

Rapport de stage

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



Rapport de stage

Plateforme NxTIC

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

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 dans l'équipe LISST-CERS 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 et équipes dans les sections 2.1 et 2.2. La section 2.3 précisera dans quel recherche le stage s'inscrit, la section 2.4 identifiera les objectifs du stage ainsi qu'une approche détaillée des fonctionnalités. Enfin, nous verrons le déroulement du stage dans la section 2.5.

2.1 L'équipe LISST-CERS

Le LISST (Laboratoire Interdisciplinaire, Solidarités, Sociétés, Territoire) est un laboratoire de l'Université Toulouse Jean Jaurès, équipe membre 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, 48 enseignants-chercheurs et plus de 110 doctorants. 17 membres de l'unité assument des fonctions de soutien technique, scientifique et administratif.

Le LISST, laboratoire CNRS situé à la Maison De la Recherche (MDR) de l'Université Toulouse II, regroupe trois équipes : le Centre Interdisciplinaire d'études Urbaines (CIEU), le Centre d'Étude des Rationalités et des Savoirs (CERS) et le Centre d'Anthropologie Sociale (CAS). Nous faisons partie de la troisième de ces équipes : le CERS. (figure 2.1-1) Ses membres sont enseignants-chercheurs et chercheurs, recrutés sur concours national et ont un statut de fonctionnaire. L'équipe LISST-CERS est composée de sociologues et étudie, entre autres, les phénomènes sociaux relatifs aux moyens de communication modernes. La description officielle de l'équipe est disponible en figure 2.1-2.

Notre stage fut encadré par Julien FIGEAC, chargé de recherches CNRS. Nous avons aussi travaillé avec Michel GROSSETTI directeur de recherche, Guillaume FAVRE, maître de conférence, Johann CHAULET, chargé de recherche ainsi que Benoît TUDOUX ingénieur de recherches.

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 sur Internet, et de posséder une base de données. L'IRIT représente une sécurité supplémentaire quant à la protection des données sensibles.

L'IRIT, Institut de Recherche en Informatique de Toulouse, est une unité mixte de recherche (UMR) 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). L'IRIT est structuré en 7 thèmes de recherche qui regroupent les 21 équipes du laboratoire (figure 2.2-1). Les organigrammes sont disponibles en figures 2.2-1 et 2.2-2.

2.3 Recherche sur les réseaux sociaux

L'étude dans laquelle s'inscrit le stage vise à compléter un papier de Michel GROSSETTI publié en 2001 intitulé « Que font les réseaux sociaux aux réseaux sociaux? Réseaux personnels et nouveaux moyens de communication ». Ce document traite de l'impact de l'apparition d'Internet sur les réseaux sociaux (au sens de relations de personnes). De même, dans les années 80, Barry WELLMAN un sociologue approche le fait que les communications se sont individualisées avec l'expansion des moyens de communications. Où avant, un numéro de téléphone correspondait à un foyer, il se resserre de plus en plus sur un individu isolé. Suite à ces études et peu d'autres, le domaine des réseaux sociaux, bien qu'actuel, n'a été que peu desservi par le monde de la Recherche.

L'étude à laquelle nous avons participé vise à compléter l'article de Michel GROSSETTI, elle s'appuie des hypothèses suivantes :

- (1) Internet a participé à spécialiser les liens et accentuer les phénomènes d'homophilie par son choix d'interlocuteurs.
- (2) Plus on se positionne vers le haut d'une hiérarchie, moins les réseaux sociaux sont denses.
- (3) Plus un réseau est dense, plus il y a de ségrégation par la spécialisation des liens.

Plus clairement, l'hypothèse de départ est qu'Internet et les moyens de communications qui en découlent comme les réseaux sociaux, de par la possibilité de choisir ses contacts en fonction de critères, rend le réseau social des individus plus sélectif et restreint. De plus, les contacts d'un individu ne se connaissent pas forcément car Internet favorise les rencontres sans se soucier de la position géographique ou de la probabilité que les individus se rencontrent en face à face. Aussi, plus on est haut placé dans une hiérarchie, plus la sélectivité des contacts sera précise et par conséquent les contacts entre eux ont encore moins de chance de se connaître.

