

Kévin DESPOULAINS
Gaël GENDRON
Corentin GUILLOUX
Valentin FOUCHER
Enzo CRANCE
Charlotte RICHARD
Laure DU MESNILDOT
Timothée NEITTHOFFER

Élèves ingénieurs de l'INSA Rennes Année universitaire 2018/2019

Rapport de spécification fonctionnelle Projet 41NFO

> Logiciel de génération de données d'apprentissage pour la reconnaissance d'écriture manuscrite

#### **Encadrants**

Bertrand COÜASNON (IRISA)

Erwan FOUCHÉ & Julien BOUVET (Sopra Steria)





Projet en collaboration avec

Jean-Yves LE CLERC

(Archives départementales d'Ille-et-Vilaine)

Sophie TARDIVEL (Doptim)





# Table des matières

1	Introduction	1
2	Rappel du contexte  2.1 Les acteurs du projet  2.2 Le prérimètre fonctionnel  2.3 Les éléments en entrée  2.4 Le périmètre de qualification  2.5 Le calendrier  2.6 Analyse des risques	2 2 3 3
3	Organisation : la gestion de projet3.1 Découpage du projet3.2 Cycle de production3.3 Mode de pilotage	5
4	L'estimation         4.1 Estimation de la partie analyse	
5	Planification : à partir de Microsoft Project5.1 Hiérarchie des tâches5.2 Affectation des ressources par tâche5.3 Jalons5.4 Diagramme de Gantt	9 10
6	Conclusion	12

## Introduction

Dans le cadre de notre projet de 4ème année sur le développement d'un générateur de bases d'apprentissage pour des reconnaisseurs d'écriture manuscrite, nous avons effectué une planification initiale des tâches à réaliser. Ce document est le troisième et dernier rapport de la phase d'analyse du projet, il présente et explique nos choix vis-à-vis de la gestion de projet, de la planification et présente les estimations que nous avons pu faire.

Dans l'objectif de vous présenter nos choix, nous rappellerons dans un premier temps le contexte, en expliquant quels sont les acteurs, pourquoi ont-ils fait appel à nous et comment nous comptons répondre leur besoin. Dans un second temps, nous décrirons la gestion que nous comptons employer afin de mener ce projet à bout, au travers des méthodes, de l'organisation ainsi que du mode de pilotage. Dans une troisième partie dédiée à l'estimation, nous expliquerons comment nous avons découpé le projet en tâches afin d'obtenir une estimation précise, que ce soit pour le développement ou pour la partie gestion. Enfin, la dernière partie permettra de mettre en valeur le travail de planification et d'estimation réalisé, en présentant l'analyse effectuée grâce à Microsoft Project.

# Rappel du contexte

### 2.1 Les acteurs du projet

Notre projet est encadré par Bertrand COÜASNON, enseignant à l'INSA de Rennes et membre de l'équipe de recherche IntuiDoc à l'IRISA, ainsi que Erwan Fouché et Julien Bouvet, ingénieurs au sein de l'entreprise Sopra Steria. Nous réalisons ce projet en relation Jean-Yves LE CLERC, conservateur et responsable de la numérisation aux archives départementales d'Ille-et-Vilaine, et Sophie Tardivel, responsable et data scientist chez Doptim.

Notre équipe est composée de huit étudiants INSA, tous en quatrième année au département informatique : Enzo CRANCE, Kevin DESPOULAINS, Laure DU MESNILDOT, Valentin FOUCHER, Gaël GENDRON, Corentin GUILLOUX, Charlotte RICHARD et Timothée NEITTHOFFER.

### 2.2 Le prérimètre fonctionnel

Ce projet nous a été proposé par Bertrand Couasnon et a pour objectif de générer de manière automatique des bases d'apprentissage pour des reconnaisseurs d'écriture manuscrite à partir de documents numérisés et de leur transcription numérique. Les reconnaisseurs d'écriture ont besoin d'apprendre à reconnaître les caractères manuscrits par des exemples pour apprendre. Notre but est donc de concevoir un logiciel et de le développer afin qu'il puisse générer ces exemples dans des bases d'apprentissage. Au travers d'une interface (IHM), l'utilisateur pourra également modifier les exemples générés dans la base afin de les corrigés s'il ont mal été générés. Il pourra également les supprimer de la base.

