>ère</sup> parution (entier > 0), Auteur (clé secondaire table auteur) ]
      - [ 1264, « Notre Dame de Paris », 1831, 12 ]
      - [ 1298, « Ruy Blas », 1838, 12 ]

## SGBD : ensemble de tables

Id	Titre	Date1ere	Auteur	Type
1	La confession d'un ...	1836	13	1
...	...	...	...	...
1264	Notre Dame de Paris	1831	12	1
...	...	...	...	...
1298	Ruy Blas	1838	12	2
1299	Faust	1808	5	2
...	...	...	...	...

Id	Nom	Prénom
...	...	...
12	Hugo	Victor
13	De Musset	Alfred
...	...	...

Id	Libellé
1	Roman
2	Pièce
3	Recueil de poésie

Schéma relationnel =  
définition des tables et les liens

## SGBD : langage SQL

- Structured Query Language
  - Langage de gestion de BD normalisé
    - Gérer le schémas
      - Ajouter, supprimer, modifier des tables
    - Gérer les données
      - Lire, ajouter, supprimer, modifier des enregistrements
  - Forme générale pour lire les enregistrements
    - SELECT ... FROM ... WHERE ...
- Données à récupérer (génération d'une table résultat)      Sources de données (tables)      Conditions

## Exemples de requêtes SQL

```
SELECT Titre, Date1ere
FROM Oeuvre
WHERE Auteur = 12 ;
```

Titre	Date1ere
Notre Dame de Paris	1831
Ruy Blas	1838

```
SELECT DISTINCT Nom, Prénom
FROM Oeuvre, Auteur, TypeOeuvre
WHERE Oeuvre.Auteur = Auteur.Id AND Oeuvre.Type =
TypeOeuvre.Id
AND TypeOeuvre.Libellé = « Roman »
```

Nom	Prénom
Hugo	Victor
Von Goethe	Johann Wolfgang
...	...

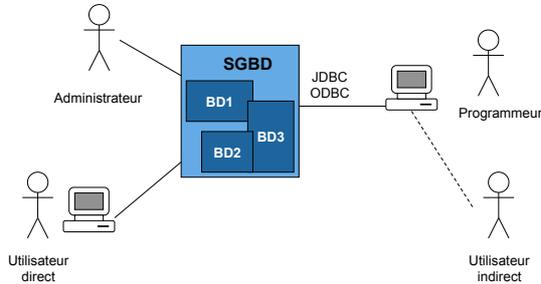
## SGBD : fonctionnalités et propriétés

- Administration centralisée / partage
  - Gestion de plusieurs bases de données
  - Client / serveur
- Indépendance à l'application
  - on ne gère que les données
- Rapidité
- Fiabilité
  - Récupération des erreurs

## SGBD : fonctionnalités et propriétés

- Limitation de la redondance
  - Une information à un endroit et un seul !
  - Notion de normalisation
- Vérification de l'intégrité
  - Contraintes sur certaines valeurs (ex. un entier est toujours > 0)
- Gestion des accès concurrents
  - La base doit toujours rester cohérente
- Sécurité
  - Utilisateur
  - Droits : gestion fine par tables/types d'opérations

## SGBD et utilisateurs



## SGBD : produits

- Gratuits (Open Source)
  - MySQL, Postgress...
- Payants
  - Access (limité)
  - SQL Server
  - Oracle
  - Informix
  - SyBase
  - ...

## Base de données semi-structurées

- Gérer des données moins structurées que dans le relationnel
  - bases de documents XML
  - requêtes : Xquery
- Produits
  - SGBD dédiés
  - Extensions de SGBD relationnel

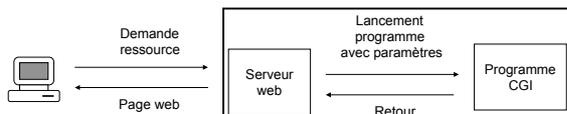
## Éléments des SIR : serveurs applicatifs

- Principe général
  - serveur dédié au *calcul applicatif*
  - rend des services à des clients
  - lien possible avec une BD
- Deux grandes catégories
  - programmes CGI sur serveurs web
  - programmes sur serveurs d'applications

