Projet R&D IA NLP - Analyse des résultats

# Objectifs

## Objectifs système expert :

Loïc – Giles (Moundir - Stéphane)

* Analyser les résultats (statistique des résultats global et sur le niveau d’identification par jour)
* Préparation de données
  + Corriger les erreurs
  + Affecter manuellement les non reconnus

## Objectifs système IA :

### Mise en place et analyse du système IA :

Loïc – Gilles (support/consultant : Moundir - Stéphane)

* Automatiser le traitement
  + Définition des besoins (2 jours)
    - Besoins sur les données (Moundir – Gilles - Loïc)
    - Besoins de l’environnement de test (Moundir – Gilles - Loïc)
    - Besoins pour la liaison entre Model – environnement de test – Talk Messagerie ( Gilles - Loïc)
  + Mise en place d’un environnement de Test (4 jours)
    - Mise en place de la réception / création des données ( Gilles - Loïc)
      * Préparation des données (Moundir – Gilles - Loïc)
    - Préparation de l’environnement ( Gilles - Loïc)
    - Test isolé Model – Environnement de test (Moundir – Gilles - Loïc)
  + Branchement du Model – Environnement de test – Talk Messagerie (Gilles – Loïc)
* Méthodologie pour analyser les résultats
  + Définition des besoins (5 jours)
    - Définition des KPI (Moundir – Gilles – Loïc - Stéphane)
    - Définition de(s) méthodologies(s) (Moundir – Gilles – Loïc - Stéphane)
  + Mise en place de(s) méthodologie(s) (Code/Requête) (Gilles – Loïc) (~2 jours)
* Mettre en place le choix du meilleur prétendant (voir Q:\Production\r&d\développement\Talk messagerie\destinataire\Le choix du meilleur prétendant.docx)
  + Définition des besoins (4 jour)
    - Définition des KPI (Moundir – Gilles – Loïc - Stéphane)
    - Définition de(s) méthodologie(s) (Moundir – Gilles – Loïc - Stéphane)
  + Mise en place du choix (Code/Requête) (Gilles – Loïc - Stéphane) (~2 jours)
* Analyse des résultats et amélioration des modèles (∞)
  + Analyse des résultats (Moundir – Gilles – Loïc - Stéphane)
  + Définition des axes d’améliorations (Moundir – Gilles – Loïc - Stéphane)
* Méthodologie pour un entrainement dynamique des modèles
  + TODO

### R&D et amélioration système IA :

Moundir (support/consultant : Loïc – Giles - Stéphane)

* Définition des critères de validation
  + Création d’un datastore d’adresses postales normées:
    - Chacune des adresses est associée à la ou les clés destinataires correspondantes.
    - Ces adresses seront construites de la table “destinataire” via les clé étrangères la reliant aux tables “voie”, “typevoie” et “commune”
  + Pré-enregistrement des différents encodages vecteur du datastore créé:
    - Ce sont des vecteurs qui sont supposés stocker l’information de similarité sémantique des adresses de destinataires.
    - Ils seront utilisés par la suite comme des bases de référence parmi lesquelles nous allons récupérer les adresses normées les plus similaires.
  + Choisir et mettre en place le récupérateur d’adresses similaires:
    - C’est généralement un algorithme des plus proches voisins (kNN).
    - Ce dernier devra être choisi selon sa capacité à récupérer rapidement un nombre défini d’adresses voisines à l’adresse EDI.
  + Pré-enregistrer l’arbre de recherche des adresses voisines de chaque type d'encodage du datastore via l’algorithme kNN. Le pré-enregistrement devra nous éviter de calculer l’arbre à chaque fois qu’une adresse est reçue.
  + Mettre au point le pipeline qui permet :
    - de prendre une adresse EDI, l’encoder en différents types de vecteurs,
    - les passer par le kNN qui s'en chargera de récupérer les k premiers indices des vecteurs d’adresses similaires pour chaque type.
  + Mettre en place le processus de vote qui s'en chargera de:
    - tirer une liste finale d’adresses voisines parmi les différentes listes de k voisins des types d’encodage.
* Utiliser les fonctions déjà existantes pour la récupération des noms et adresses postales des livraisons via l’EDI du jour.
* Effectuer un nettoyage naïf et rapide des champs Adresse1 et Adresse2. Le but sera d’enlever quelques symboles parasitaires de l’adresse EDI.
* Trouver le destinataire associé à travers la comparaison entre le nom associé à l’adresse EDI et les noms associés aux adresses normées récupérées.
* Processus général de la solution IA
  + Pipeline de récupération
  + Création de liste finale
  + Trouver le nom destinataire
    - Finalisation de la fonction pour rapprocher les destinataires
* Etapes du vote pour la construction de la liste finale

# Ressources

* Loïc
* Gilles
* El Moundir
* Stéphane
* Dylan