

Cahier des charges

Isabelle Richard, Elias Abou Haydar, Mikael Ahl et Jeremie Rouach

March 14, 2013

0.1 Introduction

0.1.1 Objectif du document de specification

Ce document de spécification produit des informations spécifiques et nécessaires pour définir efficacement les fonctionnalités, l'architecture et la conception du système afin de donner la direction à l'équipe de développement sur l'architecture du système à développer. Le document de spécification du produit est créé pendant la phase de planification du projet. Son public visé est le chef de projet, l'équipe de projet et l'équipe de développement et en partie le client. Les spécifications techniques et fonctionnelles de ce document sont réservées au chef de projet, l'équipe de projet et l'équipe de développement.

0.1.2 Portée du produit

Le logiciel iFind permet de rechercher un fichier dans un ensemble de répertoires ciblés du système. Cette recherche peut se faire soit en indiquant le nom du fichier, soit en donnant une liste de mots contenus dans ce fichier.

0.1.3 Définitions, acronymes et abréviations

UI Acronyme de “user interface” (interface utilisateur).

Corpus Un corpus est un ensemble de documents, artistiques ou non (textes, images, vidéos, etc.) , regroupés dans une optique précise.

Fichier Contenant virtuel auquel est assigné un nom unique, permettant de classer et de réunir en une même entité une séquence de données. Le fichier est stocké dans un système de fichier et les données qu'il contient sont généralement structurées en suivant un même format.

Document En informatique, le mot *document* est généralement synonyme de fichier. On parle ici de document électronique. Un document électronique est un contenu de médias électroniques (autres que les programmes d'ordinateur ou des fichiers système) qui sont destinés à être utilisés soit dans une forme électronique ou comme sortie imprimée.

Indexation L'indexation permet de regrouper en un seul endroit toutes les données souhaitées. On crée des indexes, ce qui permet d'y accéder plus rapidement.

Index Un index est, en toute généralité, une liste de descripteurs à chacun desquels est associée une liste des documents et/ou parties de documents auxquels ce descripteur renvoie. Lors de la recherche d'information d'un usager, le système accèdera à l'index pour établir une liste de réponses.

Moteur de recherche Un moteur de recherche est un code logiciel qui est conçu pour rechercher des informations ou retrouver des ressources associées à des mots quelconques. Ici, on parle de moteur de recherche de type “ Desktop ” , car son champ d'action est limité à l'ordinateur sur lequel l'application est installée.

Requête En informatique, une requête est une demande de traitement. Dans notre cas, le terme est employé dans le contexte des bases de données, une requête correspondant à l'interrogation d'une base pour en récupérer une certaine partie des données.

Base de données Une base de données, usuellement abrégée en BD ou BDD, est un ensemble structuré et organisé, permettant le stockage de grandes quantités de d'informations afin d'en faciliter l'exploration (ajout, mise à jour, recherche de données). Autrement dit, il s'agit d'un conteneur informatique permettant de stocker dans un même endroit l'intégralité des informations en rapport avec une activité. Une base de données permet de stocker un ensemble d'informations de plusieurs natures ainsi que les liens qu'il existe entre les différentes natures.

Expression régulière Les expressions régulières (aussi appelées expressions rationnelles) sont de chaines de caractères permettant de décrire un ensemble de variables par l'utilisation d'une syntaxe précise qui se retrouvent dans de nombreux langages et outils.

Démon Un démon ou daemon désigne un type de programme informatique, un processus ou un ensemble de processus qui s'exécute en arrière-plan plutôt que sous le contrôle direct d'un utilisateur.

0.1.4 References

IEEE Std 830-1998 Recommended Practice for Software Requirements Specifications

0.2 General Overview and Design Guidelines

Cette section décrit les principes et les stratégies qui seront utilisées comme des lignes directrices lors de la conception et de la mise en œuvre du système.

0.2.1 Perspective du produit

iFind utilise une base de données construite à l'aide d'un moteur d'indexation et mise à jour dès qu'un fichier est modifié. La requête est envoyée au moteur de recherche via une interface graphique (GUI) (voir Figure 1). On utilise une interface graphique, pour permettre à l'utilisateur de rechercher ce dont il a besoin. La GUI est constituée d'un champ de saisie, d'un bouton "Chercher" ainsi que d'un explorateur qui permettra d'ouvrir les fichiers trouvés (extension). De base, l'explorateur contiendra un tableau dans lequel on affichera les résultats. Le champ de saisie reçoit une requête sous forme d'expression régulière ou des mots simples.

Dans cet exemple, la recherche envoie tous les fichiers contenant les mots "toto" ou "abc".

