

Rapport de Projet : Traitement Automatique du Texte en IA

Sujet :

Identification des opinions exprimées dans les avis et
commentaires d'un ordinateur.

Yann MARTIN D'ESCRIGNE
Yohann TOGNETTI

Table des matières

1	Introduction	2
1.1	Présentation générale du projet	2
1.2	Choix du sujet	2
1.3	Description des tâches	2

1 Introduction

1.1 Présentation générale du projet

Dans le cadre de notre cours de Traitement automatique du texte en IA, il nous a été demandé d'effectuer un projet de notre choix impliquant une intelligence artificielle travaillant sur des données textuelles. La restriction sur le sujet se limitant à l'existence d'un jeu de données conséquent.

1.2 Choix du sujet

Notre choix de sujet fut **l'identification des opinions exprimées dans les avis et commentaires d'un ordinateur**. Plus précisément, comment retrouver parmi un ensemble de commentaires les parties de l'ordinateur visées, sur quels aspects et quel en est l'avis général qui en ressort.

1.3 Description des tâches

Ce sujet est fortement inspiré du "SemEval-2015 Task 12" sur la partie "Aspect Based Sentiment Analysis (ABSA) : Laptop Reviews" et partage donc le même objectifs avec des simplifications. Les différentes entités et catégories qui vont suivre proviennent également des annotations du sujet de SemEval. Chacun des commentaires est fragmentés phrase par phrase afin de faciliter le travail de l'IA.

Tout d'abord il s'agit de retrouver dans la phrase le(s) entité(s) ciblée. Elles peuvent être l'ordinateur comme un tout, ses parties physique (clavier, écran..) , des logiciels ou OS (Windows, navigateur, jeux...) ou bien même une compagnie et ses services. (DELL, Apple, le support technique, la livraison...). Rien n'interdit d'avoir plusieurs entités dans la même phrase.

Voici la liste des entités possible :

- DISPLAY (=moniteur, écran),
- CPU (=processeur),

- MOTHERBOARD (=carte mère),
- HARDDISC (=disque dur),
- MEMORY (=mémoire, RAM),
- BATTERY (=batterie),
- POWER_SUPPLY (=chargeur, unité de chargement, cordon d'alimentation, (power) adaptateur),
- KEYBOARD (=touche, clavier, pavé numérique),
- MOUSE (=souris, pavé tactile)
- FANS_COOLING (=ventilateur, système de refroidissement),
- OPTICAL_DRIVES (=lecteur CD, DVD ou Blue-ray),
- PORTS (=USB, HDMI, VGA, lecteur de carte),
- GRAPHICS (=carte graphique, carte video),
- MULTIMEDIA_DEVICES (=son, audio, microphone, webcam, haut-parleur, casque, écouteurs).

Une fois obtenue il faut également trouver sur quelle(s) catégorie(s) de l'entité le commentaire porte. Cela peut être un aspect général , sa prise en main, ses performances, son design et ses fonctionnalités, etc... Là encore, rien n'interdit d'aborder plusieurs catégorie pour la même entité au sein d'une même phrase.

Voici la liste des catégories possible :

- GENERAL,
- PRICE (=prix),
- QUALITY (=qualité),
- DESIGN_FEATURES (=design et fonctionnalités),
- OPERATION_PERFORMANCE,
- USABILITY (=ergonomie, prise en main),
- PORTABILITY (=portabilité),
- CONNECTIVITY (=connectivité),
- MISCELLANEOUS (=divers).

L'entité E et la catégorie C qui s'y rapporte forme ainsi un couple E#C. Il est à noter la possibilité qu'aucun couple E#C se rapporte à une phrase d'un commentaire.