

# Améliorer le traitement des demandes SAV grâce au NLP

CONSTANT, Pierre, [pierre.constant@amplitude-laser.com](mailto:pierre.constant@amplitude-laser.com), Amplitude Laser, Bordeaux

Thématique : Structuration des données

Résumé : *Nous avons développé un pipeline d'analyse des demandes clients SAV (reçues par email), permettant à la fois d'accélérer le temps de traitement et de gagner en qualité. Ce pipeline utilise différentes techniques de NLP afin d'identifier le type de demande client, ainsi qu'une approche basée sur l'identification de mots clés (tagging) par des experts, pour rechercher les demandes similaires déjà rencontrées par le passé.*

Mots clés : NLP, tagging, email, données non structurées

## 1. Introduction

Amplitude est une ETI industrielle concevant et fabriquant des lasers ultrabrefs pour l'industrie (ex : usinage de précision, semi-conducteur), médicale (ex : ophtalmologie) et la science (ex : imagerie, physique des hautes énergies).

Quand un client envoyait une demande SAV par mail, Amplitude traitait celle-ci en s'appuyant essentiellement sur l'expertise du technicien, qui n'exploitait pas de manière efficace l'historique des interventions SAV.

Selon l'expertise et l'expérience du technicien traitant la demande, les délais et la qualité de traitement pouvaient varier.

Nous avons donc cherché à améliorer le traitement des demandes SAV : à partir de la demande client (reçue par email), comment identifier des demandes similaires dans notre historique de données, et proposer ces informations au technicien pour l'aider à répondre à la demande client (qualité et délai de la réponse).

Ce travail a été mené avec le CATIE dans le cadre d'un projet de R&D collaboratif.

## 2. Méthodologie

### a. Sources de données :

Les sources de données utilisées dans le cadre de ce travail sont des données textuelles issues :

- Des emails envoyés par les clients puis traités par le SAV
- L'historique des Tickets SAV (enregistrés dans le CRM d'Amplitude)

Dans les deux cas, ces données sont non structurées, en plusieurs langues (anglais, français, « franglais »), extrêmement « jargonantes » (termes spécifiques au laser de façons générale, à Amplitude en particulier).

Autre point d'importance : une des forces d'Amplitude est de personnaliser ses produits pour répondre à la demande du client. Cette agilité induit une forte hétérogénéité de produits installés, et par conséquent de demandes / tickets SAV selon le niveau d'expertise du technicien :

- Certains tickets sont très renseignés, mais d'autres très peu
- Certains termes de jargon sont proches en termes de signification (ex : front-end, frontend, oscillator, osc) mais plus ou moins utilisés

Enfin, nous avons constaté qu'un problème identique s'expliquait différemment selon les clients et leur niveau d'expertise, leur historique avec Amplitude, etc.

## b. Spécifications techniques :

Nous avons cherché à répondre à différentes spécifications dans le cadre de ce travail :

Spécification	Solution technique
Identification d'une première demande client (email)	Application de diverses règles sur les emails, basées essentiellement sur : 1/ la longueur de la discussion 2/ la présence ou non d'un numéro de ticket dans l'objet, et si oui du statut du ticket (fermé ou non)
Classification du type de demande client : Information/ Technical Request	Application d'un modèle de type TF-IDF entraîné sur des emails de première demande client en français et en anglais.
Identification du problème rencontré par le client	Application de modèles de « question answering » à l'état de l'art pour le français (CamemBERT) et l'anglais (DistilBERT). En posant des questions du type (« Quel est le problème » ?)
Extraction de tickets similaires par le passé	Quelques centaines de couples première demande / tickets CRM ont été « taggués » par des experts du SAV dans le but d'identifier les termes saillants (ou « tags »). Ces personnes ont utilisé un outil open source, le « tagging » constituant à identifier des mots ou groupes de mots qui décrivaient le ou les problèmes mentionnés dans l'email/ le ticket. Plusieurs centaines de tags ont été identifiés.  Ces tags ont ensuite été regroupés en familles : chaque famille représente un type de problème spécifique, et la famille contient les tags qui décrivent ce problème, et ce à la fois d'un point de vue client, ou Amplitude.  Une fois ces familles obtenues, la CRM a été « tagguée » : identification des tags dans chaque ticket et de la / des famille(s) associées. Un score a également été attribué à chaque tag sur la base de son occurrence dans le corpus.

	<p>Enfin, pour chaque nouvelle demande, l'email du client est analysé pour y rechercher les tags, puis les tickets les plus similaires selon un principe de score sont retournés (nombre de tags en commun, poids des tags ; un seuil est appliqué pour s'assurer d'une similarité forte).</p> <p>NB : dans le cas d'une nouvelle demande où la similarité n'est pas assez forte avec des tickets similaires, nous ne retournons pas de tickets à l'utilisateur (objectif de pertinence forte)</p> <p>Des statistiques descriptives sur ces tickets similaires peuvent alors être calculées (ex : proportion d'interventions à distance, durée moyenne de réparation etc.)</p>
--	--

Pour chaque spécification, la performance de l'outil est mesurée sur la base du niveau de satisfaction des utilisateurs. (email correctement classifié/ non, tickets similaires pertinents/ non, etc.)

### **c. Infrastructure :**

3 conteneurs réalisent actuellement les spécifications mentionnées ci-dessus :

- Interactions avec les API Microsoft 365, notamment gestion du webhook permettant d'être notifié à chaque réception d'email sur la boîte email cible.
- Application du pipeline mentionné ci-dessus (identification nouvelle demande ou non, classification du type de demande, identification des tickets similaires etc.)
- Interface graphique permettant aux utilisateurs d'accéder aux différents résultats.
  - NB : cette interface permet également aux utilisateurs de nous donner leur niveau de satisfaction (email correctement classifié/ non, tickets similaires pertinents/ non, etc.)

## **3. Originalité / perspective**

### **a. Originalité :**

La base de tags est une façon de structurer la connaissance dans un contexte où les données sont peu structurées, constituées d'un vocabulaire très spécifique, et relativement en faible volume (surtout compte tenu de la diversité de nos produits).

Cette problématique est probablement assez commune au sein des PME/ ETI (à fortiori dans un contexte B2B avec une forte composante technique).

## **a. Prochaines étapes :**

- Base de tags : tagguer plus de couples emails/ tickets, et mettre en place un système permettant d'identifier de nouveaux tags en continu
  - NB : la documentation technique des produits Amplitude, voir des articles scientifiques portant sur les lasers pourraient également constituer une source de données pertinente pour l'identification de tags
- Amélioration des modèles :
  - Classification Information/ Technical Request : remplacement du TF-IDF par un modèle plus élaboré, quand le volume de données le permettra
  - Question Answering : un seul modèle actuellement disponible pour le français (qui est donc appliqué par défaut). Une veille sur ce sujet est à faire
- A réfléchir avec les responsables métiers : mécanisme de priorisation des emails (par exemple en utilisant des techniques de sentiment analysis ainsi que d'autres règles métier)

## **Références**

A propos d'Amplitude : <https://amplitude-laser.com/>

A propos du CATIE : <https://www.catie.fr/en/home/>