2.4 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 :

- (1) Enquête sur les relations dites «vraies» permettant d'étudier les relations en fonction des contextes (collègues, famille, sport, vacances, etc).
- (2) Enquête sur les relations téléphoniques, enquête identique à la première, mais basée sur les contacts contenus dans le téléphone du sujet.
- (3) Enquête sur les relations Facebook, enquête identique à la première, mais basée sur les amis Facebook.
- (4) 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.
- (5) Utilisation des données reportées pour effectuer des analyses statistiques afin de confirmer ou infirmer l'hypothèse de la recherche.

Remarque : Les étapes (1), (2) et (3) sont basées sur le même formulaire : le ResTIC.

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

- (1) 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.
- (2) Utilisation de l'API Facebook afin d'éviter le report des amis Facebook.
- (3) 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.

2.5 Déroulement du stage

J'ai réalisé mon stage au sein du LISST dans un Bureau partagé avec Romain ORTEGA, un autre étudiant de l'IUT. 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 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 souvent présent pour répondre à nos questions et nous en poser. Cette organisation nous a permis de progresser dans la relation client, en entretenant une communication régulière pour que le client soit toujours satisfait et que nous ne soyons jamais en dehors de ses attentes.

Le mardi 2 juin, nous avons eu l'opportunité de participer à une formation pour le logiciel Gephi. Ce logiciel permet à partir de données relationnelles la représentation de données sous forme de réseau. Cette formation a duré toute la journée et fut animé par Denis PARADE et organisé par l'URFIST de Toulouse. Elle nous a permis de découvrir et prendre en main la majorité des fonctionnalités du logiciel. Cette formation fut très utile car j'avais moi-même la nécessité de concevoir développer une représentation de données sous forme de réseau (section 4.2.4). Elle m'a notamment aidé à comprendre certains phénomènes comme la disposition spatiale, qui positionne de manière espacée les points sur le graphique et en fonction de leur importance. J'ai de plus ajouté à la plateforme la possibilité d'exporter les données au format Gephi car les sociologues l'utilisent pour leurs études (figure 2.5-1).

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 2.5-2 et 2.5-3)

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, 3.1.1-2 et 3.1.1-3*).

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 pour 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 graphiquement 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 les éléments se modifient 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 antécédents. Son point fort est qu'elle peut générer une documentation complète, qui est disponible à l'adresse <http://docs.nxtic.apiary.io> (figure 3.1.3-1 et 3.1.3-2). L'API intègre un système de sécurité afin que n'importe qui ne puisse pas y accéder. Elle est facilement compréhensible pour des développeurs souhaitant reprendre le projet grâce à sa documentation interactive (figure 3.1.3-2). De plus, cela rend possible l'envoi et la réception de données avec une autre application. Par exemple, si l'étude évolue et que les chercheurs ont besoin d'un programme pour récupérer directement les données NxTIC et les afficher sur Gephi. Le développeur en charge de ce programme pourra, grâce à l'API récupérer facilement les données sans avoir à modifier ce que j'ai fait.

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 au titre d'*assistant ingénieur d'étude*. 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. Ensuite, nous détaillerons les méthodes de structuration du développement dans la section 4.2. Enfin, nous détaillerons les différents cycles de développement dans la section 4.3.

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

J'ai entretenu pendant la période de stage un compte rendu régulier par mail, dès qu'un problème se présentait, ou dès qu'une fonctionnalité était terminée afin de la valider ou de proposer des modifications. Les sociologues ne pouvant être toujours présents au laboratoire, nous avons décidé de mettre en place plusieurs moyens de suivi complémentaires. Ceux-ci sont présentés dans les sections suivantes.

4.1.2 Partage de fichiers

Un dossier partagé sur le service Google Drive a été mis en place dès le premier jour afin de partager les documents utiles au projet. Notamment les questionnaires originaux en version papier afin de les reporter sur la plateforme, les journaux d'appels afin de pouvoir les lire, etc.

