



STEAMER

Charles-Lewis Jaggi
Fabien Mottier
Dylan Netuschil



Sommaire

- Introduction
- Démonstration
- Technologies
- Données
- Algorithme
- Optimisations
- Problèmes rencontrés
- Conclusion



Introduction

- Suggestion de jeux vidéo
- Plateforme steam
- En fonction de plusieurs critères
 - Review
 - Prix
 - Langue
 - Genre
 - Tag
 - Type de jeu
- Les jeux de ses amis



Démonstration





Technologies

- mongoDB (noSQL)
- Python
- Flask
- numpy
- Sklearn
- API REST



mongoDB®



Flask

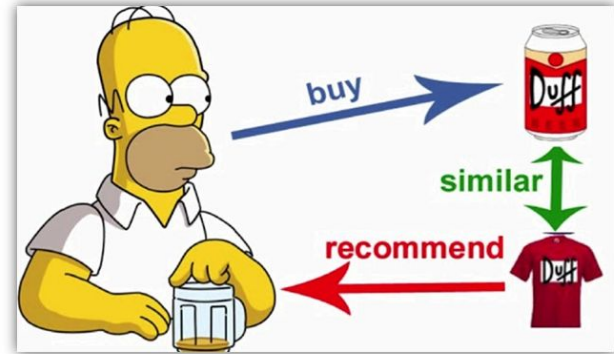
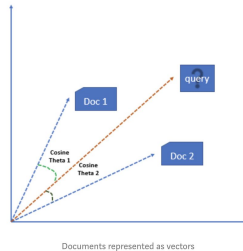


Données

- Kaggle [csv]
 - Nettoyage des données
 - Pré-processing
 - Import dans une base de données MongoDB
- Steam API
 - Jeux du joueur
 - Jeux des amis du joueur
 - Ajout à la base de donnée si pas déjà présent

Algorithme

- Recommendation Objet (Content-based)
 - Base de données
 - Cosinus de similarité
 - TF-IDF
 - Ranking des caractéristiques
 - Reviews positives
 - Caractéristiques communes





Algorithme

- Recommandation Personnalisée
 - Steam Web API
 - Bibliothèque de l'utilisateur
 - Jeu moyen représentatif
- Recommandation Sociale
 - Steam Web API
 - Bibliothèque des amis
 - Coefficient de Pearson r
 - Possession de jeu

Movie ratings

	Amy	Jef	Mike	Chris	Ken
The Piano	-	-	+		+
Pulp Fiction	-	+	+	-	+
Clueless	+		-	+	-
Cliffhanger	-	-	+	-	+
Fargo	-	+	+	-	?



Optimisations

- Algorithme
 - Pré-processing
 - Limitation des caractéristiques 440 -> 57
- Filtres
 - Langues
 - Prix
 - Niveau de review
 - Posséder une caractéristique commune



Optimisations

- Multithreading
 - Python : Package multiprocessing
 - Génération de plusieurs processus python en parallèle
- Récupération complet pour les grands jeux de donnée
 - Accès base de données
 - Utilisation de `as_pymongo()`



Problèmes rencontrés

- Lenteur de l'algorithme
- Temps pour accéder à l'API
- Le nombre d'accès à la base de données
- Difficultés d'évaluation



Conclusion

- Projet très intéressant
 - Pleins de concepts
 - Algorithme en temps réel
 - Compromis
- Améliorations possibles
 - Fournir plus de retours aux utilisateurs
 - Suivi du comportement utilisateur
 - Utiliser plus de données



Questions

