

LOGICIEL DE GESTION DE DOCUMENTS PDF : PROJET INFO 1

L. POINSOT

Contact client :

Laurent POINSOT (laurent.poinsot@lipn.univ-paris13.fr)

Résumé : Ce document est le cahier des charges du projet INFO 1. Il s'agit d'une part de développer un logiciel de gestion de documents pdf et d'autre part de rédiger un certain nombre de documents concernant ce développement. Ce document est constitué de trois parties. Dans la première partie sont listées les fonctionnalités devant obligatoirement être implantées par votre logiciel. La seconde partie décrit le travail attendu et l'organisation des équipes projet. Enfin un calendrier prévisionnel constitue la troisième et dernière partie de ce document dans laquelle sont données les dates de livraison des documents attendus.

L'objectif de ce projet est de réaliser un **logiciel de gestion de documents pdf**.

Ci-dessous sont listées les fonctionnalités **obligatoirement** supportées par le logiciel. Votre système devra fonctionner comme cela est indiqué ci-dessous. La note finale sera fonction de la qualité de votre logiciel (en particulier le fait que les caractéristiques de base soient correctement implantées), des documents livrés (voir la suite du présent sujet) et de la présentation finale de votre production.

I. LE LOGICIEL DE GESTION DE DOCUMENTS PDF : FONCTIONNALITÉS *obligatoires*

- 1 Le logiciel de gestion de documents PDF est un logiciel qui doit fonctionner sous linux dans les salles machine de l'École. Le cœur du programme (le code ne faisant pas partie de l'IHM) est codé en langage C ou C++ (sauf dérogation, après accord explicite de la part du client ¹).

1. De façon générale, si vous souhaitez ne pas suivre l'une des consignes de base, il vous faut négocier cela auprès du client en prenant soin d'argumenter vos choix techniques afin d'obtenir leur éventuel accord.

- 2 Toutes les fonctionnalités spécifiées dans la suite doivent être effectuées *via* une interface homme-machine IHM (dont vous devrez décrire l'utilisation ; se référer à la suite du présent document).
- 3 Le logiciel doit permettre à un utilisateur d'ajouter ou de retirer des documents de sa bibliothèque. La gestion de répertoires et de sous-répertoires doit également être permise. La possibilité doit être offerte à l'utilisateur de créer ou de détruire des répertoires, bref de gérer l'arborescence de sa base de documents pdf.
- 4 À chaque document de la base doit être associée une fiche électronique contenant des informations relatives au document : titre, auteurs, date de publication, un résumé, par exemple. L'IHM doit permettre l'édition des fiches (par exemple pour rajouter des informations).
- 5 La recherche de documents dans la bibliothèque *par mots-clefs* doit être implantée : les documents PDF contenant certains mots-clefs (précisés dans la recherche) seront ainsi sélectionnés. L'utilisateur doit pouvoir consulter la liste des résultats correspondants à sa recherche, mais aussi créer (à l'aide d'une simple option à déterminer) un répertoire contenant les fichiers résultats de cette recherche. **Attention, les documents pourront éventuellement être écrits en langue anglaise.**
- 6 En plus de la recherche de documents par mot-clefs, une recherche de fichiers par poids (inférieur, supérieur, égal), par date, par nombre de pages, *etc.* devra être implantée.
- 7 Le logiciel doit implanter la recherche de mots-clefs dans un document PDF. **Attention, c'est différent de la recherche par mot-clef!** Cette fonctionnalité doit avoir comme résultat la liste des mots-clefs **pertinents** contenus dans le document (on doit donc éviter les articles (*le, la, les, ...*), les prépositions (*à, dans, de, sous, sur, ...*) et autres conjonctions courantes (*mais, ou, et, donc, or, ni, car, ...*) que ce soit en français ou en anglais). **Il vous faudra décrire avec le plus grand soin l'algorithme que vous utiliserez.** L'utilisateur doit pouvoir récupérer la liste des mots-clefs (avec le nombre de leurs occurrences, ainsi que les numéros des pages où ils apparaissent) à l'aide d'un copiage-collage.

