

Rapport de stage

Conception et développement d'une plateforme d'acquisition et
de visualisation de données relationnelles



Tuteur : xxxxxxxx

MARQUÈS
4 avril - 17 juin 2016

Rapport de stage

Conception et développement d'une plateforme d'acquisition et
de visualisation de données relationnelles
- NxTIC

LISST - UMR 5193
Université Toulouse II - Le Mirail
Maison de la Recherche
5, allées Antonio-Machado
31058 TOULOUSE Cedex 9

Note de confidentialité

Le rapport dans son intégralité n'est soumis à aucune règle de confidentialité, mais certaines annexes peuvent contenir des informations de réelles enquêtes passées par de réels sujets et sont donc soumis à une politique de confidentialité qui interdit la copie et la propagation des données sensibles.

Remerciements

Je tiens à remercier spécialement Monsieur Guillaume CABANAC, chercheur à l'IRIT, qui m'a proposé pour ce stage, ainsi que pour ses conseils et sa disponibilité.

Je souhaite remercier Julien FIGEAC, mon maître de stage, pour son investissement, sa disponibilité et sa patience tout au long du stage.

Enfin, je tiens à remercier les équipes des laboratoires LISST et IRIT que j'ai pu rencontrer tout au long du stage pour leur aide, leur accueil, leurs conseils et les échanges toujours plus enrichissants.

Table des matières

1 Introduction

2 Contexte de recherche dans un laboratoire de sociologie

3 Analyse et conception

4 Développement

5 Conclusion

6 Bilan technique et personnel

Bibliographie

Chapitre 1

Introduction

Pour finaliser ma formation au sein de l'IUT Informatique, j'ai réalisé un stage de 11 semaines en milieu professionnel. Suite à la recommandation de Guillaume CABANAC, j'ai eu l'opportunité de faire ce stage au Laboratoire Interdisciplinaire, Solidarités, Sociétés, Territoire (LISST) et au sein de l'Institut de Recherche Informatique de Toulouse (IRIT). Ce stage fut une réelle expérience professionnelle, et m'a permis une initiation au milieu de la Recherche. Il m'a donc permis de m'initier à la sociologie, son fonctionnement, son vocabulaire, et ses méthodes.

L'objectif du stage était de concevoir et développer une plateforme d'acquisition et de visualisation de données relationnelles : NxTIC. Plus précisément, une plateforme permettant aux sociologues de mener des enquêtes sous formes de sondages, et de générer des graphiques résultant des données saisies. De plus, il devait être possible d'extraire ces données à la vue d'analyses statistiques poussées. L'atout de cette plateforme pour les chercheurs, est d'optimiser le temps de passage des sondages, l'optimisation ergonomique et des temps de traitement fut par conséquent un des enjeux majeurs.

Nous présenterons dans le second chapitre le contexte de travail, les différents laboratoires et leurs fonctionnements. Dans le troisième chapitre, nous aborderons la première étape du stage : l'analyse et la conception suite à la demande des chercheurs. Puis, dans une quatrième partie, présenterons le cycle de développement et le modèle d'organisation utilisé au cours du stage. Enfin, pour conclure sur l'apport du stage, nous établirons un bilan technique et personnel.

Chapitre 2

Contexte de recherche et développement dans un laboratoire de sociologie

Dans ce chapitre, nous présenterons le contexte du stage en commençant par présenter les différents laboratoires dans les sections 2.1 et 2.2. La section 2.3 identifiera les objectifs du stage ainsi qu'une approche détaillée des fonctionnalités. Enfin, nous verrons le déroulement du stage dans la section 2.4.

2.1 Laboratoire Interdisciplinaire, Solidarités, Sociétés, Territoire (LISST)

Le LISST est un laboratoire du Labex SMS (Structuration des Mondes Sociaux), qui étudie les comportements relationnels autour de l'axe de recherche : *Innovation et savoirs : réseaux, médiation, territoires*. Le LISST comprends 85 membres permanents, dont 20 chercheurs CNRS et 48 enseignants-chercheurs et plus de 110 doctorants. 17 membres de l'unité assument des fonctions de soutien technique, scientifique et administratif. Le LISST se situe à la Maison De la Recherche de l'Université Toulouse II. Mon stage fut encadré par un chercheur du Centre National de Recherche Scientifique (CNRS) ainsi que des sociologues et statisticiens.

2.2 Institut de Recherche en Informatique de Toulouse (IRIT)

Il m'a été accordé un accès à l'IRIT afin de diffuser la plateforme et de posséder une base de données.

L'IRIT, Institut de Recherche en Informatique de Toulouse, est une unité mixte de recherche en Informatique. Il comprend à ce jour 238 enseignants-chercheurs, 38 chercheurs, et 283 doctorants à travers différents centres de recherche, tels que le Centre National de Recherches Scientifiques (CNRS).

