

**Thomas Allard**

Coordonné par Dr Michal Iglewski  
Supervisé par Dr Alain Charbonneau et Dr Jean Quirion

**INF4173 Projet synthèse**

Plan de travail

Université du Québec en Outaouais

Hiver 2008

# Table des matières

Table des matières .....	2
INF4173 Projet synthèse .....	3
Plan de travail .....	3
Introduction .....	3
Contexte .....	3
Problème .....	3
Objectifs .....	3
Étapes et tâches .....	4
Calendrier .....	6
Réalisation .....	9
Méthodologie .....	9
Bibliographies .....	9
Études de solutions .....	9
Analyse théorique .....	9
Expérimentation .....	10
Prototype .....	10
Dimensions humaines .....	10
Coûts .....	10

# INF4173 Projet synthèse

## *Plan de travail*

### Introduction

Le travail à faire dans le cadre de ce cours sera coordonné par D<sup>r</sup> Michal Iglewski, Directeur du Département d'informatique et d'ingénierie (DII) de l'Université du Québec en Outaouais (UQO), supervisé par D<sup>r</sup> Alain Charbonneau, professeur au DII de l'UQO ainsi que D<sup>r</sup> Jean Quirion, professeur au Département d'études langagières(DEL) de l'UQO.

Ce plan de travail débute par une mise en contexte du logiciel Barçah ainsi qu'une description détaillée du problème à résoudre.

### Contexte

Barçah a été développé par D<sup>r</sup> Jean Quirion et D<sup>r</sup> David Nadeau, en collaboration avec le Conseil national de recherches du Canada (CNRC), afin de répondre au cinquième critère du processus de planification terminologique (Auger 1982), soit une fonction d'évaluation et de mesure qui permet d'évaluer le progrès réalisé jusqu'au moment présent de programmes de planification langagière et terminologique. Une telle évaluation requière une méthode pour mesurer le degré auquel la terminologie a été implantée.

### Problème

Barçah a été développé en utilisant une technologie plus ou moins désuète et souffre de certaines limites reliées au moteur de recherche utilisé ainsi que plusieurs lacunes portant sur son utilisation quotidienne, entre autre, une sécurisation faible ou inexistante. Le problème à résoudre dans le cadre de ce cours est donc singulièrement l'*amélioration* générale de Barçah.

### Objectifs

Plus particulièrement, le problème à résoudre peut être divisé en trois objectifs principaux :

1. mise à jour de Barçah ;
2. réalisation des tâches majeures visant à compléter Braçah ;
3. réalisation des tâches mineures visant à compléter Braçah.

Le premier objectif, considéré comme étant l'objectif principal, comporte la mise à jour technologique de Barçah, spécifiquement le passage du Microsoft .NET Framework 1.1 couplé à Microsoft SQL Server 2000 au .NET Framework 3.0 couplé à SQL Server 2005. Pour ce faire, une familiarisation du logiciel Barçah devra être effectuée auparavant.

Le temps le permettant, les deux autres objectifs (tâches majeures suivies des tâches mineures) seront entamés à la suite de la mise à jour. En particulier, deux tâches du second objectif auront la priorité, soit la quête d'un nouveau moteur de recherche

visant à remplacer « Coveo Enterprise Search (CES) » ainsi que la sécurisation de l'interface de Barçah.

## Étapes et tâches

Les trois objectifs mentionnés ci-haut dans la rubrique Objectifs se divisent en tâches afin de mieux délimiter le travail à effectuer.

La Table I présente les deux tâches du premier objectif. La Table II précise les tâches majeures appartenant au deuxième objectif tandis que la Table III indique les tâches mineures appartenant au troisième objectif. Chacune de ces tables contient le nombre estimé d'heures requises pour implanter chaque tâche y figurant. Il est à noter que les estimations des heures de travail figurant à la Table III ont été effectuées par l'un des créateur de Barçah, soit Dr David Nadeau, de sorte que celles-ci devront être augmentées significativement étant donné que je ne connais nullement ce logiciel. Il en va de même pour les tâches de la Table II car aucune estimation antérieure n'existait et que j'ai donc dû moi-même en faire une première évaluation. Je serai en mesure de préciser la valeur de celles-ci lorsque j'aurai apprivoisé Barçah. Il est à noter que la dernière rangée de chaque table démontre le total d'heures à investir pour chaque objectif. Il est aussi à noter que les tâches ne sont pas présentées dans l'ordre dans lequel elles seront exécutées.