8 Votre logiciel doit posséder une architecture client-serveur. Un logiciel serveur contiendra la base de données des documents pdf et la partie client accédera au serveur afin de réaliser les opérations sur la base de données. Plusieurs clients doivent pouvoir se connecter sur le serveur.

Option : Le logiciel pourra fonctionner sous un système Windows. Il s'agit là d'une **option** et non d'une fonctionnalité de base. Vous pouvez l'implanter dans le seul cas où les fonctionnalités précédentes ont été correctement réalisées. Évidemment la réalisation de cette option sera valorisée comme un bonus dans la note finale.

II. ORGANISATION ET TRAVAIL ATTENDU.

- **Organisation du travail** : Tous les projets s'effectuent par **groupes de quatre** (le nombre exact de membres sera validé par le client). Le sujet étant le même pour tous, il est fortement conseillé à chaque groupe de faire preuve d'autonomie, d'esprit d'initiative et d'originalité. La notation valorisera les originalités apportées au logiciel (en plus des fonctionnalités de base décrites précédemment). **Attention cependant à ce que les fonctionnalités de base, telles que demandées par les clients dans le présent document, soient correctement implantées.** Chaque groupe doit être parfaitement cloisonné de façon à ne souffrir d'aucun transfert de compétences désiré et/ou subi. Le « copiage-collage », l'*espionnage industriel*, la collaboration ou toute autre preuve explicite - ou jugée comme telle - de diffusion ou fuite d'information au profit d'un autre groupe sera sévèrement punie au niveau de la note finale que ce soit pour la source comme pour le bénéficiaire de la fuite ! Au contraire, imagination et indépendance sont encouragées et valorisées dans la note finale.
- **Planification du projet** : Ce projet se déroule en quatre étapes successives. À l'issue de chaque étape (exception faite de la phase de codage/soutenance) doit être rendu un document (par mél en version électronique au format PDF) rendant compte de façon synthétique (ce ne sont pas des romans !) du travail effectué lors de cette étape (voir " Calendrier prévisionnel " pour de plus amples informations). Au terme de l'ultime étape - la livraison finale - aura lieu une présentation orale du travail effectué et le bon fonctionnement du logiciel sera démontré à l'aide d'une batterie de tests. Les groupes sont contraints de suivre le calendrier prévisionnel du

déroulement du projet (présenté dans la dernière partie de ce document). Ils sont avertis que la qualité des documents transmis ainsi que la ponctualité de leur livraison prendront une part **prépondérante** dans la note finale (en d'autres termes, les groupes ne doivent pas se contenter de coder au risque d'être déçus de la note finalement attribuée). **Attention** : Les livraisons doivent se faire, **au plus tard à 23h59**, le jour indiqué dans le calendrier prévisionnel. Tout retard (ne serait-ce que d'une minute) sera sanctionné par une baisse de la note.

- Tout changement important effectué par un groupe par rapport à ce qu'il avait prévu est soumis à l'approbation du client. Par exemple, si un groupe juge impossible l'implantation d'une fonctionnalité originale ou qu'il doit abandonner la conception d'une fonctionnalité supplémentaire, alors il doit prévenir le client. Il attendra la validation, ou non, de ce changement par le client.
- **Toute correspondance ou demande d'entretien se fait exclusivement par mél ou lors d'un des rendez-vous d'avancement du projet. Autrement dit, vous n'êtes pas autorisés à taper directement à la porte des bureaux du client sans avoir effectué de demande d'entretien préalable.**
- Le client s'engage à répondre aux documents livrés après un délai maximum d'une semaine suivant leur livraison afin de ne pas ralentir l'avancement du projet.

