

# PROJET DE RECHERCHE

Moteur de recommandation de séries par analyse de sous-titres

A C H R A F J D A Y N ° 3 8 0 2 4 1 0



### **SOMMAIRE**

- □ Nature des données
- ☐ Traitement des données
- ☐ Approche basée sur le contenu
- □ Approche basée sur le filtrage collaboratif
- **□** Difficultés
- □ Approche hybride



### NATURE DES DONNÉES



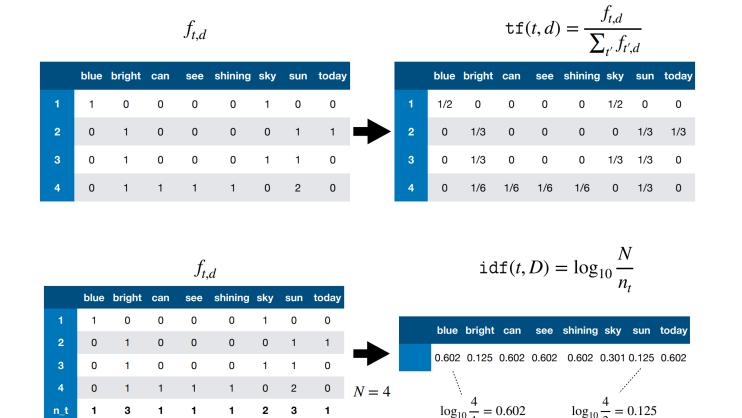
**Sous-titres** 

- Tokenisation
- Normalisation
- ☐ Lemmatisation
- □ Stemming
- ☐ Stop-words



- Nom utilisateur
- ☐ Nom de la série
- Note personnelle
- Note globale
- ☐ Genre

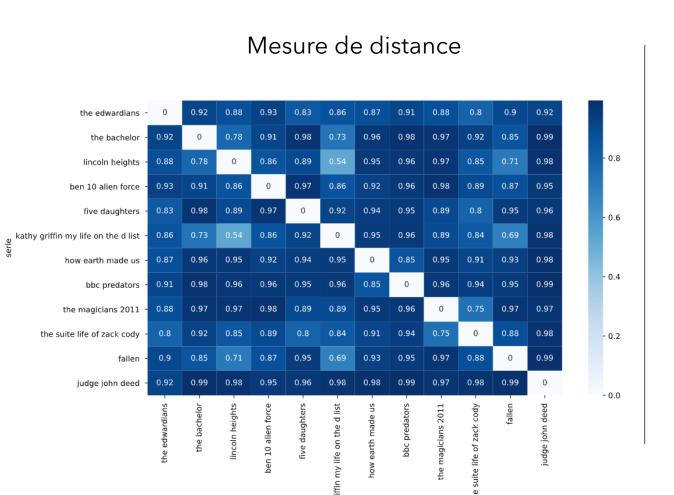
### TRAITEMENT DES DONNÉES



 $\texttt{tfidf}(t, d, D) = \texttt{tf}(t, d) \cdot \texttt{idf}(t, D)$ 

	blue	bright	can	see	shining	sky	sun	today
1	0.301	0	0	0	0	0.151	0	0
2	0	0.0417	0	0	0	0	0.0417	0.201
3	0	0.0417	0	0	0	0.100	0.0417	0
4	0	0.0209	0.100	0.100	0.100	0	0.0417	0

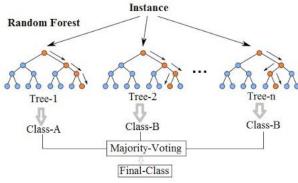
### **APPROCHE BASÉE LE CONTENU**



#### Prédiction des genres

- ☐ Fusion de base données
- ☐ TF-IDF
- ☐ Le classifieur Random Forest :
  - Bootstraping
  - Création
  - Aggregation
  - Optimisation

