

Table des matières

Avant-propos	15
Mohand BOUGHANEM et Jacques SAVOY	
Chapitre 1. Introduction à la recherche d'information	19
Mohand BOUGHANEM	
1.1. Introduction.	19
1.2. Concepts de base de la recherche d'information	19
1.2.1. Du document à la base documentaire	20
1.2.2. Du besoin en information à la requête.	20
1.2.3. Pertinence	21
1.3. Architecture d'un système de recherche d'information	21
1.4. Phase d'indexation	24
1.4.1. L'analyse lexicale.	24
1.4.2. L'élimination des mots vides	25
1.4.3. La lemmatisation	25
1.5. La pondération des termes	26
1.5.1. Loi de Zipf.	26
1.5.2. Conjecture de Luhn.	26
1.5.3. Pondération $tf * idf$	27
1.6. Les modèles-piliers de la recherche d'information.	28
1.6.1. Le modèle booléen	29
1.6.2. Le modèle vectoriel.	30
1.7. Le modèle probabiliste	32
1.7.1. Les modèles de langue pour la recherche d'information	35
1.8. Evaluation des systèmes de recherche d'information	36

1.8.1. Evaluation de la performance d'un système de recherche d'information	36
1.8.2. Les mesures à x documents et la précision moyenne	39
1.8.3. Collections de référence, un exemple : TREC	40
1.9. Conclusion	41
1.10. Bibliographie	42

Chapitre 2. Dépister de l'information sur le web. 45

Jacques SAVOY

2.1. Introduction.	45
2.2. Rechercher des pages	47
2.2.1. Fonctionnement	47
2.2.2. Evaluation	51
2.3. Utiliser les hyperliens	52
2.3.1. L'activation propagée	54
2.3.2. L'algorithme de Kleinberg	55
2.3.3. L'algorithme PageRank	58
2.4. Les internautes et leurs requêtes.	60
2.4.1. Analyse des interactions entre internautes et moteurs.	61
2.4.2. Les requêtes les plus fréquentes	63
2.5. Les résultats de Google sont-ils biaisés ?	64
2.6. Répondre à d'autres types de requête.	67
2.7. Conclusion	69
2.8. Bibliographie.	70

Chapitre 3. L'importance des probabilités *a priori* pour la recherche des pages d'entrée 75

Wessel KRAAIJ, Thijs WESTERVELD et Djoerd HIEMSTRA

3.1. Introduction.	75
3.2. Contexte et ouvrages reliés.	76
3.2.1. Optimisation des systèmes de recherche d'information	76
3.2.2. Caractéristiques non reliées au contenu des pages web.	77
3.2.3. Probabilités <i>a priori</i> et modèles de langue en recherche d'information	78
3.3. Modèles de langue et probabilités <i>a priori</i>	79
3.3.1. Estimation des probabilités <i>a priori</i>	79
3.3.2. Combiner des probabilités <i>a priori</i>	81

3.3.3. La combinaison des modèles de langue	83
3.4. Les probabilités <i>a priori</i> pour la recherche de pages d'entrée	83
3.4.1. Longueur du document.	84
3.4.2. Nombre des liens entrants	85
3.4.3. Forme de l'URL.	86
3.5. Evaluation.	88
3.6. Expérimentations	89
3.6.1. Combinaison des modèles avec les textes d'ancrage	90
3.6.2. Combinaison des probabilités <i>a priori</i>	91
3.7. Discussion	92
3.8. Conclusion	95
3.9. Bibliographie.	96
Chapitre 4. XML et recherche d'information.	99
Karen PINEL-SAUVAGNAT et Claude CHRISMENT	
4.1. Introduction.	99
4.1.1. Historique des langages de balisage et des approches de recherche d'information liées.	100
4.1.2. Documents semi-structurés et XML	101
4.1.3. Recherche d'information structurée : problèmes et enjeux.	104
4.2. Techniques d'indexation des documents semi-structurés	108
4.2.1. Indexation de l'information textuelle	109
4.2.2. Indexation de l'information structurelle.	111
4.3. Langages de requêtes	114
4.3.1. Langages de requêtes : premières approches	115
4.3.2. Evolution des langages de requêtes : prise en compte de la composante recherche d'information	116
4.3.3. Conclusion sur les langages de requêtes XML.	118
4.4. Appariement requêtes/granules documentaires	118
4.4.1. Approches par propagation des termes des documents	119
4.4.2. Approches par propagation des scores des éléments	123
4.4.3. Remarques concernant le traitement de la structure.	127
4.5. Evaluation : la campagne d'évaluation INEX.	128
4.5.1. Collection	128
4.5.2. Requêtes	128
4.5.3. Tâches	129
4.5.4. Jugements de pertinence	130
4.5.5. Evaluation	130
4.6. Conclusion	131
4.7. Bibliographie.	132

