

Haehnel Jonathan
Knobloch Thomas
Armand Julien

« Préparation à la fouille de données »

Tuteur : Nicolas Lachiche
Ref : T305ALAN

Dossier des risques



Année 2010-2011

IUT Robert Schumann – Informatique

Sommaire :

Page de garde	<i>Page 1</i>
Sommaire	<i>Page 2</i>
Présentation du projet	<i>Page 3</i>
Sélection des risques et calcul du niveau de risque	<i>Page 4</i>
• Risques sélectionnés	
• Niveau de risque	
Analyse des risques et actions préventive	<i>Page 5-6</i>
Fiche du projet T3	<i>Page 7</i>

Introduction:

Sujet :

Le projet consiste à réaliser une application java permettant de préparer les données pour plusieurs programmes de fouille de données. Cet utilitaire devra pouvoir manipuler et extraire des données de certaines bases de données et certains fichiers. Une interface graphique simple mais ergonomique est attendue.

Demandeur :

Mr Lachiche est le responsable du projet, mais également le maitre d'ouvrage.

Contexte :

Le projet s'insère dans le contexte de la fouille de données (recherche d'un modèle à partir d'un échantillon de données). Il va permettre de faciliter ce processus en sélectionnant à l'avance certaines données dans les bases et dans les fichiers. A la main, ce type d'opération est plutôt long et laborieux, il est donc important d'automatiser cette tâche.

Limite du sujet :

Le système permettra d'extraire des données pour l'élaboration d'un modèle, mais en aucun cas, l'apprentissage de ce modèle n'est attendu dans ce projet.

Type de sujet :

Ce projet demande une petite analyse sur la structure des fichiers et des bases de données à intégrer dans notre application, mais c'est surtout du développement. Il faut en effet développer une interface graphique permettant de se connecter à des bases de données, d'ouvrir des fichiers spécifiques pour générer de nouvelles bases ou fichiers.

Notre projet devra pouvoir s'intégrer dans un projet déjà existant, les bases de données et les fichiers d'entrés existent déjà.

Sélection et calcul du niveau de risque :

a) Sélection des risques probables :

Parmi les 20 risques présentés en cours, nous avons déjà fait une sélection préalable des risques les plus probables. En effet, nous travaillerons et étudierons les **8 risques ci-dessous** durant toute l'étude.

- Définition imprécise d'un projet
- Taille d'un projet et durée
- Complexité des projets
- Sophistication
- Maîtrise des demandes
- Répartition des charges
- Méthodes / Normes
- Gestion des incidents / Réaction suite aux dérives

Pour information, les risques écartés ont un niveau de risque de 3% seulement.

b) Calcul des niveaux de risques:

<i>Paramètres généraux</i>				
Risques	Gravité	Probabilité	Indetectabilité	Pourcentage
Définition imprécise d'un projet	5	3	1	15%
Taille d'un projet et durée (délai)	4	5	2	40%
Complexité des projets	2	2	3	12%
Sophistication	1	3	2	6%
<i>Paramètres liés à la maîtrise d'œuvre</i>				
Risques	Gravité	Probabilité	Indetectabilité	Pourcentage
Maîtrise des demandes	3	3	1	9%
Répartition des charges	5	5	2	50%
Méthodes / Normes	3	3	3	27%
Gestion des incidents / Réaction suite aux dérives	5	1	1	5%

Analyse des risques et actions préventives :