**Table I** : estimation du temps nécessaire à la réalisation des tâches du premier objectif

Tâche	Heures
Apprivoiser le logiciel Barçah.	25
Mise à jour de Barçah (passage du Microsoft .NET Framework 1.1 à au Microsoft .NET Framework 3.0 et de Microsoft SQL Server 2000 à Microsoft SQL Server 2005).	50
Total	75

**Table II** : estimation du temps nécessaire à la réalisation des tâches majeures existantes du deuxième objectif

Tâche	Heures
Déterminer si Barçah peut fonctionner facilement avec un autre moteur de recherche que CES (évaluation de moteurs de recherche qui pourraient être utilisés par Barçah).	À dét. <sup>1</sup> 30 ?
Déterminer comment un corpus peut être ajouté.	À dét. 5 ?
Sécurisation de l'interface (restriction d'accès), c'est-à-dire implanter des vues par utilisateur afin que les différents utilisateurs n'aient accès qu'aux seuls domaines dans lesquels ils sont impliqués. Par exemples, les utilisateurs de l'OQLF ne devraient pas avoir accès aux domaines du BT et vice-versa.	À dét. 5 ?
Peaufiner l'algorithme de désambiguïsation lexicale. Peut-être trouver une solution pour que le logiciel soit 100 % automatique ?	À dét. 40 ?
Possibilité de sauvegarder un domaine sur un support externe. Cela est nécessaire pour respecter la limite de documents de Coveo Enterprise	À dét. 10 ?

<sup>1</sup> À déterminer au fur et à mesure que le projet va se dérouler.

Search. Les domaines terminés peuvent être sauvegardés (ex. : sur CD-ROM), ce qui laisse de la place pour de nouveaux domaines. Au besoin, un domaine peut-être rechargé pour analyse, modifications, etc. Note : actuellement, la sauvegarde peut être effectuée manuellement en prévoyant quelques heures. L'idée serait d'automatiser la sauvegarde.	
On a aussi parlé récemment d'accepter une langue comme le catalan, ce qui n'était pas naturel dans Barçah (tout comme l'anglais ou n'importe quelle autre langue).	À dét. 25 ?
Appliquer des modèles de désambiguïsation d'un domaine terminé (par exemple, nanotechnologies 2005) à un nouveau domaine (par exemple, nanotechnologies 2006).	À dét. 15 ?
Total	130

**Table III** : estimation du temps nécessaire à la réalisation des tâches mineures existantes du troisième objectif

Tâche	Heures <sup>2</sup>
Format « concordancier », trier avant et après.	80
Interface de sélection de mots indicateurs de sens (cliquer dans la liste courte).	40
Pouvoir travailler sur un ou plusieurs sous-corpus, selon les arrivages.	40
Coefficient d'implantation sur 1 pour chaque terme dans chaque sous-corpus (voir fichier Excel & ajouter le statut).	10
Nombre total d'occurrences réelles pour chaque terme (voir fichier Excel & ajouter le statut).	10
Possibilité de cocher une liste d'occurrences et d'attribuer une valeur (p. ex. valide) à l'ensemble des occurrences sélectionnées ?	10
Statuts suivants : R = recommandé, N= normalisé, R.O. = recommandé par un organisme international Année, (Ex. : 71 ou 04) = terme qui est en vedette, soit dans le vocabulaire ou dans le GDT.	5
Quand on quitte les termes à l'étude pour revenir aux notions à l'étude, possibilité que le programme se souvienne d'où on était rendu plutôt que de nous ramener au début de l'index ?	5
Termes répétés, lorsque valide dans un, invalide dans l'autre.	5
Nombre de ces termes catégoriquement implantés qui sont les termes officialisés ou en vedette (parmi ceux qui n'ont aucune concurrence terminologique).	2
Nombre de notions au total.	2
Nombre de termes (anglais et français) au total.	2
Nombre de termes officialisés (recommandés et normalisés), s'il y a lieu.	2
Nombre de termes trouvés dans le corpus (dans BD, <code>if view &amp;&amp; occurrenceoverall != 0</code> )	2
Nombre de notions désignées par un seul terme (aucune concurrence terminologique).	1
Nombre de notions trouvées dans le corpus (nombre de notions pour lequel au moins 1 terme a 1 occurrence).	1
Nombre de termes anglais trouvés dans le corpus.	1
Nombre de termes français trouvés dans le corpus.	1
Préciser la base de calcul du % complété.	1
<b>Total</b>	<b>220</b>

