amadeus

NATURAL LANGUAGE PROCESSING TO PROCESS AIRLINES FARE CONDITIONS

Introduction

Comme décrit ci-contre, le moteur de recherche Amadeus est assez statique : une boucle qui va créer des millions de combinaisons et vérifier leur validité. Il y a deux manières d'invalider des combinaisons :

- pendant que l'on combine des objets (checks)
- avant de combiner des objets (pre-checks)

∀ combinaison de dates, ∀ combinaison de vols, ∀ combinaison de tarifs, ∀ contexte de tarification :
  if(check_is_KO): continue;
  else: saveValidCombination(); continue;

En analysant les filings des tarifs des compagnies aériennes, on y trouvera par exemple la condition « COMBINABLE END-ON-END DOMESTIC ». Si on se trouve dans une branche d'exploration avec de la combinatoire internationale, et si on est capable de comprendre ce filing, il devient inutile de continuer d'explorer cette branche puisque tout sera invalide, seules les combinaisons domestiques étant autorisées.

En introduisant du Natural Language Processing (NLP), on essaye donc de passer d'une approche d'exploration Generate&Check à une approche SmartGenerate&Check, ce qui permettra de réduire le nombre de checks tardifs et donc d'optimiser le temps de réponse du moteur de recommendations.

Règles de combinabilité

```
FQN4*10
CO.COMBINABILITY
   APPLICABLE ADD-ON CONSTRUCTION IS ADDRESSED IN
   MISCELLANEOUS PROVISIONS - CATEGORY 23.
  END-ON-END
    END-ON-END COMBINATIONS PERMITTED WITH AD  DOMESTIC
    FARES. VALIDATE ALL FARE COMPONENTS. SIDE TRIPS
    PERMITTED.
  OPEN JAWS/ROUND TRIPS/CIRCLE TRIPS
    FARES MAY BE COMBINED ON A HALF ROUND TRIP BASIS
    -TO FORM SINGLE OR DOUBLE OPEN JAWS/ROUND TRIPS/CIRCLE
    TRIPS.
   PROVIDED -
     COMBINATIONS ARE WITH ANY FARE FOR CARRIER AD WITHIN
     BRAZIL IN ANY RULE AND TARIFF.
```

Les fares contiennent des règles de combinabilité, ATPCO Fare Category ; et il y en a une trentaine !
La catégorie 4, par exemple, indiquera la validité d'un fare sur un ensemble de numéros de vols passant par certains points. Les catégories 6 et 7 spécifient le temps minimum/maximum qu'il est possible de rester à un certain point. La catégorie 10 définit les règles de combinabilité autorisées entre plusieurs fares…
Parmi les millions de combinaisons de fares et les milliards de combinaisons globales à considérer pour trouver la recommandation la moins chère sur un itinéraire donné, seul un faible pourcentage sera valide. Toutes les autres se verront invalidées par des checks de catégorie.

```
007 - MAXIMUM STAY

BETWEEN GOT AND PAR FOR KFLYAF  FARES

  TRAVEL FROM LAST STOPOVER MUST COMMENCE NO LATER THAN 1
  MONTH AFTER DEPARTURE FROM FARE ORIGIN
  OR - TRAVEL FROM LAST STOPOVER MUST COMMENCE NO LATER THAN
     3 MONTHS AFTER DEPARTURE FROM FARE ORIGIN.
       NOTE -
        EXTENSION UP TO 3 MONTHS PERMITTED AGAINST PAYMENT
        OF DKK/NOK/SEK 600 FOR THE INBOUND SECTOR.
```

```
007 - MAXIMUM STAY

BETWEEN MRS AND CZECH REPUBLIC FOR RAPCZFR  FARES

  TRAVEL FROM LAST INTERNATIONAL STOPOVER MUST BE COMPLETED
  NO LATER THAN 1 MONTH AFTER DEPARTURE OF THE FIRST
  INTERNATIONAL SECTOR.
```

```
007 - MAXIMUM STAY

FOR -PRITWE  TYPE FARES

  OUTBOUND -
    IF TRAVEL OCCURS SUN
      TRAVEL FROM LAST STOPOVER MUST BE COMPLETED THE SAME
      DAY AS DEPARTURE FROM FARE ORIGIN.

  OTHERWISE
    TRAVEL FROM LAST STOPOVER MUST COMMENCE NO LATER THAN 3
    DAYS AFTER DEPARTURE FROM FARE ORIGIN.
```

amadeus

Objectifs du TER

Le but de ce travail d'étude et de recherche est de traiter la partie « free text » des conditions tarifaires en utilisant du NLP. Il faudra comprendre les règles de combinabilité écrites par les compagnies aériennes avant de lancer des combinaisons d'objets, ce qui permettra de générer de manière intelligente les différentes combinaisons pour construire des recommandations. Ensuite, l'Etudiant devra essayer de créer des partitionnements (clusters) de conditions tarifaires : cela pourrait considérablement améliorer le temps de réponse du moteur de recommandations, ou du moins créer des opportunités de développement ultérieur.

Note : les ensembles de textes contenus dans les champs free text des filings ne sont pas annotés, et il sera difficile d'y appliquer de l'apprentissage supervisé. Une idée de généralisation consiste à représenter les concepts et les relations comme une ontology / un graphe de connaissances et essayer de trouver un lien entre les checks de catégorie et cette représentation. On pourra également utiliser du bootstrapping *(to learn lexico-syntactic patterns)* ou de l'active learning *(to improve the method and increase data coverage with semi-supervised methods)*.

Livrables du TER

Il est attendu des Etudiants un rapport écrit et un logiciel livrable.

Le rapport détaillé devra à minima décrire l'approche utilisée, en faisant mention des points bloquants et des solutions de contournement, une description technique de la solution logicielle mise en place avec des exemples, ainsi que des tests de performance prouvant l'efficacité, ou pas, de la solution.

Le logiciel livrable pourra être réalisé dans un ensemble de langages différents selon les besoins des Etudiants. Il devra accepter en entrée un ensemble de données prétraitées voire au format brut et fournir en sortie des données pouvant être utilisées par le moteur de recommandations actuel. Si nécessaire à la mesure de la performance, l'intégration par les Etudiants de la solution logicielle dans le moteur de recommandations est attendue, en C++ (simplement via une communication par fichiers textes ou de manière native).