Chapitre 5. Recherche d'information multilingue	139
Jian-Yun NIE et Jacques SAVOY	
5.1. Introduction	139
5.2. Rechercher dans d'autres langues que l'anglais	141
5.2.1. Les langues européennes	141
5.2.2. Les langues d'Extrême-Orient	147
5.2.3. Principaux résultats	149
5.3. La recherche bilingue	151
5.3.1. Utiliser un système de traduction automatique	152
5.3.2. Recourir à un dictionnaire bilingue ou multilingue	152
5.3.3. Recourir à un corpus parallèle	154
5.4. Recherche d'information multilingue	159
5.4.1. Stratégies pour la traduction automatique	159
5.4.2. Stratégies de fusion des résultats	161
5.5. Vers une intégration dans un modèle unifié	164
5.6. Conclusion	166
5.7. Bibliographie	167
Chapitre 6. Les systèmes de question-réponse et le web	171
Brigitte GRAU	
6.1. Introduction	171
6.2. Les systèmes de question-réponse en domaine ouvert	173
6.2.1. Relations entre question et réponse	173
6.2.2. Composants d'un système de question-réponse	175
6.3. Le web comme source de connaissances pour trouver des réponses	180
6.3.1. Interrogation du web	182
6.3.2. Utilisation de la redondance	185
6.3.3. Le web structuré	186
6.4. Le web multilingue	188
6.4.1. Traduction des questions	188
6.4.2. Traduction des termes de la question	189
6.5. Le web comme corpus	191
6.5.1. Approches statistiques	191
6.5.2. Validation de réponses	192
6.6. Conclusion	193
6.7. Bibliographie	195

Chapitre 7. Recherche d'information contextuelle et web	201
Lynda TAMINE-LECHANI et Sylvie CALABRETTO	
7.1. Introduction.	201
7.2. De la recherche d'information orientée système à la recherche d'information orientée utilisateur.	202
7.2.1. La recherche d'information adaptative : quel bilan ?	203
7.2.2. Le web et la recherche d'information contextuelle	207
7.2.3. Verrous technologiques	208
7.3. Notions et vocabulaire de base	209
7.3.1. Contexte	209
7.3.2. Profil	209
7.3.3. RI contextuelle ou RI personnalisée ?	210
7.3.4. Pertinence contextuelle.	210
7.4. Les modèles de recherche d'information contextuelle.	211
7.4.1. Déclinaison du contexte dans les systèmes de recherche d'information	211
7.4.2. Etat de l'art des modèles de recherche d'information contextuelle	211
7.4.3. Extension du modèle probabiliste	212
7.4.4. Extension du modèle vectoriel	215
7.5. Les systèmes d'accès contextuel à l'information.	216
7.5.1. Expression explicite du contexte	217
7.5.2. Génération automatique des informations du contexte	217
7.5.3. Deviner les besoins de l'utilisateur.	218
7.6. Evaluation des modèles et techniques de recherche d'information contextuelle	218
7.6.1. Le modèle d'évaluation Cranfield : apports et limites	219
7.6.2. Vers un modèle d'évaluation adapté à la recherche d'information contextuelle ?	219
7.7. Conclusion	223
7.8. Bibliographie.	224
Chapitre 8. Le web sémantique, quel renouvellement pour la recherche d'information ?	231
Nathalie AUSSENAC-GILLES	
8.1. Introduction.	231
8.2. Que recouvre le web sémantique aujourd'hui ?	232
8.2.1. Le projet du web sémantique	232
8.2.2. La mise en œuvre du web sémantique	234