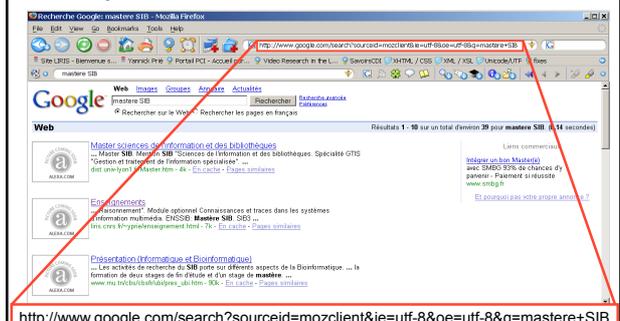


## Serveurs applicatifs : programmes CGI

- Common Gateway Interface
- Programme lancé par un serveur web
  - prend des paramètres en entrée
  - s'exécute
  - fabrique et renvoie une page web



## Exemple CGI



## Serveurs applicatifs : serveurs d'application

- Middleware
  - ensemble de programmes suffisamment génériques pour qu'on puisse les acheter tout faits et les en utiliser les fonctionnalités dans d'autres programmes
    - Ex. librairies informatiques
- Environnement d'exécution des applications côté serveur
  - Exécution + fonctionnalités diverses
    - gestion de la session utilisateur
    - gestion des montées en charge et reprises sur incident
    - accès aux sources de données
  - Ex : serveur faisant tourner
    - JSP : Java Server Pages → générer des pages web
    - Servlets : programmes Java côté serveur

## Éléments des SIR : interfaces utilisateurs

- A disposition de l'utilisateur
  - sur son poste
  - permet d'interagir avec le système d'information
- Plusieurs possibilités
  - interfaces *ad-hoc*
  - interfaces web
    - HTML
    - (X)HTML + javascript = AJAX
    - applets
    - autres (ActiveX, etc.)

## Interfaces *ad-hoc*

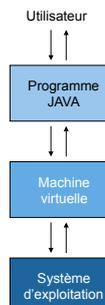
- Programme client installé sur le poste utilisateur
- Interagit avec un ou plusieurs serveurs
- Exemples
  - Endnote = client Z39-50
  - Client jeu multi-utilisateur
  - ...
- Complexité
  - aussi complexe que nécessaire
- Installation / mise à jour
  - sur chaque poste

## Interfaces HTML / AJAX

- Idée
  - la page dans le navigateur est l'interface
  - simplicité
- Exemple
  - Billets de train
- Complexité
  - Limites HTML + XML + javascript
  - En train d'évoluer
    - ex. interface gmail
- Installation
  - pas d'installation
    - afficheur d'interface = navigateur standard
    - toute la logique est côté serveur

## Excursus : JAVA

- JAVA
  - langage objet
  - programmes multi-plateformes (Mac, Unix, Windows, ...)
- Intérêt
  - programmer une seule fois pour toutes les machines du monde !
    - moins d'erreur
    - diminution des coûts
- Principe
  - exécution des programmes sur une machine virtuelle
  - une machine virtuelle par système d'exploitation
- Désavantage
  - rapidité moindre : exécuter un programme sur une machine virtuelle (un autre programme)



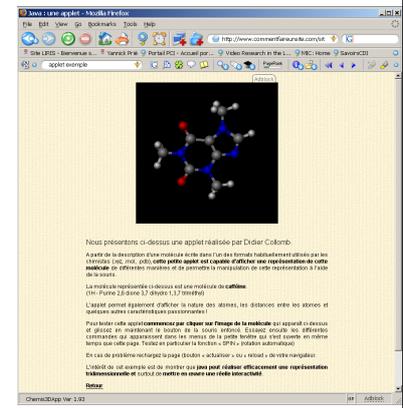
## Interfaces applets

- Applet
  - application JAVA qui s'exécute dans un navigateur
  - le navigateur délègue une partie de son affichage à l'applet
- Complexité
  - application graphique non limitée
- Installation
  - installation sur le serveur
- Limitations applet
  - communications réseaux
  - accès disques limités

## Autres interfaces

- Java web start
  - Lancer une vraie application Java depuis le web
- ActiveX
  - dans un navigateur
- Flash
- ...

