

Rapport TAL

Cédrick RIBET - Hugo BOMMART - Manfred MADELAINE

3 mai 2017

Sommaire

I	Projet TAL	2
1	A propos	3
1.1	Consignes	3
1.2	Choix du Sujet	4
1.3	un titre plus court	4
1.3.1	Sous-Section	4

Première partie

Projet TAL

Chapitre 1

A propos

1.1 Consignes

Rapport : Un rapport de 5 à 7 pages qui décrit votre travail (police et marges de taille raisonnable). Vous devez y décrire l'objectif de votre projet, la motivation derrière ce que vous avez fait, la méthode que vous avez utilisé, à quel point cela fonctionne, et les limitations de votre programme (c'est-à-dire les cas qui restent à traiter ou qui fonctionnent mal).

Vous devez également inclure une petite section qui décrit de façon brève la contribution de chaque membre du groupe. Le rapport ne doit pas inclure de code source.

Mettez en avant ce que votre programme gère bien. Les limitations de votre programme sont importantes à mentionner, mais il est aussi important de les présenter comme des points d'amélioration plutôt que comme des failles. Si vous savez comment vous auriez souhaité résoudre ces limitations si vous aviez eu plus de temps, mentionnez-le comme une des choses prévues pour l'avenir.

Sujet :

Début du projet : Le 27 Mars 2017

Dead line : **Le 6 Mai 2017 à 23h59**, avec une pénalité de -5pts par jour de retard.

Git : Un dépôt GIT a été utilisé tout au long de ce projet. Voici le lien du dépôt :

<https://github.com/hkemmel/tal>

1.2 Choix du Sujet

E-opinion - classification de reviews : L'e-opinion (analyse de sentiments) est un domaine très important pour les entreprises car les retours clients permettent de s'adapter aux marchés ou d'identifier des points d'amélioration. Un moyen simple de s'attaquer au problème est de classer les reviews comme étant positives ou négatives (voire sur une échelle plus fine). Vous pouvez également essayer d'identifier automatiquement certains aspects des reviews qui peuvent être pertinents pour une entreprise, par exemple en identifiant toutes les phrases négatives ou positives.

Enjeux :

- Vous aurez probablement un certain nombre de mots qui relèvent du domaine spécifique que vous aurez choisi. Vous pourrez donc constituer un petit lexique contenant des mots spécialisés du domaine, par exemple des aspects du produit qui est en vente.
- Détecter les phrases positives et négatives n'est pas trivial. Par exemple, l'existence d'un mot négatif dans une phrase ne signifie pas forcément que la phrase est négative. Par exemple : Ce n'est pas du tout un mauvais portable.

Vous pouvez également rencontrer des comparaisons avec d'autres produits : Le portable de la marque concurrente est bien meilleur que ce portable.

Donc une approche "sac de mots" ne sera pas forcément la méthode la plus adaptée. Mais vous pouvez toujours tester cette méthode pour l'utiliser comme point de comparaison avec d'autres méthodes plus complexes que vous essayerez.

A faire :

- Choisissez un domaine sur lequel vous voulez appliquer votre analyseur de sentiments et définissez L'OBJECTIF DE VOTRE ÉTUDE. Imaginez-vous dans la position d'une entreprise qui fournit une étude de marché pour un client. Votre tâche est de vendre votre analyseur de sentiments et de leur montrer comment cela marche et à quel point ça marche bien!
- Trouvez le corpus sur lequel vous allez travailler. Votre corpus sera constitué d'un ensemble de documents (chaque document étant une critique de produit) contenant du texte en langage naturel.
- Vous utiliserez la méthode que vous voulez pour effectuer cette tâche (probabilités, expressions régulières, grammaires CFG (CFGs avec NLTK), règles écrites à la main, etc.)

Motivations : Ce sujet se rapprochait le plus des TPs que nous avons fait en cours et des affinités en TAL de chacun. En effet,

1.3 section

1.3.1 Sous-Section

Début c'est génial!

Suite azaa

ou

nouveau paragraphe

* Un lapin.

1. un canard

2. une pintade

Un canard : bestiole qui fait coin.

Un ours : bestiole qui fait mal.

un canard ¹

un ornithorynque ¹⁸

un ours ³²

1. bestiole qui fait coin
18. bestiole qui fait rire
32. bestiole qui fait mal