2.3 Objectifs et missions du stage

Le développement de cette plateforme s'inscrit dans le cadre d'une recherche en sociologie qui vise à étudier les convergences et divergences de rapports sociaux entre les différents supports de communications : appels, sms, Facebook, face à face, etc. Cette enquête prend la forme d'un sondage, puis d'une étude statistique sur les données acquises. Le déroulement de l'enquête d'un sujet s'articule autour de 5 étapes:

- Enquête sur les relations dites «vraies» permettant d'étudier les relations en fonction des contextes (collègues, famille, sport, vacances, etc).
- Enquête sur les relations téléphoniques, enquête identique à la première, mais basée sur les contacts contenus dans le téléphone du sujet.
- Enquête sur les relations Facebook, enquête identique à la première, mais basée sur les amis Facebook.

- Report des données dans une feuille Excel, ou logiciel équivalent, aboutissant à des graphiques permettant de donner au sujet un aperçu de ses relations sociales. Il peut ensuite compléter les enquêtes en cas d'oubli, et avoir un retour visuel sur sa situation.
- Utilisation des données reportées pour effectuer des analyses statistiques afin de confirmer ou infirmer l'hypothèse de la recherche.

L'objectif global du système est d'optimiser le temps d'une enquête avec les fonctionnalités suivantes :

- Importation de journal d'appel permettant de récupérer tous les contacts, détecter lesquels sont les plus appelés et avec lesquels il y a le plus de sms.
- Utilisation de l'API Facebook afin d'éviter le report des amis Facebook.
- Exportation des données dans un format compatible avec les logiciels de graphiques et de statistiques.

La plateforme permettrait aux chercheurs un gain de temps important quant au passage des enquêtes qui est environ de 4 par semaine, et ce, uniquement pour la première étape. Il permettrait aussi ensuite d'éviter le report de toutes les données sur un logiciel. L'objectif fixé au début du projet était de passer à 2 enquêtes complètes par jour.

2.4 Déroulement du stage

J'ai réalisé mon stage au sein du LISST dans un Bureau partagé avec un autre étudiant de l'IUT, Romain ORTEGA. Nous avons aussi un bureau au sein de l'IRIT. Nos projets respectifs étant liés, la phase de conception comprend une partie commune car nous partageons la même base de données.

Romain devait concevoir et développer un *générateur de questionnaires* permettant aux sociologues de créer n'importe quel questionnaire avec les éléments spécifiques à leur domaine : fiches relation, matrices d'interconnaissances, générateurs de noms. Les sujets devaient passer la première étape de l'enquête sur la plateforme de Romain, à propos des relations dites «vraies». Ensuite, sur ma plateforme, les sociologues pourraient récupérer les enquêtes passées et les compléter avec les données des contacts téléphoniques, et les données des amis Facebook. Pour ainsi pouvoir visualiser, toujours sur la plateforme, des graphiques relatifs aux données et enfin, exporter les données pour effectuer des analyses statistiques.

J'ai pu tout au long du stage, échanger avec d'autres chercheurs, doctorants, enseignants de la Maison De la Recherche et de l'IRIT, ce qui m'a permis de mieux maîtriser mon sujet et d'étendre un peu plus mon expérience sociale professionnelle. Ils m'ont été d'une aide précieuse et les échanges furent tous plus enrichissants les uns que les autres.

Durant toute la période de stage, nous avons eu l'opportunité d'avoir des interlocuteurs disponibles et à l'écoute. Notre maître de stage était présent toutes les semaines pour répondre à nos questions et nous en poser. Cette organisation nous a permis de progresser dans la relation client-informaticien, en entretenant une communication régulière pour que le client soit toujours satisfait et que nous ne soyons jamais en dehors de ses attentes.

Nous parlerons dans le troisième chapitre de la phase d'analyse et conception qui s'est étendue durant la première semaine, puis tout au long du stage. (Figure 2.4-1 et 2.4-2)

Chapitre 3

Analyse et conception

Dans ce chapitre, nous présenterons la phase d'analyse et conception qui s'est déroulée la première semaine et tout au long du développement. Nous aborderons dans la section 3.1 les objectifs du projet et les solutions retenues pour les atteindre. Enfin, nous détaillerons les différentes contraintes et les solutions trouvées pour s'y adapter dans la section 3.2.

3.1 Objectifs du projet

Premièrement, l'objectif principal étant le gain de temps, il a été choisi de mettre en œuvre un système de chargement et de navigation optimisé pour éviter que le chargement et la navigation ne soient trop lents (section 3.1.1). En accord avec le gain de temps, l'ergonomie visant à simplifier l'utilisation et limiter la réflexion a été pensée pour guider au mieux l'utilisateur (section 3.1.2). Enfin, la plateforme avait pour but d'être *flexible*, c'est à dire qu'elle doit pouvoir être complétée et réutilisée par le futur (section 3.1.3).