## Exemple Applet



## Architecture des SI

- On a vu
  - données → SGBD
  - traitements sur les données → serveurs applicatifs
  - interface utilisateur → ad-hoc, web, applets...
- Architecture d'un SI
  - répartition des trois grands rôles
  - utilisation ou non du réseaux et des principes client / serveur

## Deux exemples de SI ...

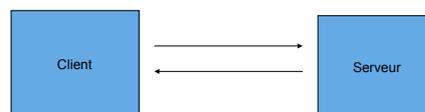
- Un SI très simple
  - gestion de DVD à la maison
    - une machine
    - une base de données des DVD
    - une interface graphique à la BD
- Un SI très complexe
  - gestion des impôts en France
    - des milliers de machines, plusieurs types de réseaux
    - plusieurs grosses base de données
    - de multiples interfaces graphiques
    - grosses contraintes de fiabilité

## ... et trois niveaux

- Données
  - Les vidéos, titres, auteurs, supports, prêts, commentaires, etc.
  - Les contribuables, les déclarations, les règles de calcul, les décrets, etc.
- Logique applicative / métier
  - Ajout d'un nouveau DVD, ajout d'un commentaire sur un film, prêt d'un DVD à un ami, statistiques sur les années, etc.
  - Ajout contribuable, déclaration, simulation de changement de loi, redressements fiscaux, etc.
- Présentation
  - Interface de commentaire, visualisation de la base, des courbes, des prêts, etc.
  - Interface pour le fonctionnaire des impôts, pour le statisticien, pour le contribuable, etc.

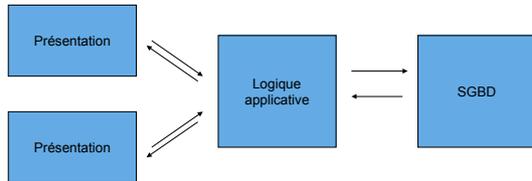
## Architecture 2 tiers

- La plus simple : deux applications
- Client = interface utilisateur
- Répartition éventuelle de la logique applicative entre client et serveur



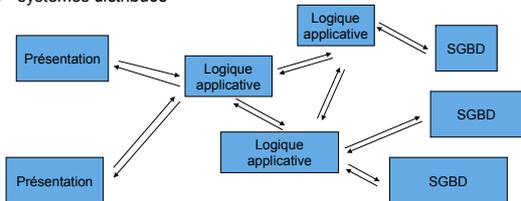
## Architecture 3 tiers

- Client-serveur généralisé
- Répartition des rôles
- Deux ou trois machines physiques



## Architecture n-tiers

- Généralisation
  - des données dans de multiples bases
  - une logique applicative partagée entre composants sur plusieurs serveurs
  - systèmes *distribués*



## Analyse des SIR : notion de rôles

- Un utilisateur
  - joue toujours un rôle par rapport à un système d'information quand il l'utilise
  - peut jouer plusieurs rôles / plusieurs utilisations
- Exemple site web de publication
  - *administrateur*
    - vérification fonctionnement, amélioration, etc.
  - *utilisateur identifié responsable*
    - création utilisateurs, validation articles
  - *utilisateur identifié normal*
    - utilisation du SI (lecture / ajout articles)
  - *visiteur*
    - utilisation limitée du SI (lecture, commentaires)

## Analyse des SIR : fonctionnalités générales

- Identification et gestion d'utilisateurs, groupes, etc.
- Administration (plusieurs niveaux)
- Saisie et contrôle de saisie
- Interrogation
- Statistiques
- Fonctionnalités liées à la collaboration
- ...

## Fonctionnalités des SID2

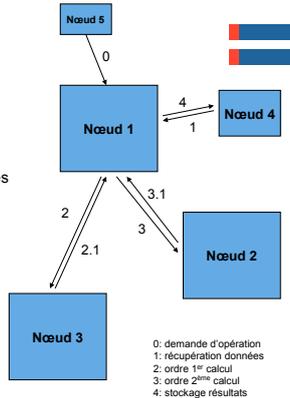
- Données = documents, méta-données (notices)
- Fonctionnalités
  - Identification et gestion d'utilisateurs, groupes, etc.
    - utilisateurs de documents, autorisations
  - Administration (plusieurs niveaux)
    - ajout documents, choix des méta-données
  - Saisie et contrôle de saisie
    - orientée par les données
  - Interrogation
    - dans les documents et notices
  - Statistiques
    - sur la recherche, l'utilisation du système
  - Fonctionnalités liées à la collaboration
    - possibilité de corriger un résumé
  - Autres fonctionnalités
    - prêt
    - ...