<i>Paramètres généraux</i>		
Risques	Commentaire	Actions préventives
Définition imprécise d'un projet	Le projet ne concerne que la phase de préparation à la fouille de donnée, il faut donc bien délimiter les champs du projet au risque de sortir du cadre du projet.	Ce risque peut arriver dans la mesure où le sujet peut être mal interprété et donc sortir du sujet du T3. La mesure à prendre pour éviter ce risque sera une bonne explication préalable du sujet par le tuteur.
Taille d'un projet et durée (délai)	Vu qu'on doit faire le projet de A à Z, le projet est assez imposant en terme de volume de travail, il sera très important de respecter les délais intermédiaires que nous nous sommes fixé afin de finir dans les temps.	On risque de dépasser les délais si les dates intermédiaires fixées par le chef de projet et le tuteur ne sont pas respectées. Pour éviter cela, il va falloir respecter chaque échéance afin de finir dans les temps.
Complexité des projets	Les connaissances demandées par le projet peuvent nous dépasser.	C'est possible dans la mesure où le projet utilise peu de connaissances acquises en cours, pour éviter cela, il va falloir bien se documenter sur le sujet, poser de bonnes questions au tuteur et donc des rendez-vous réguliers.
Sophistication	L'envie d'intégrer d'emblée des fonctionnalités supplémentaires peut conduire à l'élaboration de traitements super-flux qui compliquent l'application.	On peut être amené à compliquer l'application en voulant « faire bien », cherchant à créer une application la plus complète possible. Cependant, cette recherche de l'exhaustivité peut entraîner une complication de l'application pour éviter cela, il va falloir bien délimiter le projet dans le cahier des charges et s'y tenir dans le développement.
<i>Paramètres liés à la maîtrise d'œuvre</i>		
Risques	Commentaire	Actions préventives
Maîtrise des demandes	Le chef de projet peut rapidement être débordé par son rôle de chef de projet et de développeur.	Le chef de projet risque d'être dépassé par les tâches de gestion du projet à défaut de celles de développement car le travail de chef de projet peut rapidement devenir volumineux (exemple : répartition des tâches, suivi de la progression du projet, création du planning ... etc) pour éviter cette surcharge, les autres membres du groupe doivent l'aider dans son rôle de chef de projet et y participer.

Répartition des charges	Un chef de projet trop impliqué peut s'accaparer plus de travail que celui qu'il distribue aux autres intervenants du projets. La répartition des charges peut ne pas être respecté par l'ensemble des membres du projet.	La plus forte motivation du chef de projet sur le sujet peut entrainer un problème dans sa répartition du travail. Il peut alors avoir tendance à se donner plus de travail, ce qui peut s'avérer dangereux car il doit en plus assumer son rôle de chef de projet. Pour éviter cela il va falloir répartir de manière équitable le travail à effectuer et s'en tenir à ce qui aura été prévu.
Méthodes / Normes	Les carences dans les domaines mis en œuvre pour l'élaboration du projet peuvent conduire a des solutions peu performantes voir déficientes.	Le projet utilise des connaissances jusque là inconnues cela risque donc d'entrainer l'élaboration de solutions peu performantes, pour éviter cela, nous allons devoir effectuer beaucoup de recherches personnelles afin de trouver les solutions les plus performantes.
Gestion des incidents /Réaction suite aux dérives	Un manque de réactivité de la part du groupe quand aux éventuelles dérives peut entraîner un effet cumulatif de celles-ci.	Si on ne réagit pas rapidement aux dérives cela risque d'entrainer de nouvelles dérives pus importantes. Pour éviter cela, il va falloir réagir rapidement à toute dérives pouvant subvenir au cour du projet et si nécessaire, en informer le tuteur.

Fiche du projet T3 :

Tuteur : Nicolas. Lachiche (T305BLAN)	Titre : Préparation de données pour la fouille
---------------------------------------	--

Contexte	<p>La fouille de données consiste à chercher un modèle à partir d'un échantillon de données. Par exemple, à partir d'un échantillon de 1000 clients, il faut trouver un modèle prédisant qui est susceptible d'acheter un nouveau produit parmi l'ensemble des autres clients.</p> <p>Les données sont en général stockées dans une base de données relationnelle. Il est souvent nécessaire d'extraire un sous-ensemble des tables et des colonnes avant de procéder à l'apprentissage d'un modèle.</p> <p>De plus, certains programmes de construction de modèles utilisent des fichiers. Il existe plusieurs formats pour ces fichiers.</p>	
	Le projet	
Travail à réaliser	à Principaux résultats attendus	Une interface graphique facile à utiliser
		Une interface avec PostgreSQL et MySQL
		Une interface avec les fichiers de ACE, HiFi et Tertius
Les outils	Nom	Description
	Java	