

# 目 次

訳者前書き	i
序文	iii
<b>第 1 章</b> ウェブ探索エンジンについて	<b>1</b>
1.1 情報検索の手短な歴史	1
1.2 伝統的な情報検索のまとめ	7
1.2.1 検索エンジンのブール代数モデル	7
1.2.2 検索エンジンのベクトル空間モデル	8
1.2.3 検索エンジンの確率モデル	10
1.2.4 メタ検索エンジン	10
1.2.5 検索エンジンの比較	11
1.3 ウェブ情報検索	12
1.3.1 ウェブ検索での課題	12
1.3.2 ウェブ検索プロセスの構成要素	15
<b>第 2 章</b> クローラー、インデックス付け、およびクエリーの処理	<b>19</b>
2.1 クローリング	19
2.2 内容インデックス	24
2.3 クエリー処理	27
<b>第 3 章</b> 人気度によってウェブページをランク付けする	<b>31</b>
3.1 1998 年の場面	32
3.2 2つの主題	33

3.2.1	PageRank .....	34
3.2.2	HITS .....	36
3.3	クエリー独立性 .....	38
<b>第 4 章</b>	<b>Google の PageRank の数学</b>	<b>39</b>
4.1	PageRank のもとの総和公式 .....	40
4.2	総和方程式の行列表現 .....	41
4.3	繰り返しプロセスでの問題 .....	44
4.4	マルコフ連鎖理論のさわり .....	45
4.5	基本モデルに対する初期の調整 .....	46
4.6	PageRank ベクトルの計算 .....	50
4.7	Google 行列のスペクトラムの定理と証明 .....	59
<b>第 5 章</b>	<b>PageRank モデルのパラメータ</b>	<b>61</b>
5.1	$\alpha$ 成分 .....	61
5.2	ハイパーリンク行列 $\mathbf{H}$ .....	63
5.3	テレポーテーション行列 $\mathbf{E}$ .....	65
<b>第 6 章</b>	<b>PageRank の感度</b>	<b>75</b>
6.1	$\alpha$ に関する感度 .....	75
6.2	$\mathbf{H}$ に関する感度 .....	82
6.3	$\mathbf{v}^T$ に対する感度 .....	83
6.4	感度に対する他の分析 .....	83
6.5	感度定理とその証明 .....	87
<b>第 7 章</b>	<b>線形システムとしての PageRank 問題</b>	<b>93</b>
7.1	$(\mathbf{I} - \alpha\mathbf{S})$ の性質 .....	93
7.2	$(\mathbf{I} - \alpha\mathbf{H})$ の性質 .....	94
7.3	PageRank の疎な線形システムの証明 .....	97
<b>第 8 章</b>	<b>PageRank の大規模実装における問題点</b>	<b>99</b>
8.1	ストレージの問題 .....	100
8.2	収束基準 .....	104
8.3	正確さ .....	105
8.4	ぶら下がりノード .....	106
8.5	戻りボタンのモデル化 .....	111

<b>第 9 章 PageRank の計算の高速化</b>	<b>117</b>
9.1 適応型ベキ乗法	118
9.2 外挿	119
9.3 凝集	124
9.4 他の数値手法	127
<b>第 10 章 PageRank ベクトルの更新</b>	<b>129</b>
10.1 2つの更新問題とその歴史	130
10.2 ベキ乗法を再スタートする	132
10.3 近似凝集を用いる近似更新	134
10.4 正確な凝集	136
10.5 正確凝集 vs. 近似凝集	138
10.6 繰返し凝集による更新	141
10.7 分解を決定する	143
10.7.1 収束の速さによる分割	144
10.7.2 スケールフリーネットワークと Google の PageRank	145
10.8 結論	146
<b>第 11 章 ウェブページのランキングのための HITS 手法</b>	<b>149</b>
11.1 HITS アルゴリズム	149
11.2 HITS の実装	151
11.3 HITS の収束	154
11.4 HITS の例	157
11.5 HITS の長所と短所	159
11.6 計量書誌学への HITS のかわり	161
11.7 クエリー独立な HITS	162
11.8 HITS の高速化	165
11.9 HITS の感度	166
<b>第 12 章 ウェブページをランキングするための他のリンク手法</b>	<b>171</b>
12.1 SALSA	171
12.1.1 SALSA の例	171
12.1.2 SALSA の長所と短所	176
12.2 混在型ランキング手法	177
12.3 トラフィック量に基づくランキング	178