## Analyse des SIR : types de données

- Différents nœuds du système
  - différentes données (types de données)
  - gérées différemment (format de stockage, accès)
  - générées différemment (entrées utilisateur, calcul, ...)
- Exemple
  - notices bibliographiques (titre, auteur)
    - gestion SGBD dédié
  - index de recherche / mots-clé
    - gestion SGBD dédié
  - données utilisateur (nom, mdp)
    - gestion SGBD général utilisateurs intranet
  - préférences utilisateur (langue, taille fenêtre)
    - gestion fichier machine utilisateur
  - ...

## Analyse des SIR : types de traitements

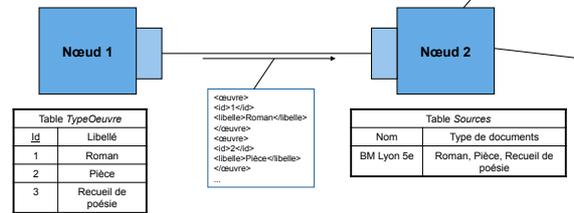
- Différents nœuds du système
  - différents traitements
  - plus ou moins complexes
  - plus ou moins vitaux
  - réalisés par des programmes variés
    - scripts interprétés
    - programmes compilés binaires
    - ...
- Exemple
  - validation d'accès
  - recherche d'information
  - récupération d'information
  - transformation de données
  - accès aux documents
  - ...



0: demande d'opération  
1: récupération données  
2: ordre 1<sup>er</sup> calcul  
3: ordre 2<sup>ème</sup> calcul  
4: stockage résultats

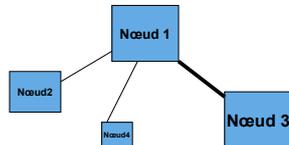
## Analyse des SIR : circulation des données

- Entre deux nœuds du système
- Implique souvent un transcodage



## Analyse des SIR : types de machines/systèmes

- A chaque nœud du SI, des ordinateurs
  - système d'exploitation
  - puissance de calcul
  - possibilités d'exécution de programmes : scripts, binaires, etc.
  - possibilité de stockage, de sauvegarde, de récupération
  - traitement parallèle des connexions
  - ...
- Entre les nœuds, un réseau
  - technologie
  - débit
  - fiabilité
  - ...



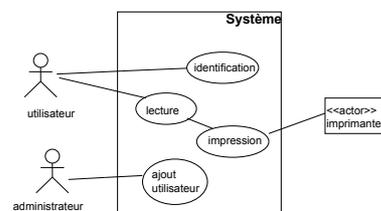
## Bilan

- Composants d'un SID
  - SGBD(s)
  - Serveur(s) applicatif(s)
  - Interfaces(s) utilisateurs
- Architecture
  - N-tiers (réseau, client/serveur, multiples machines et systèmes)
- Rôles et fonctionnalités
  - standards
  - spécifiques
- Flux d'information
  - circulation des données
  - traitements variés

## Quelques diagrammes UML

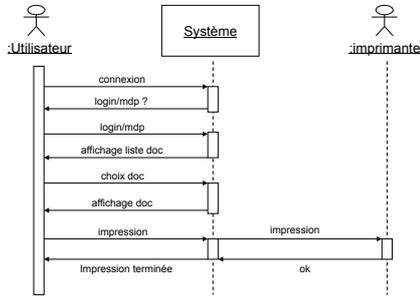
- Unified Modelling Language
- Language graphique normalisé de description de systèmes
  - de l'expression des besoins utilisateur
  - à la description du code et des exécutables
- UML 2 : 13 types de diagrammes
- Intéressant pour nous
  - Diagrammes de cas utilisation → rôles
  - Diagrammes de séquence → entrée/sortie des informations
  - Diagramme de déploiement → localisation des machines et des services