#### Random Forest Simplified



## APPROCHE BASÉE SUR LE FILTRAGE COLLABORATIF













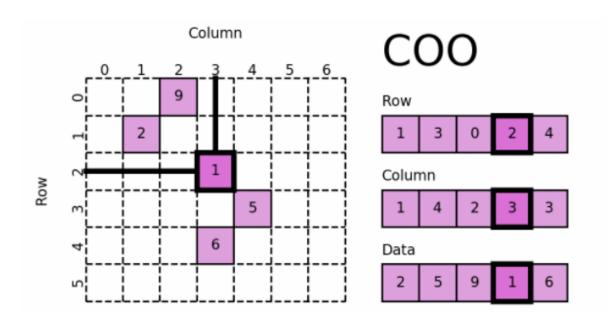


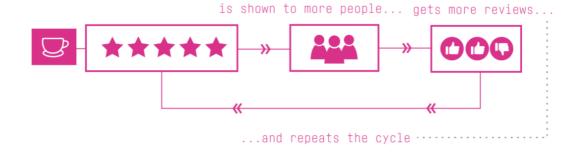
2	4	2	4	5	3
5	4	4	ന	5	1
3	4	5	5	2	3
4	1	2	5	3	4
4	3	4	3	5	2
4	5	5	1	2	3

### **DIFFICULTÉS**

Sparsité

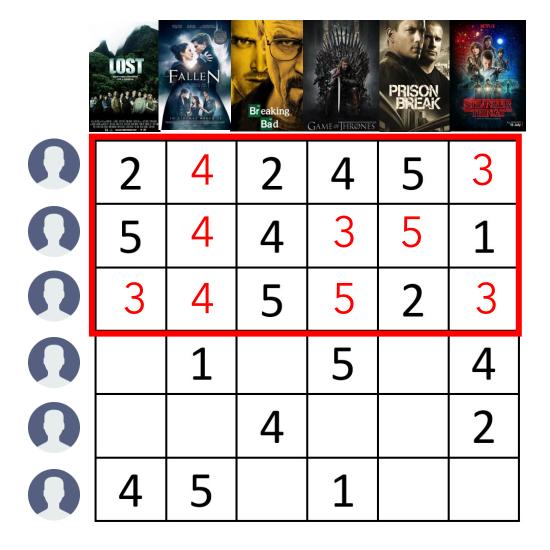
Démarrage à froid





### **APPROCHE HYBRIDE**

the edwardians -	0	0.92	0.88	0.93	0.83	0.86	0.87	0.91	0.88	0.8	0.9	0.92
the bachelor	0.92	0	0.78	0.91	0.98	0.73	0.96	0.98	0.97	0.92	0.85	0.99
lincoln heights	0.88	0.78	0	0.86	0.89	0.54	0.95	0.96	0.97	0.85	0.71	0.98
ben 10 alien force	0.93	0.91	0.86	0	0.97	0.86	0.92	0.96	0.98	0.89	0.87	0.95
five daughters -	0.83	0.98	0.89	0.97	0	0.92	0.94	0.95	0.89	0.8	0.95	0.96
မှု kathy griffin my life on the d list	0.86	0.73	0.54	0.86	0.92	0	0.95	0.96	0.89	0.84	0.69	0.98
how earth made us	0.87	0.96	0.95	0.92	0.94	0.95	0	0.85	0.95	0.91	0.93	0.98
bbc predators -	0.91	0.98	0.96	0.96	0.95	0.96	0.85	0	0.96	0.94	0.95	0.99
the magicians 2011 -	0.88	0.97	0.97	0.98	0.89	0.89	0.95	0.96	0	0.75	0.97	0.97
the suite life of zack cody	0.8	0.92	0.85	0.89	0.8	0.84	0.91	0.94	0.75	0	0.88	0.98
fallen -	0.9	0.85	0.71	0.87	0.95	0.69	0.93	0.95	0.97	0.88	0	0.99
judge john deed  -	0.92	0.99	0.98	0.95	0.96	0.98	0.98	0.99	0.97	0.98	0.99	0
	the edwardians –	the bachelor –	lincoln heights -	ben 10 alien force –	five daughters –	kathy griffin my life on the d list –	how earth made us -	bbc predators –	the magicians 2011 -	the suite life of zack cody –	fallen –	judge john deed -



### **CONCLUSION**