Ce projet s'installe dans la continuité des efforts de transformation numérique des archives. En effet, les modes de consommation de l'information changent et tendent à se diriger vers le tout numérique : c'est dans cette approche que les archives départementales d'Ille-et-Vilaine font appel à nos services afin de permettre à tous d'accéder aux informations archivées depuis chez soi via une application web, évitant ainsi de se déplacer dans les locaux des archives.

#### 2.3 Les éléments en entrée

#### 2.3.1 Les données en entrée de l'application

Les données en entrée de l'application sont les documents manuscrits numérisés, leur transcription et des annotations. Ces documents pourront être de différents types : des articles de presse, des actes de naturalisation ou de mariage, etc. Ces données devront être formatées correctement pour être utilisées par notre programme.

#### 2.3.2 Les sources d'information existantes

Pour nous aider tout au long du projet nous pouvons nous aider d'une base existante contenant des documents numérisés ainsi que leur transcription et des annotations. Elle servira donc à tester notre programme tout au long de son développement. Nous avons également des outils de détection de lignes dans des documents fournis par Bertrand Coüasnon.

### 2.4 Le périmètre de qualification

Le résultat attendu par le client est une application pouvant être utilisable par n'importe quelle personne disposant d'un reconnaisseur d'écriture manuscrite et désirant générer une base d'apprentissage rapidement à partir de documents déjà retranscrits. Le niveau de qualité attendu est donc assez élevé, il faudra privilégier la qualité des fonctionnalités plutôt que leur quantité. C'est pourquoi nous sommes partis sur un logiciel simple, ne fonctionnant sans aucune connexion internet. En plus des attentes fonctionnelles, la qualité du code doit aussi être soignée. Avoir un code propre, lisible et compréhensible est un point-clé, en effet cela permettra aux promotions suivantes de continuer le projet si toutes les fonctionnalités n'ont pas été implémentées, ou si de nouvelles idées d'amélioration émergent dans le futur.

Afin d'arriver à ce niveau de qualification, nous allons devoir mettre en place plusieurs stratégies. Tout d'abord nous ferons des tests unitaires sur l'application afin de nous assurer de la robustesse du système de gestion des données. Ensuite, du côté client, les tests seront surtout faits directement sur l'interface utilisateur pour détecter des problèmes fonctionnels d'une part et problèmes d'affichage d'autre part.

#### 2.5 Le calendrier

Les ressources humaines ne seront pas les mêmes tout au long du projet. En effet certains membres du groupe partent en mobilité lors du second semestre de l'année : Kevin DESPOULAINS, Gaël GENDRON et Corentin GUILLOUX. Il restera donc cinq membres pour la partie développement de l'application.

Livrables du projet	Date de rendu
Rapport de pré-étude associé à la phase d'analyse	Jeudi 25 octobre 2018
Rapport de spécification fonctionnelle	Vendredi 30 novembre 2018
Dossier de planification	Mardi 18 décembre 2018
Rapport de conception logicielle	Vendredi 15 Février 2019
Page HTML résumant le projet	Vendredi 29 mars 2019
Rapport final	Mardi 7 mai 2019
Livraison du projet	Jeudi 9 mai 2019

Présentations du projet	Date
Soutenance de planification	Jeudi 20 décembre 2018
Première soutenance de projet	Vendredi 21 décembre 2018
Dernière soutenance de projet	Jeudi 9 mai 2019
Showroom des projets	Vendredi 10 mai 2019

## 2.6 Analyse des risques

Lors de la phase de développement de l'application, nous avons envisagé plusieurs risques potentiels qui pourraient retarder le bon déroulement du projet. Nous avons donc prévu dès maintenant des solutions, présentées ci-dessous, afin de pouvoir continuer à avancer normalement en cas de problème. Risque: Absence non prévue (maladie, etc.)

