

Sommaire

Table des matières

Préface

Introduction : contexte, enjeux et positionnement de la recherche d'information

Chapitre 2 Représentation et indexation

2.1 Prétraitements linguistiques

2.1.1 Segmentation

2.1.2 Normalisation

2.1.3 Filtrage par un anti-dictionnaire

2.2 Les deux lois de base en recherche d'information

2.2.1 Loi de Heaps

2.2.2 Loi de Zipf

2.3 Représentation documentaire

2.3.1 Modèle vectoriel

– Pondération des termes

– Index inversé

– Indexation dans des collections statiques et dynamiques

Chapitre 3 Recherche d'information

3.1 Modèles de recherche

3.1.1 Modèles booléens

3.1.2 Modèles vectoriels

3.1.3 Modèles probabilistes

3.1.4 Approche axiomatique de la RI

3.2 Expansion de requêtes

3.3 Mesures d'évaluation

– résultats ordonnés / non ordonnés

– précision, rappel, F-mesure, etc.

– courbe ROC et AUC

Chapitre 4 Recherche sur le Web / architectures Web

Structure du Web, HTML, protocoles

Robots d'indexation

Index distribués

Algorithmes liés aux liens

Chapitre 5 Classification / catégorisation de documents

Formalisme

Sélection de variables

Modèles génératifs (Bernoulli, multinomial)

Modèles discriminants

Modèle logistique, SVM (séparateurs à vaste marge)

Mesures d'évaluation pour la classification

Chapitre 6 Partitionnement / clustering (

Algorithmes principaux (à plat, hiérarchique)

Validation des classes

Applications à la recherche d'information

Chapitre 7 Modèles à thèmes latents / modèles probabilistes avancés

Analyse sémantique latente

Modèle LDA (Latent Dirichlet Allocation)

Extensions probabilistes

Chapitre 8 Outils, logiciels et passage à l'échelle / Big Data

Présentation de quelques frameworks / moteurs open source

Problèmes de passage à grande échelle

Optimisations et défis du Big Data

Conclusion

Bibliographie

Annexes / Exercices corrigés

Index