

Contextualisation de messages courts l'importance des métadonnées

Cossu, JV; Gaillard J; Torres-Moreno, JM; El-Beze, M

LIA, EPM

Projet ImagiWeb

Analyse de la dynamique des images sur le web

- Catégorisation automatique des messages
- Résumer et visualiser une image

Méthodes probabilistes et naïves sans traitement TAL

Vodkaster

Réseau social du cinéma

- 80 000 MC (140 caractères) depuis Juin 2010
- MC notées entre 0,5 et 5
- 1 800 utilisateurs
- 7 800 films

Répartitions inégales des MC

-> Proposer de nouvelles fonctionnalités

Moteur de recommandation

Problèmes du moteur de recommandation

- Recommandations inappropriées
- « **cold start** » pour les utilisateurs/produits peu représentés
- Difficultés pour justifier une recommandation

Solution -> exploiter la MC

Extraction d'information

On connaît pour chaque MC

- la note attribuée
- son auteur
- le film concerné

-> Extraction probabiliste des termes discriminants

Pour chaque MC

- Prise en compte des métadonnées

-> Amélioration de la prédiction de note

Perspectives et applications

Segmentation de la population

Segmentation des critiques dès l'apprentissage

Détection des points positifs/négatifs

Séparation entre neutres et nuancées