Impact : Plus de travail pour les autres, nécessité de rattraper le retard pour l'absent

Action 1 : Prévenir au plus tôt dans la mesure du possible pour minimiser les effets de surprise Action 2 : Bien commenter le code pour pouvoir reprendre facilement la partie de l'absent

Risque: Surcharge de travail avec les cours

Impact : Moins de temps pour travailler sur le projet

Action : Prendre de l'avance sur les périodes moins chargées

Risque : Bug incompréhensible, zone de blocage

Impact : Perte de temps et risque de retarder les autres membres de l'équipe

Action: En parler avec les autres pour ne pas rester coincé seul

Risque : Blocage dû à la parallélisation des tâches (exemple : le développeur de la partie front-end en

attente de la mise en place de la partie base de données)

Impact : Perte de temps et membres du groupe non utilisés

Action: Ne pas attendre en silence, en parler au chef de projet pour faire une autre tâche en attendant

Risque : Problème relationnel entre des personnes du groupe

Impact : Création de conflits, perte de communication et donc d'informations

Action: Mettre les choses à plat, ne pas faire monter les tensions en gardant le silence

Risque: Problème relationnel entre des personnes du groupe

**Impact**: Création de conflits, perte de communication et donc d'informations

Action 1: Mettre les choses à plat, ne pas faire monter les tensions en gardant le silence

Action 2 : En parler au chef d'équipe, qui sera chargé de désamorcer les tensions

Risque: Mauvaise interprétation de ce qui est demandé par le client

Impact : Perte de temps à développer inutilement

Action: Bien communiquer au sein de groupe et avec le client pour être sûr d'avoir la même vision des

choses

# Organisation: la gestion de projet

### 3.1 Découpage du projet

Le projet peut être découpé en quatre blocs distincts. Nous avons donc réparti chacun des membres de l'équipe sur ces blocs. Cette façon de découper le projet possède des avantages comme des inconvénients.

Le premier avantage est que nous pouvons faire avancer les différentes parties du projet en parallèle et donc de gagner un temps précieux sur sa réalisation globale. Le second avantage est que tous les membres de l'équipe n'ont pas à se former sur la totalité des technologies que nous allons utiliser. Cela permet de gagner encore plus de temps. Le premier inconvénient est que seulement certaines personnes seront compétentes sur le blocs. Par conséquent, lors d'absences des mêmes membres d'un bloc, ce dernier sera gelé. Il faudra aussi penser à prendre cet inconvénient en compte lorsque trois d'entre nous partiront en mobilité. Le second inconvénient est qu'il peut être difficile de regrouper les blocs une fois qu'ils seront développés.

Voici les blocs que nous avons défini :

- traitement des données en entrées;
- création et communication avec la base de donnée;
- IHM;
- interface avec les reconnaisseurs.

## 3.2 Cycle de production

Étant donné le contexte scolaire dans lequel est réalisé ce projet, baser notre gestion de projet sur un cycle en V itératif nous a semblé particulièrement adapté. Par ailleurs, notre façon de travailler lors de la phase d'analyse nous a conforté dans cette idée. Nous proposons donc de fonctionner en suivant cette méthode pour les livrables jusqu'à la phase de conception, puis en suivant un cycle en V itératif pour la réalisation logicielle.

Nous avons ainsi défini deux cycles ou itérations. Le premier ayant déjà débuté et se terminant début février, le second débutant à la suite du premier et allant jusqu'à la fin du projet. L'objectif du premier cycle est d'anticiper l'inconvénient d'incompatibilité entre les blocs. Nous allons donc, avant la fin de ce cycle, développer une base de chaque bloc qui devra être capable de communiquer avec les autres blocs. Si la communication peut se faire sans problèmes, nous pourrons nous lancer dans le développement approfondi de nos parties. Dans le cas contraire, il faudra d'abord résoudre les problèmes de compatibilité.

