

**INSA de LYON – Doc'INSA**

# **Serveur de thèses en texte intégral**

Dossier d'initialisation

**Rédaction** : Marc-Etienne HUNEAU

**Version** : 2.16

**Révision** : mardi 9 décembre 1997

## **Objet du projet - Contexte**

Le projet concerne la mise en place d'un serveur de thèses en texte intégral à la bibliothèque Doc'INSA, dépositaire des thèses produites à l'INSA de LYON.

### **Contexte**

#### **Les thèses de Doc'INSA**

Doc'INSA, dépositaire officiel de toutes les thèses produites au sein des laboratoires de l'INSA de LYON, reçoit chaque année environ 150 thèses. Ces thèses, conservées en deux exemplaires, peuvent être consultées à la bibliothèque. Dans le cadre du prêt entre bibliothèques, des reproductions totales ou partielles de ces thèses (photocopie) sont envoyées aux bibliothèques demandeuses.

Actuellement, plus de 2000 thèses sont conservées à Doc'INSA.

#### **Intérêt de la publication électronique des thèses**

Les thèses sont généralement des documents à durée de vie relativement courte. A de rares exceptions près, les thèses ont une durée de vie « utile » de quelques années. Il importe donc que ces thèses soient mises à disposition des lecteurs éventuels le plus rapidement possible (dès leur réception).

Par ailleurs, mettre en place un accès informatique à ces thèses (via internet) permettra probablement de réduire significativement le volume de photocopies à réaliser, les documents électroniques étant imprimables par l'utilisateur.

### **Positionnement du projet**

#### **Base documentaire de Doc'INSA**

Par manque de place, Doc'INSA ne propose pas un accès direct aux documents. Il est donc impératif d'utiliser la base de données documentaire afin d'identifier un ouvrage, avant de le consulter.

Ceci présente un avantage : l'accès aux thèses « électroniques » ne nécessitera pas une démarche fondamentalement différente de la procédure actuelle (de la part du lecteur). Il serait alors judicieux que la base documentaire soit interfacée avec le serveur de thèses (et que des liens vers les thèses trouvées soient proposés à l'utilisateur). L'interface de recherche est actuellement accessible via une interface HTML, depuis un navigateur internet (netscape, internet explorer...). Les thèses pourraient être consultées directement via ces navigateurs.

## **Choix déjà effectués**

Un certain nombre de choix (notamment techniques) ont déjà été faits. Le matériel a été acheté (un serveur et un poste de travail pour l'élaboration des documents électroniques), et le format PDF (Adobe Acrobat) a été adopté pour la publication. Ce format permet en effet de proposer des documents électroniques complexes (images, équations...) d'un poids raisonnable, et est adapté au *Web* (hyperliens, chargement incrémental).

Ceci n'exclut pas une réflexion quand à la pérennité des dits choix, et l'évolution des techniques doit être prise en compte. Des possibilités de rétro-conversion (par exemple en HTML ou dans le futur XML) doivent être conservées, et la solution proposée ne doit pas être figée.

## **Phases à réaliser**

L'objectif fixé pour ce projet est de disposer d'une solution fin 1998. Le service (accès télématique aux thèses) et la chaîne de production des documents électroniques (conversion au format PDF, inscription dans la base) devront être opérationnels.

Les phases à réaliser vont donc de l'étude préalable à la réalisation. Si la réalisation n'est pas terminée dans le cadre de ce PFE, elle pourra être confiée à un stagiaire (par ex. : 4IF).

# Résultats attendus

## Livrables (documents et composants logiciels) demandés

### Documents

Le projet est découpé en trois phases principales :

- Etude préalable,
- Etude détaillée,
- Réalisation.

Chacune de ces phases donnera lieu à un rapport.

En sus, chaque réunion donnera lieu à un compte-rendu, tout incident sera recensé et recevra une fiche d'incident (afin de constituer une « base de connaissances » utilisable par d'autres équipes pour un projet similaire).

Tous les documents concernant ce projet (*à l'exception des documents à usage interne - mémos, comptes-rendus*) seront mis en forme afin de constituer un ensemble cohérent pouvant être mis à disposition du monde extérieur (via internet, par exemple).

Cette mise à disposition sera soumise à l'accord de Doc'INSA.

### Composants logiciels

Les éléments à réaliser sont les suivants :

Une chaîne de production, souple mais automatisée tant que faire se peut, afin de préparer les documents électroniques,

Le service à proprement parler (mise en place du serveur et des outils de recherche).

La chaîne de production sera un ensemble comprenant des éléments logiciels, mais aussi des documents type, des règles de fonctionnement, des procédures à suivre. La publication de documents électroniques soulève des questions relatives au droit d'auteur, et il est à prévoir que des règles soient à soumettre à l'approbation des auteurs.

Le phasage du projet précise qu'une maquette pourra être réalisée. Dans la mesure du possible, les composants logiciels définitifs seront également développés.