3.1.1 Module d'optimisation réseau

Pour optimiser le chargement et fluidifier la navigation sur le site, j'ai réutilisé un module (programme) que j'ai développé et conçu auparavant et qui a évolué à travers mes différents projets personnels et professionnels (le projet tuteuré du S3 par exemple). Le principe de ce module est de charger toutes les ressources principales au premier chargement et de ne charger que le minimum nécessaire à chaque changement de page (*figures 3.1.1-1 et 3.1.1-2*).

3.1.2 Charte graphique et conception graphique

Pour réduire le temps de prise en main et d'utilisation de la plateforme, il était crucial que celle-ci soit claire et intuitive. Notamment la page de saisie des données qui compte en moyenne 1150 questions à remplir par questionnaire, soit 2300 au total. Deux aspects majeurs ont donc été pensés pour augmenter le temps du passage d'un questionnaire.

Premièrement, le questionnaire a été conçu sous forme de *timeline*, soit une frise chronologique permettant une logique dans la saisie.

Et en seconde partie, le questionnaire a été conçu pour être entièrement *dynamique*, c'est à dire qu'il ne nécessite aucun rechargement de page, et que tout se modifie en temps réel. Par exemple, si l'on ajoute un nouveau contact, sa *fiche relation* se génère automatiquement. (*figure 3.1.2-1*)

3.1.3 API et flexibilité

Une API est un système permettant de structurer toutes les fonctionnalités d'un système et de les rendre accessible depuis l'extérieur. J'ai donc basé toute la plateforme sur une API que j'avais développée, conçue et améliorée au travers de projets antérieurs. Son point fort est qu'elle peut fournir une documentation complète, qui est disponible à l'adresse <http://docs.nxtic.apiary.io>. (figure 3.1.3-1)

3.2 Contraintes du projet

Le projet a été soumis à quelques contraintes, notamment une contrainte de confidentialité que nous verrons dans la section 3.3.1. Nous avons aussi, malgré l'organisation initiale, eu quelques problèmes au niveau de la durée du stage face à l'envergure du projet. (section 3.3.2)

3.2.1 CNIL et confidentialité

Les enquêtes recueillant des données sensibles, nous avons dû élaborer un dossier pour faire enregistrer la plateforme à la CNIL. Lors de la lecture des closes du dossier et de recherches, nous avons remarqué que les informations sensibles l'étaient du moment que l'on pouvait identifier les personnes et leur associer des données personnelles et relationnelles. Nous avons pensé dans un premier temps *crypter* ou *hasher* les noms, prénoms, et autres identifiants afin de les rendre inexploitable sans un mot de passe du sujet en question. Puis nous avons tout simplement décidé de n'enregistrer qu'un pseudo pouvant soit être un prénom, soit un diminutif, car il ne représente pas un moyen direct ni indirect d'identifier les personnes en soi. Les données complètes comme le numéro de téléphone ou les noms des contacts récupérés via un journal d'appel sont donc affichés pendant la saisie des données, mais à aucun moment enregistrées. La responsabilité en terme de confidentialité dépend alors de la personne qui remplira les questionnaires. C'est ainsi que la plateforme paraît correcte quant aux directives de la CNIL. Cependant si le dossier est rejeté, il sera aussi possible de ne pas enregistrer les pseudos et de garder uniquement un identifiant numérique. Cette alternative n'est pas à écarter, mais elle engendrerait une perte de donnée, si un sujet veut modifier les données saisies précédemment il n'aura aucun moyen sûr de reconnaître ses contacts ou amis.

3.2.2 Durée limitée du stage

Le projet est un projet ambitieux qui nécessite beaucoup de tests, de modifications, on remarquera qu'il n'a été que partiellement planifié sur le diagramme Gantt (figure 2.4-1) et qu'il dépasse la période de stage. J'ai donc essayé en premier lieu de planifier le projet pour qu'il soit réalisable dans la période de stage, notamment avec une grande partie du travail réalisée hors des horaires et des jours ouvrés. C'est en voyant les demandes évoluer au cours du stage, ceci allongeant significativement le temps passé sur certaines fonctionnalités (notamment sur la version 0.2, cf. figure 2.4-1), que le laboratoire a pris la décision d'établir un CDD pour septembre 2016-2017. Ceci me permettra de prendre plus de temps pour effectuer les tests, pour finir le développement ainsi que pour prendre en compte toutes les nouveautés proposées par les clients.

Chapitre 4

Organisation et développement

Dans ce chapitre, nous présenterons l'organisation adoptée et les différents cycles de développement. Premièrement, nous aborderons rapidement les moyens d'organisation et de communication mis en place pour fluidifier la compréhension et les échanges tout au long du stage dans la section 4.1. Enfin, nous détaillerons les différents cycles de développement dans la section 4.2.