### 3.3 Mode de pilotage

Afin de gérer la liste des tâches, Trello nous permettra de les attribuer aux différents développeurs selon la planification prévue et de suivre en temps réel leur avancement. Microsoft Project nous permettra de calculer le temps passé sur chacune d'entre elles, afin de comparer notre avancement réel avec les prévisions. Comme nous avons déjà effectué la planification, nous serons en mesure de comparer notre avancement réel face à celui supposé, afin de déceler d'éventuels retards, et d'y trouver des solutions comme, par exemple, laisser une fonctionnalité moins importante de côté pour se focaliser sur l'essentiel ou alors demander un allongement de délai pour la réalisation. Ces risques ainsi que les actions à réaliser en réponse ont été décrits précédemment dans ce rapport.

Nous estimons que deux cycles sont suffisants car, une fois les problèmes de compatibilité entre les blocs résolus, les améliorations apportées à chaque blocs seront internes à ceux-ci et donc n'induiront que très peu (si ce n'est aucun) problème de compatibilité supplémentaire. Faire plus de cycles n'est donc pas nécessaire mais pourra être envisagé si le besoin se fait ressentir au cours du second semestre. Ce mode de pilotage nous permettra ainsi d'avoir un logiciel fonctionnel au 9 Mai 2017, date de livraison du projet, utilisable par le client et pouvant être amélioré dans le futur.

## L'estimation

### 4.1 Estimation de la partie analyse

L'estimation des charges de la partie non développement se résume en deux grands points : le temps passé en réunion ainsi que le temps passé pour la rédaction des rapports et à la préparation des soutenances. Cette partie n'a pas été la plus complexe à estimer car nous avons pu nous baser sur les temps passés lors des trois premiers mois du projet où nous nous sommes concentrés sur la compréhension du cahier des charges et la planification du projet.

Pour les deux grands points, le calcul a été sensiblement le même. Dans un premier temps, pour les réunions, nous nous sommes basés sur le temps hebdomadaire passé en réunion depuis le début du projet, au mois de septembre, que nous avons multiplié à la fois par le nombre de semaines restantes et par le nombre de personnes dans le groupe. Dans un second temps, nous avons calculé le temps nécessaire à la réalisation des prochains rapports. Dans ce cas, nous avons multiplié le temps passé en moyenne par chaque personne pour l'écriture d'un rapport par le nombre de personnes dans le groupe et par le nombre de rapports qu'il nous reste à rendre. Nous avons pu calculer la charge des rapports de cette façon, car nous avons réparti leurs réalisations de façon homogène entre les différents membres du groupe.

## 4.2 Estimation de la partie développement

Cette partie a été plus compliquée à estimer car nous avions peu de repères. Pour nous rapprocher au maximum d'une estimation fiable, nous avons estimé le temps de développement de chaque partie en considérant les points suivants :

- connaissance des technologies utilisées (donc temps d'apprentissage nécessaire estimé);
- temps estimé de conception;
- temps estimé de développement .

#### 4.2.1 Estimation du temps d'apprentissage

Les différentes technologies qui seront utilisées ont été définies dans le rapport de spécification. Nous savons donc déjà quelles technologies devront être apprises avant de commencer le développement. Le temps d'apprentissage a été défini grossièrement en nous basant sur notre expérience d'auto-formation sur des langages ou d'autres types de technologie.

#### 4.2.2 Estimation du temps de conception

La conception ayant majoritairement été réalisée précédemment, cette partie est presque inexistante. Cependant, il est possible que la conception réalisée soit éloignée des réalités techniques. Nous avons donc prévu un temps de remodélisation si nous nous rendons compte à un moment qu'une modélisation n'est pas la meilleure.

