

#### Sujets pour un stage IA/ ML En traitement automatique du langage

Novembre 2022

#### ideactiv en quelques mots

- ideactiv est le 1er moteur de recherche d'activités et événements culturels et touristiques. Il permet de trouver en quelques instants quelque chose à faire ou à voir, correspondant à vos gouts, tous les jours et partout en France (et bientôt au-delà ;-).
- ideactiv est disponible sur Android, iPhone, et web (ideactiv.com). Il compte déjà ~10000 utilisateurs réguliers.
- ideactiv préfigure un « deep search engine », c'est-à-dire un moteur de recherche qui permet une recherche complexe (« je recherche une exposition de peinture en soirée vers le 15 novembre à Lunéville »), parce qu'il « comprend » les documents qu'il indexe (et les requêtes des utilisateurs).

# Quelques exemples de défi NLP. a. Interpréter une date.

Savoir interpréter des dates comme :

- Jusqu'au 30 avr 2023 mar 20h mer 20h
- 21 Mar 06 Jui 23 A 17h les mercredis 21 et 28 mar, 4 et 11 avr, 2, 16 et 23 mai, 6 jui
- Tous les jours de 10h a 19h sauf le dimanche et les jours feries, jeudi jusque 21h
- 29 décembre 2019 2 janvier 2020 Tous les jours a 14h30 et 16h Mardi, Mercredi, Jeudi, Vendredi, Samedi, Dimanche, Jours feries : 14h30 et 16h. Lundi : Ferme. Vacances scolaires de 10h a 12h30 et de 14h30 a 17h.
- Ferme ouverte tous les jours : en juin / juillet / aout de 8h a 21h, en septembre de 8h a 20h et en octobre de 8h a 19h. Derniere entree 18h. Mise a jour le 27/05/2020

## Quelques exemples de défi NLP. b. Interpréter une adresse.

Savoir interpréter des adresses comme :

- BAYEUX LE RADAR, Espace d'art actuel
- Les Etangs de Corot 15 rue de Versailles Paris
- Theatre de rue devant la salle polyvalente l'Escale route de Romille Geveze
- Salle Gerard Philipe Cour des associations Aile sud de la ferme 91800 Boussy-Saint-Antoine
- Le Cube RD 8 78000 Versailles
- Site de Brest 16, rue du Chateau, 29200 Brest Site du Relecq-Kerhuon 50, rue Eddy Chetler, 29480 Le Relecq-Kerhuon
- Regroupement de 3 lieux culturels sur Echirolles Centre du graphisme Musée de la Viscose Musée Géo-Charles 38130 Échirolles
- Dans le jardin du restaurant du musée de l'imprimerie 78000 Versailles

#### Quelques exemples de défi NLP. c. Déterminer si une activité est en ligne.

A partir du titre, du lieu, de la description d'une activité, déterminer si elle est en ligne.

Ce n'est pas toujours évident :

- Parfois un lieu est donné alors que l'activité est en ligne.
- Une activité de réalité virtuelle est bien réelle.
- Un concert peut proposer une réservation en ligne sans être en ligne.
- Un événement peut s'appeler « Bricolez de chez vous » ou « Le magicien vient chez vous » et l'interprétation de « chez vous » est différente.

### Quelques exemples de défi NLP. d. Attribuer des catégories à une activité.

A partir du titre, des mots clefs, de la description de l'activité, lui attribuer un certain nombre de catégories : **exposition**, **spectacle**, **théâtre**, **science**, **art**, **musique**, **jazz**, **famille**, etc.

L'algorithme actuel fonctionne correctement, mais il peut être enrichi.

Et il ne sait pas traiter certains cas difficiles : par exemple la description d'un spectacle pour enfant qui ne contient aucun mot de type « spectacle » ou « enfant », mais un texte comme « Pierrot aime secrètement Colombine et ensemble ils vont essayer de décrocher la lune. »

#### Quelques exemples de défi NLP. e. Regrouper correctement des adresses.



Voici un exemple réel d'adresses qui se ressemblent indexées par ideactiv.

Le système devrait pouvoir décider s'il y a un seul endroit au 16 rue Saint Denis, et quel est le meilleur nom à lui donner.

Il y a des centaines de situations comparables, toutes différentes et les décisions sont difficiles à formaliser en règles.

Cette étape est actuellement manuelle.

# Aller plus loin avec vous sur ces sujets

- ideactiv aimerait aider à monter un projet de recherche pour ajouter du ML à ses algorithmes actuels.
- ideactiv peut apporter à un projet : la connaissance accumulée des problèmes, des données en très grande quantité (1,5 millions d'activités, 1 million d'endroits à ce jour), des jeux de test, la performance de ses algorithmes actuels, un solide background mathématique, et du temps et de la disponibilité pour participer à l'élaboration des stratégies et à la mise au point des solutions.

Intéressé(e)? Parlons-en! Thomas Chenevier thomas@ideactiv.com