

## Filtre de spam

POURMOHAMMAD Maryam, CHABANE Djamila, HAJEB Ghamdan  
Département d'Informatique

### C'est quoi le spam ? Quel est son impact ?

- Le **spam** également connu sous le nom de **POURRIEL** en français
- Consiste en l'**envoi en masse** de courriers électroniques **non sollicités**
- **249 milliards** de courriers sont échangés dans le monde chaque jour, dont **238 milliards** étant du spam, soit plus de **81%**
- Généralement **3 à 5 minutes** sont consacrées pour éliminer le spam chaque jour
- Le spam coûte **300€ par employé et par année!**



### Se protéger ? Éviter le spam ? L'Anti-spam !

- Respecter un ensemble de règles ...
- Utiliser des technologies anti-spam ;
  - **Analyse de contenu**
  - Répertorier les adresses e-mails par **catégories**
  - Filtrages des courriers : **Bayésien, par mots clés, ...**



Des taux de réussite allant jusqu'à **95%** !

### Filtrage Bayésien du spam

- Très répandu et efficaces
- Principe de fonctionnement :
  1. L'utilisateur prédéfinit les premiers spams
  2. Analyse du contenu des e-mails entrant
  3. Stocke les mots pertinents
  4. Détermine par probabilité, si c'est du spam ou pas

$$P_s = \frac{p(\text{foo} | \text{spam}) \times p(\text{spam})}{p(\text{foo} | \text{spam}) \times p(\text{spam}) + p(\text{foo} | \text{ham}) \times p(\text{ham})}$$