## Diagramme de cas d'utilisation

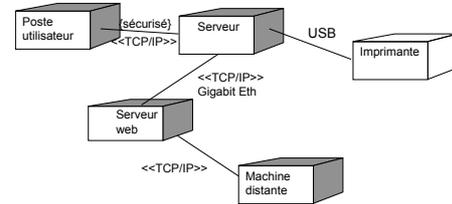


Rôles à l'extérieur du système  
Cas d'utilisation = grandes classes d'interaction

## Diagrammes de séquence



## Diagramme de déploiement



## Discussion

- Problèmes liés à la répartition
  - Réseau
    - engorgement, coupure, etc.
  - Machines
    - problème, maintenance, etc.
- Evolution des SI/SID
  - Systèmes collaboratifs
    - synchrones, asynchrones
  - Systèmes pervasifs
    - Mobilité
  - SAAS
    - Software As A Service

## Remerciements

- <http://www.w3schools.com/sql>

## Exemples de systèmes d'information documentaires distribués

SIB  
WEB  
Bibliothèque numérique

## Objectif : comprendre des systèmes d'information

- En termes
  - de fonctionnalités et de rôles
  - d'architecture
  - des traitements
  - de localisation et circulation des données
- On va s'intéresser
  - au web
  - à un SI de bibliothèque
  - à une bibliothèque numérique

## Exemple 1 système d'information de bibliothèque

- Bibliothèque municipale
- Personnel : un dizaine de personnes
- Connexion Internet
- Catalogue : plusieurs dizaines de milliers de documents
- Postes de consultation

## SIB : fonctionnalités / rôles

- Fonctionnalités
  - Administration
  - Gestion
    - Achats, facturation
    - Entrée des notices
  - Consultation
    - Locale, distante, identifiée ou non
  - Prêt
    - Poste bibliothèque
- Rôles
  - administrateur
  - employé
  - membre identifié
  - visiteur

## SIB : données

- Notices
- Description utilisateurs
- Prêts
- Comptes
- ...

## SIB : traitements

- Traitements basiques de modification des données
- Recherche de fiches
- Gestion des retards
- Statistiques

## SIB interfaces utilisateurs

- Interface web pour accès web
- Interface spécialisée pour autres accès

## SIB : architecture

- Un serveur central avec une base de données
  - notices, prêts, etc.
- Un serveur web pour l'accès distant
  - site web de la bibliothèque + accès catalogue
- Des terminaux visiteur
  - consultation notices
- Des terminaux guichets
  - notices, prêts, ajouts membres...
- Des terminaux dans les bureaux
  - administration générale, finances
- Terminal administrateur
  - éventuellement déporté

## SIB : catalogue en Z39-50

- Utilisation du serveur web pour répondre aussi aux requêtes Z39-50 (autre port)
- Adaptation des requêtes Z39-50 au SGBD des notices
  - Questions
  - Format réponses

## Exemple 2 Le web comme SID2

- Web
  - gigantesque base documentaire
  - outils pour chercher dans cette base
    - moteurs de recherche

## Web : fonctionnalités et rôles

- Administrateurs
  - des moteurs de recherche
  - des sites web
- Utilisateurs
  - navigation
  - utilisation de services (moteur de recherche)
- Fonctionnalités
  - recherche sur moteur
  - recherche annuaire
  - navigation / recherche
  - visualisation

## Web : données

- Documents de tous types
  - HTML
  - PDF
  - PPT
  - Images
  - ...
- Descriptions de documents
  - Méta-données
    - annuaires
    - moteurs de recherche
  - Pages et liens

## Web : traitements

- Recherche d'information
  - mots-clés → sites réponse classés
- Construction des index
  - Robots

## Web : interfaces utilisateurs

- Interfaces web
- Applications sur la machine utilisateur
  - Application seule : word, acrobat, etc.
  - Plug-in : acrobat, etc.
- Interfaces pour gestion annuaires

## Web : architecture

- Serveurs web
  - fichiers disponibles à certaines URLs
- Bases de données
  - moteurs de recherche
  - réponse aux requêtes
- Robots
- Machines utilisateurs
  - système d'exploitation + navigateur + applications
- Réseau internet

## Exemple 3 bibliothèque numérique

- Gestion de documents électroniques
- Réfléchissons ensemble
  - Rôles par rapport au système
  - Fonctionnalités
  - Architecture
  - ...