## Calendrier

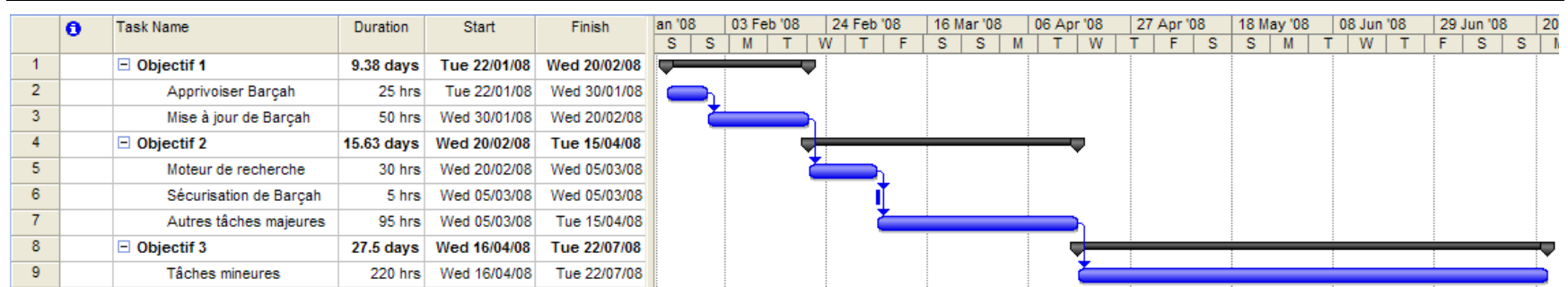
La planification pour réaliser les objectifs de ce plan de travail est basée sur une contribution hebdomadaire de dix heures, ce qui résulte en un total de 150 heures. Par contre, il est fort possible que ce total pourra être augmenté, ce en fonction du nombre d'heures supplémentaires à ma disposition.

<sup>2</sup> Selon les estimations de Dr David Nadeau.

Il est clair que le nombre d'heures au total tel qu'approximé ci haut dépasse les 150 heures normalement dépensées pour un projet synthèse typique. Par contre, il est possible que je puisse fournir le double, soit 300 heures sur un total de 425 heures et ainsi donc être capable d'entamer les tâches majeures et mineures. Toutefois, il est aussi possible que certaines tâches soient extrêmement difficiles à résoudre, ce qui pourrait diminuer le nombre total de tâches accomplies à la fin du projet.

Voici un bref aperçu graphique du calendrier associé à ce plan de travail ; voir la Figure 1. Il a été créé à l'aide de Microsoft Project 2007. Le dit plan de projet en format numérique sera utilisé tout au long de la session afin de suivre les progrès hebdomadaires. Il est à noter que ce calendrier dépasse la fin de la session mais il faut tenir compte que toutes les tâches ne seront pas réalisées. Entre autre, ce calendrier pourra être utilisé en cas de continuation du travail avec Barçah.

**Figure 1** : aperçu graphique du calendrier associé au plan de travail





## Réalisation

Les prochaines sections étayeront quelques aspects supplémentaires qui seront très probablement encourus lors de la réalisation de ce plan de travail. Plus particulièrement, le paradigme utilisé sera une application Web utilisant le langage de programmation Microsoft ASP.NET en C# secondé par Microsoft Internet Information Services (IIS) ainsi que Microsoft SQL Server 2005.