#### 4.2.3 Estimation du temps de développement

Cette partie de l'estimation est la plus compliquée à estimer. En effet, il n'est pas possible de prévoir les problèmes que nous allons rencontrer. C'est pourquoi, nous nous sommes basés sur le temps de développement d'autres projets que nous avons réalisé en y ajoutant une marge au cas où nous serions confrontés à un problème.

# Planification : à partir de Microsoft Project

#### 5.1 Hiérarchie des tâches

Le projet est découpé en 6 tâches principales. Les tâches représentent les différentes étapes techniques par lesquelles nous devrons passer pour développer le projet.

- préparation des données;
- stockage des données;
- interface avec le reconnaisseur;
- IHM :
- généralités;
- écriture des rapports et préparation des soutenances .

Le projet se déroule de manière parallèle en règle générale, mais chaque bloc a un fonctionnement interne séquentiel, chaque grande tâche est divisée en sous-tâches.

Tout d'abord, la partie **préparation des données** consiste à traiter les documents que l'utilisateur rentre dans le logiciel pour pouvoir les découper et les mettre dans un format intégrable à la base de données. Ensuite, la partie **stockage des données** consiste à créer une base de données (choix de la manière de stocker les données établi dans le rapport de spécifications) pour enregistrer les exemples d'apprentissage données par l'utilisateur et fournir une interface pour extraire ces données. La partie interface avec le reconnaisseur consiste à fournir une interface entre le stockage et le système de reconnaissance d'écriture manuscrite. La partie **IHM** est la partie fournissant l'interface entre le logiciel et l'utilisateur. La partie **généralités** est une partie regroupant des travaux globaux sur le logiciel (lier les blocs entre eux, concevoir un logiciel évolutif, etc.). Enfin, la partie **rapports et soutenances** concerne tout le travail en lien avec la rédaction des rapports et des soutenances, cette partie représentant une partie non négligeable du travail et surtout la partie la plus contrainte aux dates de rendus.

### 5.2 Affectation des ressources par tâche

Nous avons fait le choix de spécifier une ressource générique, un membre de l'équipe dans notre cas. Ce choix nous permet de garder une certaine liberté sur l'attribution des tâches durant le développement en nous permettant notamment d'affecter les tâches à un autre développeur à un instant donné. Pour le moment les développeurs ont été répartis sur chacun des blocs mais si certains blocs avancent plus vite que d'autres, il sera possible de réattribuer les rôles. Nous avons essayé de mettre au moins deux personne par tâche pour que personne ne se retrouve seul face à un problème. Pour les tâches plus simples ou plus rapides nous avons mis une seule personne.

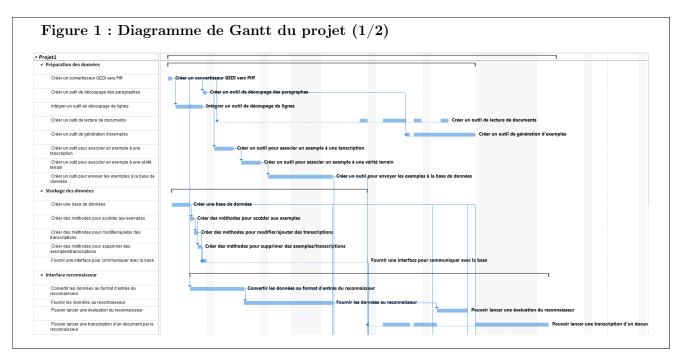
Ainsi Laure et Charlotte travaillent sur la partie IHM, cette partie étant importante et comprenant suffisamment de travail pour nécessiter deux personnes dessus s'y consacrant entièrement. Timothée travaille sur la partie interface avec le reconnaisseur, cette partie est plus petite que les autres et ne nécessite qu'une personnes pour être réalisée. La partie stockage des données est une partie différente des autres car la base de données a déjà été terminée lors du premier semestre. Ainsi, Timothée et Valentin travailleront dessus au début du second semestre pour vérifier son bon fonctionnement puis Valentin travaillera dessus seul pour terminer l'interface permettant d'y accéder. De même, le développement de la partie préparation des données a été entamé durant le premier semestre. Bien qu'étant une partie importante du projet, elle est déjà bien avancée. De fait, au début du second semestre, Enzo travaillera seul dessus, puis il sera rejoint par Valentin pour l'aider.