4.1.3 Trello et Slack

Trello est un service en ligne qui a été introduit 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. (figure 4.1.3-1)

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

Apiary.io 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. Ce service utilise un langage appelé *api Blueprint* qui est l'enrichissement d'un langage répandu : le *markdown*. Le *markdown* est un langage permettant de mettre en page un document simplement, avec l'utilisation de symboles particuliers. Par exemple, plusieurs tirets « ---- » afficheront une ligne horizontale. Le langage *api Blueprint* prend, en plus, en compte les requêtes et réponses HTTP afin de tester les fonctionnalités de l'API.

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 et pour pouvoir récupérer le projet depuis n'importe où.

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 Système de versions et pré-versions

Pour simplifier le développement j'ai opté pour un *système de versions*, ce qui m'a permis de découper le projet en étapes logiques. Ainsi, j'ai pu informer le client de l'avancement du projet et des différentes fonctionnalités à implémenter, notamment avec l'appui de *Trello* (figure 4.1.3-1) pour se repérer.

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

Note : La liste des versions est disponible en *figure 4.2-1*.

4.2.1 Version 0.1 – Structure et base de la plateforme

La version 0.1 fut la première étape du développement, elle a pour objectif la création de la base du site, la définition de son fonctionnement global et la conception de la première interface. J'ai donc, en premier lieu, implémenté la base de données qui a été conçue lors de la phase de conception (figure 4.2.1-1). Suite à ça, j'ai établi le premier prototype de la charte graphique (figure 4.2.1-2). Pour ce qui est du fonctionnement global, j'ai utilisé mon propre framework (figure 4.2.1-3) qui a évolué au cours de mes différents projets. Il comprend un système de routage, ce qui permet de charger telle ou telle page en fonction de l'URL. Il permet aussi la gestion des utilisateurs, la gestion d'une base de données, il intègre l'API, etc. J'ai ensuite réalisé le système de navigation avec le *module d'optimisation réseau* (section 3.1.1). La gestion des utilisateurs fut la dernière étape de cette version, j'ai réalisé un premier prototype et implémenté tout le système de connexion, de création et de suppression des utilisateur (figure 4.2.1-4 et 4.2.1-5). La version 0.1 a commencé le 8 avril et a pris fin le 15 avril, avec un total de 157 heures.

4.2.2 Version 0.2 – Acquisition des données

La version 0.2 a pour objectif de créer la page d'acquisition des données, elle correspond au plus gros objectif du projet et consiste au report du questionnaire RestIC qui fut élaboré avant le stage. Cette enquête était passée en face à face, il fallait donc optimiser au maximum le temps de passage. Ce fut l'étape la plus longue et complexe du projet pour différentes raisons :

- (1) La page d'acquisition est composée de trois pages : la recherche d'un sujet, l'acquisition des données des contacts téléphoniques, et l'acquisition des données des amis Facebook.
- (2) Chaque page d'acquisition comporte en moyenne 1150 champs à saisir.
- (3) Tous les questionnaires composant l'enquête sont liés les uns aux autres (figure 3.1.2-1).
- (4) Elle comprend l'export des données issues d'un journal d'appel.

La première étape fut la récupération des données d'un journal d'appel. Le journal d'appel est un fichier généré à partir d'une application mobile. J'ai pour cela eu à ma disposition 8 fichiers d'anciennes enquêtes afin de comprendre comment il était

structuré. A partir de ce fichier, j'ai pu extraire 2 types d'informations, premièrement la liste des contacts avec leur nombre d'appels et de sms échangés, enfin le journal d'appel à proprement parlé contenant tout les échanges. Le nombre d'échanges récupérés dans un journal d'appel peut aller jusqu'à 30000 échanges, j'ai donc du réaliser une première optimisation au niveau de la lecture du fichier et de l'enregistrement des données. L'enregistrement sera traité plus en détail dans les sections 4.2.3 et 4.2.4.