III. CALENDRIER PRÉVISIONNEL (LES QUATRE ÉTAPES ET LES DOCUMENTS À LIVRER).

Remarques :

- Sur votre emploi du temps sont spécifiées les dates réservées au projet. Pour la plupart d'entre elles, le client ne sera pas présent (sauf sur rendez-vous après sollicitation d'un groupe de projet), par contre il faut que vous les utilisiez afin de faire avancer votre projet. À certaines dates (données ci-dessous), le client sera présent afin de répondre aux questions éventuelles avant ou après livraison.
- Vous pourrez également rencontrer le client aux dates indiquées ci-dessous comme " point d'avancement intermédiaire " afin de statuer sur l'état d'avancement de vos travaux et d'éviter la propagation d'éventuelles erreurs ou autres incompréhensions. **Prière de vous présenter au début du créneau horaire !**

- (1) **Description générale du logiciel.** Il s'agit de la première étape. Les groupes doivent présenter (en français, et non sous la forme de code ou pseudo-code) les différentes fonctionnalités du logiciel en précisant la manière dont ils interprètent les descriptions précédentes (que signifie " pertinent " pour un

mot-clef ? Les recherches peuvent-elles être restreintes à un sous-répertoire ? Sous quelle forme sont renvoyés les résultats d'une recherche ? *etc.*). Les groupes évoquent également les éventuelles originalités qu'ils vont y ajouter. Cette partie n'est nullement technique mais précise et expose les différents services que proposent votre logiciel.

Point d'avancement intermédiaire : 16/03/2012.

Document à livrer : le document de description générale de la bibliothèque. **Date de livraison**² : **30/03/2012** (23h59 dernier délai).

- (2) **Manuel d'utilisation et État de l'art sur la gestion en langage C ou C++ des documents PDF.** Dans le premier de ces documents vous expliquez à un usager quelconque la manière d'employer votre logiciel (dans quelle fenêtre de l'IHM faut-il écrire les mots-clefs pour une recherche ? Sur quel bouton cliquer pour lancer la recherche ? Quel est le langage de requête utilisé pour écrire les recherches ? *etc.*). Rappelez-vous que ce document est à destination d'une personne quelconque, qui ne connaît pas votre travail, et que sa simple lecture doit lui permettre d'installer, de configurer et d'utiliser le logiciel. Dans le second document, vous détaillez les moyens de gestion (lecture, recherche, comment " parser " un document PDF, *etc.*) *via* le langage C ou C++ (en utilisant des bibliothèques ou des logiciels) des documents PDF issus d'une recherche sur internet. Veuillez expliciter ceux parmi les logiciels ou les bibliothèques qui sont libres et gratuits d'utilisation.

Point d'avancement intermédiaire : 6/04/2012.

Documents à livrer : le manuel d'utilisation et l'état de l'art. **Date de livraison** : **4/05/2012** (23h59 dernier délai).

- (3) **Algorithme de recherche de mots-clefs.** Dans cette étape, vous concevez de façon très détaillée l'algorithme de recherche des mots-clefs pertinents dans un document PDF (lequel algorithme sera implanté dans votre logiciel).

Point d'avancement intermédiaire : 18/05/2012.

Document à livrer : le document descriptif de l'algorithme. **Date de livraison** : **1/06/2012** (23h59 dernier délai).

Attention à commencer la phase de codage avant la fin

2. Toutes les dates de livraison indiquées sont à respecter sous peine de sanction au niveau de la note finale.

de cette présente phase au risque de ne pas disposer d'assez de temps !

- (4) **Codage/soutenance.** La soutenance est prévue pour fin 2012 et vous sera précisé ultérieurement. La soutenance se fera sous la forme d'une présentation orale du fonctionnement du logiciel en mettant l'accent sur les originalités apportées suivie de réponses aux questions des clients (ces questions concerneront les différents documents livrés, l'utilisation et le test du logiciel, la connaissance du code).

Aucun document n'est à livrer à cette étape.

Date de soutenance : fin juin 2012, à préciser.

LE CORRECTEUR (SUSNOMMÉ LE CLIENT) SE RÉSERVE LA LIBERTÉ DE MODIFIER À SA GUISE - MAIS AUSSI APRÈS CONCERTATION AVEC LES ÉTUDIANTS - LE CALENDRIER PRÉVISIONNEL. BIEN ÉVIDEMMENT TOUTE MODIFICATION SERA ANNONCÉE AUX GROUPES AVEC UN PRÉAVIS DÉCENT *ad hoc*.