## Méthodologie

Les recommandations de Microsoft visant la mise à jour du .NET Framework 1.1 couplé à SQL Server 2000 au .NET Framework 3.0 couplé à SQL Server 2005 seront fortement considérées.

Les concepts appris dans le cadres des cours INF4043 Génie logiciel II, INF4093 Génie logiciel III, INF1553 Programmation Windows, INF3723 Systèmes d'exploitation, INF4143 Algorithmique I, INF4063 Structures des informations I, INF4163 Techniques de bases de données, INF4523 Réseaux d'ordinateurs, INF4533 Technologies Internet et INF4192 Gestion des projets informatiques seront sollicités afin de mener à terme ce plan de travail en respectant l'énoncé des objectifs à atteindre.

Une méthodologie itérative et évolutive ayant pour type le processus unifié sera recrutée afin d'éviter la « paralysie par l'analyse », fréquemment rencontrée lors de projets utilisant une méthodologie en cascade. Ceci impliquera plusieurs itérations incrémentales, chacune comprenant de plus en plus de fonctionnalités ayant un risque élevé pour l'amélioration de Barçah telles qu'énoncées dans les objectifs.

## Bibliographies

La référence ultime et unique est l'article de Jean Quirion et les citations qui s'y retrouvent : Methodology for the design of a standard research protocol for measuring terminology usage, *Terminology* 9 :1 (2003), 29-49, ISSN 0929-9971/E-ISSN 1569-9994, John Benjamin's Publishing Company.

## Études de solutions

La sollicitation d'un nouveau moteur de recherche afin de remplacer CES sera l'unique étude de solutions de ce travail puisque le logiciel Barçah est déjà utilisé et que les technologies sous-jacentes sont déjà bien établies. Il ne sera donc pas nécessaire de rechercher d'autres technologies pour ce projet. Le défi sera essentiellement de bien analyser et de bien concevoir la mise à jour des technologies utilisées par Barçah.

## Analyse théorique

La dernière tâche de la Table II aura certainement recourt à des théories mathématiques de désambiguïsation. Cette tâche nécessitera donc l'apport de D<sup>r</sup> Charbonneau et D<sup>r</sup> Quirion. Je viserai l'apprentissage des théories de désambiguïsation utilisées lorsque j'apprivoiserai Barçah.

## **Expérimentation**

Tout comme la rubrique Études de solutions, il y aura un certain degré d'expérimentation relié au remplacement du moteur de recherche CES. Autrement, il sera possible d'expérimenter de nouveaux algorithmes de désambiguïsation ainsi que de déterminer si les technologies récentes de Microsoft (i.e. .NET Framework 3 couplé à SQL Server 2005 et IIS) offrent de meilleures performances que celles couramment utilisées par Barçah.

## **Prototype**

Un prototype ne constituera pas un livrable intermédiaire lors de l'exécution de ce plan de travail, sauf à l'exception de la possibilité de divers prototypes associés à l'essai de divers moteurs de recherche afin de valider leur convenance au remplacement de CES. Le résultat de la mise à jour de Barçah constituera une nouvelle version améliorée et sera le livrable principal.

## **Dimensions humaines**

Barçah est un logiciel permettant à son utilisateur de déterminer quel est le meilleur mot à utiliser pour un contexte langagier donné. Il résulte en une amélioration de la communication entre humains. Par contre, il faudra veiller à ce que Barçah ne favorise pas un terme par rapport à un autre. En d'autres mots, Barçah se doit d'être le plus neutre possible.

## **Coûts**

Les coûts associés à l'exécution de ce plan de travail sont principalement le montant qui représente les licences pour les logiciels de Microsoft, soit Visual Studio .NET 2003, Visual Studio 2005 et SQL Server 2005. Autrement, quelques coûts de faible envergure seront nécessaires tel que l'utilisation de ressources pour l'impression des documents reliés à Barçah (articles, code source, etc.).