4.1 Organisation et communication

Dès la première semaine, j'ai élaboré avec le client, plusieurs moyens de communications ainsi que des moyens de le tenir informé de l'avancement du projet en temps réel.

4.1.1 Mails et SMS

Compte rendu régulier, dès qu'un problème se présente, ou dès qu'une fonctionnalités était ajoutée afin de la valider ou de proposer des modifications.

4.1.2 Partage de fichiers

Google Drive pour partager les documents utiles au projet. Les questionnaires originaux en version imprimée afin de les reporter sur la plateforme, les fichiers types de journaux d'appels afin de pouvoir les lire, etc. Le dossier de la CNIL.

4.1.3 Trello et Slack

Trello un service en ligne que j'ai découvert à l'IUT, il permet de gérer l'avancement d'un projet de manière visuelle et explicite. J'ai donc mis en place et tenu à jour tout au long du projet un tableau Trello. Le client pouvait alors connaître l'avancement actuel, les fonctionnalités restantes, et apporter ses modifications ou détails.

Slack est un système de chat en ligne enrichi pour le travail en groupe, et le développement. J'ai utilisé ce service tout au long du stage pour détailler tout ce que je faisais en temps réel. Il nous a aussi permis de dialoguer avec Romain pour avancer quand nous n'étions pas au même endroit.

4.1.4 Apiary.io

Apiary est un service en ligne permettant de créer la documentation d'une API, et permettant aussi de la tester, ce qui rend ce service très puissant.

4.1.5 Git et Gogs

Git est un système de gestion de versions, c'est à dire qu'il permet d'enregistrer le contenu d'un projet au fur et à mesure qu'il évolue, mais permet de revenir à un état antérieur sans devoir tout recommencer. Il est utilisé par beaucoup de développeurs pour travailler en équipe facilement, pour l'open-source, ou tout simplement pour permettre une maniabilité de l'historique très efficace. J'ai aussi utilisé Gogs, un service git en ligne, hébergé sur mon serveur afin de garder une copie en ligne en cas de problème.

4.1.6 Mise en ligne de la plateforme

Afin de permettre aux sociologues de tester la plateforme et la prendre en main tout au long du développement, j'ai provisoirement hébergé la plateforme sur mon serveur personnel en expliquant à chaque fois par mail ou en réunion la démarche à suivre pour utiliser telle ou telle fonctionnalité. Ceci afin de permettre aux sociologues de tester n'importe quand et depuis n'importe où le travail effectué.

4.2 Cycles de développement

Pour le déroulement j'ai opté pour un système de versions, ce qui m'a aidé à structurer la planification et expliquer plus facilement au client les différentes étapes du projet, nous présenterons ce système dans la section 4.2.1. Dans les autres sections, nous détaillerons chaque version abordée pendant le stage.

4.2.1 Système de versions et pré-versions

Le système de versions est un moyen de structurer l'avancement d'un projet par fonctionnalités et en étapes distinctes. Chaque version (ou étape) est désignée par un numéro. La version 1 correspond à la première version stable finale, toute version inférieure est donc une *pré-version*, ce qui équivaut à une étape du développement. Sur le même principe, les versions supérieures à 1 sont des ajouts, des corrections, des nouvelles fonctionnalités ajoutées à la version 1. Ci-dessous un schéma expliquant le nommage des versions.

Version **x** . **y** . **z**

x : Développement distinct

y : étape importante / fonctionnalité

z : partie distincte de **y**

Ci-dessous la liste des versions prévues pendant le stage.

Versi on	Sous- version	Libellé	%
0.1	0.1.0	Structure et base de la plateforme	
	0.1.1	Navigation et charte graphique	
	0.1.2	Sécurité et utilisateurs	
0.2	0.2.0	Acquisition des données	
	0.2.1	Prototype de l'interface et choix ergonomiques	
	0.2.2	Report du questionnaire ResTIC	
	0.2.3	Contenu 100% dynamique	
	0.2.4	Corrections, optimisation et modifications	
	0.2.5	Supplément modifications	
0.3	0.3.0	Export des données	
	0.3.1	Format, mise en forme des données	
	0.3.2	Conception d'un système d'accès aux données	
0.4	0.4.0	Visualisation graphique	
	0.4.1	Réseau	
	0.4.2	Système d'interaction	
0.5	0.5.0	Modification par les sujets	
	0.5.1	Création automatique de compte	
	0.5.2	Restitution partielle des données et modification	
0.6	0.6.0	Récupération des données LabSurveys	
0.7	0.7.0	Pré-production	
	0.7.1	Refonte et nettoyage du design	
	0.7.2	Élaboration des tests unitaires	
	0.7.3	Optimisation du chargement	