

Filtrage : Bayésien (1)

- Utilisation récente
- Travaux de Paul Graham (Août 2002)
 - Travaux de Robinson
 - Travaux de Robinson sur Fisher
- Principe d'apprentissage
- Classification plus que détection de SPAMs
- De très nombreux outils se sont créés

Filtrage : Bayésien (2)

- Base contenant
 - Terme ou token (mot, bi-mot, tri-mot, voire phrase)
 - Nombre d'apparitions dans chaque classe (SPAM/HAM)
 - Calcul de la probabilité d'appartenir à chaque classe
- Problèmes :
 - Apprentissage long
 - Bien choisir son apprentissage (penser à la poubelle !)
 - Technique émergente
 - Penser à nettoyer les bases

Filtrage : Bayésien (3)

- Avantages
 - Rapide (5 à 50 fois plus rapide)
 - Très efficace
 - Résiste à des erreurs d'apprentissage
 - Rarement de faux positifs
- Exemples
 - Site : Bogofilter, SpamProbe, CRM114
 - Individu : popfile, MacOS X, Mozilla