# Méthodes – modes opératoires – phasage

## Méthodes utilisées

La méthode de conception orientée objet OMT sera probablement utilisée lors de l'étude de ce projet. Ceci sera précisé ultérieurement.

## Phasage

Un planning précis sera établi au début de chaque phase principale. En voici la liste :

- 1) étude préalable (du début du PFE aux vacances de Noël),
- 2) étude détaillée (de Noël aux vacances de Pâques),
- 3) réalisation (des vacances de Pâques à juin),
- 4) validation.

# Identification des activités et tâches

## Planification du projet

Un planning est fourni en annexe du présent document.

Il est à noter que les horaires destinés à la réalisation du PFE (dans l'emploi du temps de 5IF) ne constitueront en aucun cas une limite supérieure à cette réalisation. L'étudiant effectuera conjointement à l'étude du PFE une gestion de projet comprenant des rapports d'activité.

*Des périodes de vacances de l'étudiant pourront être consacrées à l'avancement de ce projet.*

L'étudiant pourra, dans le cadre des heures consacrées à son PFE, travailler hors de l'enceinte de Doc'INSA, *notamment* en-dehors des horaires normalement consacrés au PFE.

Durant tout le déroulement du projet, et particulièrement lors de la définition précise des objectifs à atteindre et de l'état de l'art en la matière, des périodes de regroupement fréquentes seront prévues.

# Composition de l'équipe

Monique JOLY, directrice de doc'INSA, est commanditaire du projet.

Jean-Michel MERMET, responsable de l'informatique de doc'INSA, encadre le projet pour doc'INSA.

Jean-Marie PINON, enseignant, est le tuteur du PFE pour le département informatique.

Marc-Etienne HUNEAU, étudiant en cinquième année d'informatique, est en charge du projet.

# Modalités de validation et de recette

Chaque phase du projet donnant lieu à la rédaction d'un rapport, ledit rapport sera visé par les membres de l'équipe, et une réunion permettra d'en discuter le contenu.

Deux soutenances interviendront (autour des vacances d'hiver, puis en fin de PFE), et permettront également de contrôler les orientations choisies.

Des réunions pourront être sollicitées par tout membre de l'équipe.

## Annexes

### Description succincte des logiciels à livrer

Les livrables logiciels attendus sont de deux nature.

Le *BackOffice* comprendra les outils de préparation des documents électroniques (chaîne de production des thèses numériques), ainsi que les outils de leur insertion au sein de la base documentaire.

Le *FrontOffice* sera constitué par l'interface de recherche et de consultation des thèses numérisées. Celui-ci s'appuiera sur un serveur Web classique (NCSA), ainsi que sur le logiciel Doris-Web, accès *cgi* (via une requête de type web) à la base de données de doc'INSA.

Les fonctions principales de ces logiciels seront les suivantes :

- Chaîne de préparation des documents :

Cette chaîne doit permettre l'élaboration du document final (consultable via internet), à partir du document original.

Ce document original peut être de différentes natures : Document Word (monolithique ou par chapitres), fichier *PostScript* ou *Tex / Latex*. De plus, des annexes peuvent être fournies dans des formats différents (tableau Excel, fichier issu d'un programme de dessin...). Des règles pourront être fixées afin de limiter la diversité des formats source, mais une souplesse suffisante pour ne pas compliquer le travail du thésard est indispensable. Enfin, des documents photographiques ou imprimés pourront être numérisés via un scanner ; d'éventuels fichiers multimédia (vidéo, son) pouvant également être intégrés à la thèse.

Le document final sera (dans un premier temps en tout cas) au format PDF d'Adobe. La forme exacte de ce document (fichier unique ou découpage en chapitres, intégration des données multimédia ou liens hypertexte vers ces formats) sera définie durant la phase d'étude détaillée.



- Interface de consultation :

Les documents résultant de la numérisation (ou tout au moins du codage en PDF) seront proposés à la consultation via internet. Une interface d'accès à une telle base de documents (base qui sera amenée à croître au rythme de production des thèses à l'INSA – soit 150/an) doit permettre un accès efficace et simple à l'information.

De tels systèmes existent, et apportent des réponses plus ou moins satisfaisantes aux besoins en matière de recherche, de consultation ou de contrôle des accès aux documents.

Les fonctionnalités incontournables seront : la recherche (suivant différents critères) des thèses dans la base documentaire, la consultation de ces thèses.

Parmi les fonctionnalités dont l'adjonction au serveur est envisageable, on peut citer le contrôle de l'accès aux thèses (par catégorie de thèse et/ou d'utilisateur), la mise à disposition du lecteur des outils nécessaires, l'élaboration de quelques pages d'aide ou de mise en garde au lecteur, un formulaire pour le dépôt des remarques éventuelles...

## **Descriptif des tâches**

*A préciser ultérieurement.*

## **Planning**

Joint au présent document.