Ensuite, j'ai réalisé le prototype de l'interface d'acquisition afin d'éviter de perdre du temps sur l'utilisation (figure 4.2.2-1). Ceci fait, j'ai reporté le contenu du questionnaire original ResTIC et conçu le système d'interaction entre les différents questionnaires présents sur chaque page. Les sociologues ont proposés suite au résultat, nombre de modifications afin de rendre l'enquête plus complète et adaptée au nouveau format. Le conception de cette version a débuté le 15 avril et son développement a pris fin le 6 mai, pour un total de 269 heures.

4.2.3 Version 0.3 – Extraction des données

Après l'acquisition des données, la version 0.3 avait pour objectif l'extraction des données en vue d'analyses poussées ou la visualisation à travers des logiciels tels que Gephi.

La première étape fut la conception du format de sortie défini à travers les entrevues avec le personnel chargé de l'analyse. Ce format a été étudié pour avoir des fichiers contenant beaucoup de données tout en étant légers. Pour cela, j'ai utilisé un système de dictionnaire, toutes les valeurs prédéfinies (homme, femme, etc) sont remplacés par des nombres, et un fichier contenant les correspondances entre ces nombres et les valeurs réelles est fourni lors de l'extraction. Il est donc plus aisé de faire des analyses avec des nombres, tout en sachant à quoi ils correspondent avec le fichier contenant les correspondances. De plus, lors du téléchargement, les fichiers sont compressés (zip), ce qui réduit en moyenne de 75% la taille de fichiers.

Malgré le poids réduit des fichiers, la récupération des données depuis la base de données, leur mise en forme et leur compression était très longs par rapport au fonctionnement du site. Sachant que j'aurais besoin de récupérer les données en temps réel pour la version 0.4, je ne pouvais me permettre d'attendre plus de 2 secondes pour récupérer les données. Par conséquent, j'ai décidé de développer mon propre système d'accès aux données, qui s'utiliserait de la même manière qu'une base de données, mais serait plus adapté et optimisé au projet.

J'ai développé la librairie *lightdb*, un système d'enregistrement et de récupération de données, qui utilise des fichiers et quelques principes de la base de données. Afin de tester son efficacité, j'ai réalisé un *benchmark*, ce qui

consiste en une analyse des performances sur un nombre important de données (figure 4.2.3-1). Cette analyse confirme effectivement l'efficacité d'un tel système. La version 0.3 a débuté le 7 mai et s'est terminée le 19 mai, pour un total de 197 heures.

4.2.4 Version 0.4

La version 0.4 avait pour objectif toute la partie graphique représentant les données, elle n'est pas terminée à ce jour. Les graphiques étant un extra, car ils peuvent être générés à partir des données extraites, le système, à ce jour, est fonctionnel. Il y a au total 9 graphiques à afficher :

- (1) La répartition des genres (homme, femme) par rapport à toutes les communications (appels et SMS).
- (2) La répartition des types d'appels (entrant, sortant, manqué) par rapport à tous les appels.
- (3) La répartition des types de communication (appels, SMS) par rapport à toutes les communications.
- (4) La répartition des ages par rapport à la liste des contacts, et par rapport à toutes les communications.
- (5) La répartition des types de relation (parent, ami, etc) par rapport à la liste des contacts, et par rapport à toutes les communications.
- (6) La répartition des appels et SMS par rapport aux jours de la semaine.
- (7) La répartition des durées des appels.
- (8) La répartition des appels dans les heures de la journée.
- (9) Un graphique de type réseau représentant le sujet et ses contacts avec les liens entre tous les individus et positionnés en fonction de leur nombre d'apparition (figure 2.5-1).