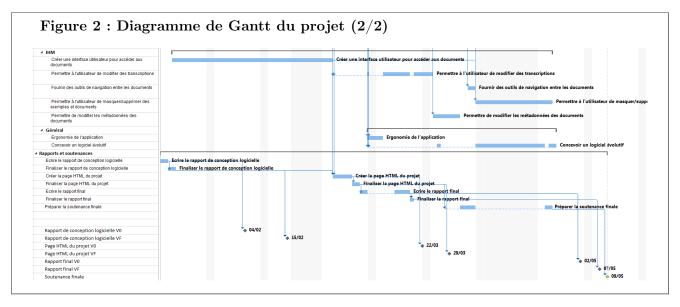
Cette répartition permet de répartir la charge de travail de manière équilibrée. Selon notre estimation du temps de travail pour chaque tâche, chaque membre du groupe a une charge de travail égale.

#### 5.3 Jalons

En plus de la planification du temps de développement, nous avons intégré à notre planning des jalons qui correspondent aux dates de rendus et de soutenance. Nous avons attribué l'ensemble des ressources sur les phases de préparation des rendus et des soutenances.

### 5.4 Diagramme de Gantt





Le diagramme de Gantt ci-dessus décrit l'ensemble de la planification du projet, de la conception au rendu final en passant par la gestion de projet ou le développement de chaque itération. On y retrouve également les jalons dont nous avons parlé précédemment.

Comme dit précédemment, notre projet est consistué de deux itérations. La première itération est constitué des tâches à effectuer avant la création de la page HTML. La seconde est composée des tâches restantes.

Comme on peut le voir sur le diagramme, les quatre blocs sont biens parallèles et découpés en tâches.

## Conclusion

Ce rapport et les deux présentations qui donnent suite à ce dernier représentent la fin de la phase d'analyse et de spécifications de notre projet de quatrième année.

Tout au long de la phase de planification, nous avons vu et estimé la durée de chacune des tâches du projet. Cela nous a permis d'établir une planification initiale des tâches, du début de la phase de conception jusqu'au rendu de la dernière version du projet, en y incluant la gestion de projet. Nous avons ainsi découpé le projet en 6 tâches principales, elles-mêmes découpées en une multitude de sous-tâches détaillées sur le diagramme de Gantt de la partie précédente. En ayant considéré 1 heure de travail par personne et par jour, ainsi qu'une semaine dédiée au projet contenant 7 heures de travail par persone et par jour, notre première estimation annonce la fin du projet au 26 avril 2018. Évidemment, nous ne sommes pas à l'abri d'imprévus, c'est pourquoi nous avons quelques jours de travail disponibles avant la fin du projet.

Cette phase, dont l'importance ne nous semblait, à priori, pas capitale, nous a finalement paru très importante : elle nous a permis de voir que la gestion de projet prenait une grande place dans la réalisation dudit projet, mais également que ce dernier n'est pas trop ambitieux par rapport à nos capacités. La majeure difficulté de la phase de planification a été d'estimer correctement la durée des tâches, chacun de nous ayant souvent une estimation assez différente. Pour palier ce problème, nous avons décidé de faire la moyenne de toutes nos estimations afin d'obtenir un résultat certainement plus proche de la réalité que nos estimations initiales.

Cette phase de planification nous a beaucoup apporté, nous nous sommes rendus compte de la difficulté de la répartition des charges ainsi que de la difficulté d'estimer la durée d'une tâche. Par ailleurs, cela nous a également permis de nous rendre compte de l'importance de bien découper toutes ces tâches en sous-tâches, afin de ne rien oublier et ainsi pouvoir terminer le projet dans les délais annoncés.