8.2.3. Les réalisations technologiques du web sémantique	235
8.2.4. La recherche d'information dans le cadre du web sémantique	237
8.3. Les ontologies	238
8.3.1. Héritage pluridisciplinaire des ontologies.	238
8.3.2. Ressources utilisées en recherche d'information.	240
8.3.3. Ontologies en ingénierie des connaissances : définitions.	240
8.3.4. Composante lexicale et éléments linguistiques dans les ontologies	242
8.3.5. Construire et réutiliser des ontologies	243
8.4. Utilisation des ontologies en recherche d'information	245
8.4.1. Evolution des enjeux de la RI.	245
8.4.2. Apports attendus des ontologies dans la recherche d'information	246
8.5. Ontologies, indexation et annotation sémantique.	247
8.5.1. Ontologies pour décrire le contenu de documents.	247
8.5.2. Indexation conceptuelle <i>versus</i> annotation	248
8.5.3. Ontologie pour l'indexation sémantique en amont de moteurs généralistes	250
8.5.4. Extraction d'information pour l'annotation sémantique	252
8.6. Ontologie pour formuler des requêtes, sélectionner et visualiser les documents résultats	253
8.6.1. Formulation de requêtes à l'aide de concepts	253
8.6.2. De la requête aux résultats dans le cas d'une indexation sémantique	255
8.6.3. Les raisonnements permis par les ontologies au cours de la RI	257
8.7. Quel web sémantique pour une recherche d'information efficace ?	259
8.7.1. Adéquation des ressources sémantiques à la recherche d'information	259
8.7.2. Annotation sémantique et maintenance d'ontologie.	259
8.7.3. Complémentarité du web sémantique et du web 2.0	260
8.7.4. Ontologies ou thésaurus structurés pour la recherche d'information ?	261
8.8. Bibliographie.	261

Chapitre 9. Recherche de documents images, audio et vidéo sur le web

Georges QUÉNOT et Philippe MULHEM

9.1. Introduction.	267
9.2. Recherche d'images fixes sur le web.	268
9.2.1. Introduction	268

9.2.2. Utilisation du contenu textuel des pages web	268
9.2.3. Utilisation du contenu visuel	271
9.2.4. Combinaison des aspects textuels et visuels	272
9.2.5. Utilisation des liens des images.	276
9.2.6. Conclusion	279
9.3. Recherche de documents audiovisuels sur le web	279
9.3.1. Introduction	279
9.3.2. Utilisation du contenu textuel des pages web	281
9.3.3. Utilisation de la transcription de la parole	281
9.3.4. Utilisation du contenu visuel	284
9.3.5. Système et interfaces utilisateur	290
9.3.6. Conclusion	291
9.4. Bibliographie	292
Chapitre 10. Métaphores de visualisation des résultats de recherche d'information sur le web	295
Nicolas BONNEL, Max CHEVALIER et Bernard DOUSSET	
10.1. Caractéristiques des interfaces de visualisation en RI	297
10.1.1. Aspects cognitifs	297
10.1.2. Dimensionnalité des métaphores de visualisation : 1D (texte), 2D ou 3D ?	298
10.1.3. Usage de la couleur	303
10.2. Les interfaces de visualisation des résultats de recherche sur le web.	303
10.2.1. Classification des interfaces de visualisation	304
10.2.2. Les interfaces adaptées à la modalité navigation.	305
10.2.3. Les interfaces adaptées à la modalité interrogation	309
10.3. Evaluation des interfaces de visualisation en RI	328
10.3.1. Méthodes analytiques.	329
10.3.2. Méthodes empiriques	330
10.3.3. Discussion	332
10.4. Conclusion	332
10.5. Bibliographie	333
Index	341