Les graphiques ont pour but d'obtenir un récapitulatif visuel des données saisies, et de même, ils peuvent être montrés aux sujets afin qu'ils prennent conscience de leur situation et éventuellement qu'ils apporte des corrections s'ils remarquent des oublis. Le point fort de cette visualisation est que tous les graphiques doivent être « connectés ». C'est à dire que si l'on sélectionne une partie d'un graphique, tous les autres graphiques se verront mis à jour en fonction de la sélection. Il a été choisi d'utiliser la librairie « highcharts.js » pour les graphiques (1) à (8), car c'est une des plus riches en terme d'interactions et parce qu'elle permet de créer un nombre important de graphiques différents. Pour ce qui est du graphique (9), je n'ai trouvé aucune librairie adaptée à la demande qui m'a été faite, j'ai donc décidé de développer moi-même ma librairie pour cette visualisation. Un prototype de cette visualisation est disponible en figure 4.2.4-1.

J'ai conçu et développé les graphiques (1) à (8), disponibles en figure 4.2.4-2. Je n'ai malheureusement pas eu le temps de faire le graphique (9). Le développement de cette version a démarré le 19 mai et n'est pas terminée, elle

prévoit un total de 94 heures. La version sera terminée lors du CDD prévu pour septembre 2016.

4.2.5 Version 0.5

La version 0.5 n'a pas pu être réalisée dans le cadre du stage. Elle vise à gérer les comptes des sujets. A la fin d'une enquête, les sujets se verront créer un compte utilisateur qui pourra leur permettre d'accéder aux news publiées par les chercheurs à propos de l'étude. Ils pourront notamment modifier les données qu'ils ont saisies ou remplir eux-même les enquêtes manquantes sans l'aide d'un sociologue. Le développement de cette version sera effectué lors du CDD.

4.2.6 Version 0.6

La version 0.6 vise à récupérer les données de l'application *LabSurveys* développée par Romain pour compléter les informations et avoir toutes les données de toutes les enquêtes. Le développement de cette version sera effectué lors du CDD.

4.2.7 Version 0.7

La version 0.7 n'a pas de fonctionnalité particulière, mais elle correspond à tout le protocole de publication du site. Elle prend en compte l'installation sur les machines de l'IRIT, les différents tests permettant de vérifier que tout se déroule comme prévu. Une phase de correction s'en suivra, complétée par une dernière optimisation du système. Il est possible que tout le design soit revu.

Chapitre 5

Bilan professionnel et personnel

Le stage de S4, représente une première expérience en milieu professionnel, il permet notamment d'intégrer une équipe et de découvrir un métier plus en détail. Nous établirons un bilan professionnel dans la section 5.1, puis nous finirons par un bilan personnel en section 5.2.

5.1 Bilan professionnel & personnel

Le stage s'est déroulé dans un contexte un particulier, j'étais seul pour la création de la plateforme, de plus, le projet a du être mené entièrement (de la conception à la production). Pour autant, j'ai beaucoup appris de la relation client sur le fait d'être à l'écoute, ne pas hésiter à demander des précisions afin de bien cibler les besoins. Ce fut pour moi une expérience enrichissante au niveau des moyens de communication établis avec le client afin qu'il ait toujours un aperçu de l'avancement.

Durant tout le stage, j'ai pu m'initier au domaine de la recherche, tant avec l'IRIT qu'avec le LISST, j'ai pu ainsi découvrir la sociologie. J'ai trouvé les phénomènes sociologiques abordés très intéressants, de même que leur mise en pratique.

5.2 Bilan technique

Cette expérience m'a permis d'évoluer dans les langages que j'ai utilisés, tels que le *php*, le *javascript*, le *html*, le *css*, etc. Cela m'a aussi permis de faire évoluer mon framework ainsi que mon module d'optimisation réseau. J'ai pu découvrir et/ou progresser dans l'utilisation des logiciels et services :

- Gephi *Représentation de réseaux*
- GitKraken *Client git enrichi*
- Apiary *Générateur de documentation enrichi pour API web*
- Trello *Gestionnaire d'organisation*
- Slack *Chat enrichi pour les équipes et le développement*
- Gnome Planner *Générateur de diagramme Gantt*
- Inkscape *Éditeur vectoriel*
- Gimp *Éditeur d'images*
- Atom *Éditeur enrichi (IDE)*
- phpunit *Gestionnaire de tests unitaires pour PHP*
- Gogs *Serveur git*
- git *Système de versions*
- apache *Serveur web*