

Programmation, GL, preuve — Projet

Stéphane Lopes

16 février 2018

Ce projet consiste à créer une application pour la gestion des documents personnels. Ce projet est à réaliser par groupe de trois ou quatre étudiants.

Si un élément n'est pas précisé dans l'énoncé, vous êtes libre de choisir la mise en œuvre qui vous convient. De même, vous êtes libres d'ajouter d'autres fonctionnalités (support d'autres types de fichiers, ...).

1 Description du projet

L'idée est de proposer un système capable de collecter et d'organiser un ensemble de documents de différents types et sous différents formats.

Un premier cas d'usage peut être la gestion des factures. Lorsque l'utilisateur récupère une facture (EDF par exemple) au format PDF, il la dépose dans l'espace *INBOX* de l'application. Cette dernière récupère le document, lui applique de la reconnaissance de caractères (OCR), extrait des métadonnées puis archive le document en ayant complété les métadonnées du PDF et en l'annotant pour des recherches ultérieures.

Un second exemple concerne la gestion d'une médiathèque (musiques, vidéo, ebooks, ...). Une vidéo est déposée dans l'espace *INBOX*. L'application analyse le titre et les métadonnées, fait une recherche sur IMDb, puis complète les métadonnées, archive et annote la vidéo et la rend disponible par l'intermédiaire d'un *media center*.

2 Description de l'application

L'application devra se présenter sous la forme d'un *shell* (donc en mode texte).

Elle proposera les fonctionnalités suivantes :

- intégration d'un document ou d'un répertoire (récursif)
- structuration en hiérarchies virtuelles (un document peut se trouver dans plusieurs hiérarchies)
- annotation des documents par des balises
- classement et nommage automatique de documents
- navigation dans les hiérarchies ou les annotations
- différents modes de recherche (nom du document, annotation, hiérarchie, *full text* pour certains types de documents)
- prise en compte des fichiers textes (recherche *full text*)
- prise en compte des fichiers PDF (récupération et mise à jour des métadonnées, OCR et recherche *full text*)

2.1 Contraintes techniques

- Le projet est à réaliser en Java (ou Kotlin) par groupe de 3 ou 4 étudiants.
- Il sera géré dans un dépôt *bitbucket* **privé**.
- La construction du projet devra s'appuyer sur l'outil *maven* (ou *gradle*).
- La plupart des classes du projet devront être associées à des tests unitaires.
- Les données de l'application devront être conservées dans un SGBD.
- Le projet devra comporter une documentation sous la forme d'un fichier *README.md* et de commentaires *javadoc* dans le code source.
- La documentation devra décrire l'usage de l'application (*manuel utilisateur*) ainsi que la conception (*manuel technique*).
- En particulier, l'accent sera mis sur les *patterns de conception* utilisés dans le projet.

3 Références

- Open Semantic Search
- Paperwork
- JAnsi pour la gestion des couleurs dans un terminal
- Java OCR and Barcode Recognition
- Apache Lucene pour la recherche *full text*