0.2.2 Fonctionnalités du produit

Lors de la première utilisation, iFind lance un démon ayant pour tâche d'indexer un corpus ciblé. Ce démon va construire une base de données à l'aide de ces in-

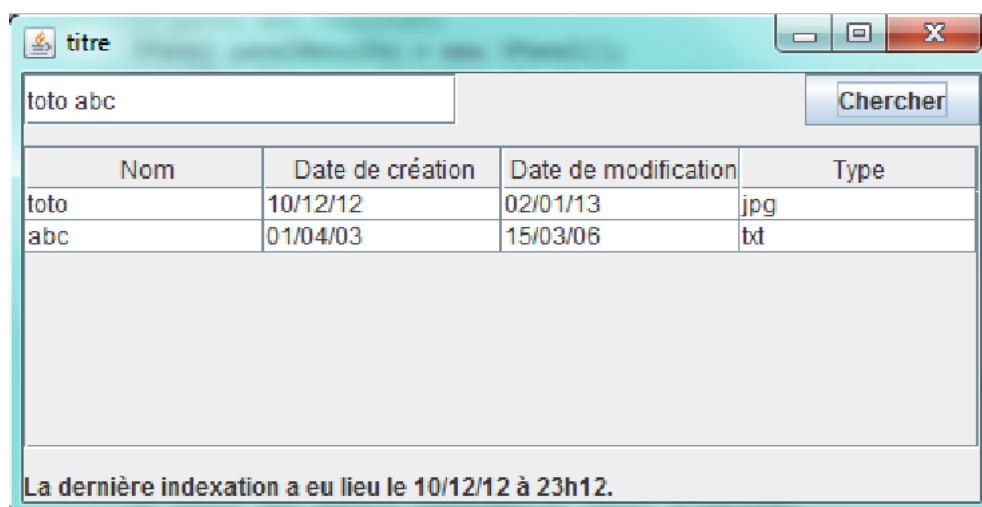


Figure 1: Recherche simple

dex. Un algorithme est appliqué pour identifier dans le corpus (en utilisant l'index), les fichiers qui correspondent le mieux aux mots contenus dans la requête, afin de présenter les résultats des recherches par ordre de pertinence. Cette base de données va jouer un rôle clé dans la phase de recherche de fichiers. Ensuite, à chaque création ou modification de fichier appartenant au corpus, le démon met à jour la base de données en fonction des modifications du fichier.

0.2.3 User characteristics

0.3 Syntaxe MR

Minuscules/ majuscules	iFind ne tient pas compte de la casse des caractères. Les requêtes 'ibm', 'Ibm' et 'IBM' renvoient le même résultat.
Lettres accentuées	Important <i>electricite</i> et <i>électricité</i> ne donnent pas le même résultat, même si les différences sont souvent minimales.
Ordre des mots	Important: <i>paris brest</i> donne un résultat différent de <i>brest paris</i> . Une plus grande importance est donnée au premier mot choisi.
Disjonction : Rechercher un mot ou l'autre, requête -or requête	-or Exemple : machin OR bidon. L'opérateur doit être saisi avec un tiret obligatoirement.
Conjonction ET	Opérateur par défaut Exemple : 'moteur recherche' recherche les fichiers qui contiennent à la fois 'moteur' ET 'recherche'. Il est également possible d'utiliser le signe + pour demander une orthographe spécifique: Exemple : +jérémie ne trouvera pas la forme "jeremie" (non accentuée)
Exclure un mot -e requête	-e Exemple : moteur -e automobile recherche les fichiers qui contiennent 'moteur' mais qui ne contiennent pas automobile.
Expressions	Non. Il n'est pas possible de faire des recherches de phrases exactes.
Troncature	Non. Il n'est pas possible de faire des recherches en utilisant la troncature sur iFind. le moteur recherche toujours exactement le mot demandé. 'mot' ne trouve pas 'mots' ni 'moteur'. L'astérisque (*) ne peut pas être utilisé. iFind tient cependant parfois compte de la troncature, sans qu'il soit possible pour l'utilisateur de décider quand.
Recherche sur le type de fichier	-f Exemple : exemples -f pdf. Plusieurs formats sont possibles.
Recherche avancée	Advanced Search, Recherche avancée Recherche sur le format de fichiers, sur la date de mise à jour, etc. Cependant, il n'y aura pas de syntaxe spécifique mais des boîtes de choix au niveau de l'interface graphique.

Utilisation normale

L'utilisateur entre une requête dans la barre de recherche.

Figure 2

Utilisation avancée

L'utilisateur peut également entrer une requête incluant des critères spéciaux sur les fichiers à rechercher :

- l'auteur
- la date de création
- la dernière date de modification
- le type
- la taille (pour les fichiers de type image)
- la durée (pour les fichiers de type musique ou vidéo)

Il est possible de paramétrer les fonctionnalités suivantes de l'indexation :

- TODO
- TODO

Figure 3

0.3.1 Contraintes générales

Contraintes sur la recherche → si un fichier a été sauvegardé mais pas quitté

0.3.2 Dépendances

Le logiciel sera utilisable sur une distribution GNU/Linux.