<b>第 13 章</b>	<b>ウェブ情報検索の将来</b>	<b>183</b>
13.1	スパム	183
13.2	パーソナル化	187
13.3	クラスタリング	188
13.4	知的エージェント	190
13.5	今後の傾向および時間に関する検索	191
13.6	プライバシーと検閲	193
13.7	図書館分類方式	195
13.8	データ融合	196
<b>第 14 章</b>	<b>ウェブ情報検索のための手引き</b>	<b>197</b>
14.1	始めるにあたって必要なもの	197
14.1.1	データセット	197
14.1.2	クローラー	198
14.1.3	コード	198
14.1.4	参考文献	198
14.2	本格的な研究のための情報源	199
14.2.1	データセット	199
14.2.2	クローラー	199
14.2.3	コード	199
<b>第 15 章</b>	<b>数学的基礎</b>	<b>201</b>
15.1	線形代数	201
15.2	ペロン-フロベニウスの定理	220
15.3	マルコフ連鎖	230
15.4	ペロン補完	246
15.5	確率的補完	255
15.6	打ち切り	257
15.7	凝集	259
15.8	逆凝集	262
<b>第 16 章</b>	<b>用語集</b>	<b>265</b>
	参考文献	<b>273</b>
	索引	<b>285</b>

## コラム

コラム：インデックス化戦争	25
コラム：検索エンジン最適化	56
コラム：検索エンジンはどのようにしてお金をかせぐか？	57
コラム：SearchKing 対 Google	69
コラム：Google 爆弾	72
コラム：RankPulse	86
コラム：Google の上場	115
コラム：Google API	128
コラム：Google ダンス	147
コラム：Google 寡占	147
コラム：固有値ベクトルのランキング	168
コラム：ランク融合と投票手法	177
コラム：Alexa のトラフィック・ランキング	181
コラム：検索の幽霊	183
コラム：KartOO のクラスタ	189
コラム：最初の脳移植	190
コラム：ブログと今後の傾向	191
コラム：Google のクッキー	194
コラム：中国での検索	195
コラム：Google のデジタル図書館計画	196

## 網かけコラム

検索エンジン理解のために	10
検索エンジンにサイトを登録する	21
Spidering Hacks	21
MATLAB クローラー M ファイル	22
インターネットアーカイブプロジェクト	27
キャンベル卿の索引義務化動議	30
ウェブグラフの地図	34
Google ツールバー	35
入リンク機能	37
誰が誰をリンクするか	38
PageRank に影響するマルコフ特性	46
PageRank 問題についての表記法	49
世界最大の行列計算	51
Google 行列の劣位固有値	54
PageRank ベキ乗法のための MATLAB M ファイル	55
サイト内の検索	56
Kaltix のパーソナル化されたウェブ検索	67
パーソナル化 PageRank ベキ乗手法のための MATLAB の M ファイル	68
PageRank の感度のまとめ	78
PageRank 問題の基本行列	84
PageRank とリンクスパム	85
PageRank 問題に対する線形システム	95
Google Hacks	96
エイトケンの外挿付きの PageRank のベキ乗法に対する MATLAB M ファイル	121

二次外挿法つき PageRank ベキ乗法に対する MATLAB M ファイル .....	122
HITS 問題の記号 .....	153
HITS アルゴリズムに対する MATLAB M ファイル .....	154
線形システムの感度 .....	204
シャーマン-モリソン (Sherman-Morrison) の一階更新の公式 .....	205
ジョルダンの定理 .....	208
対角化可能性 .....	209
行列関数 .....	210
一般行列のスペクトル定理 .....	210
無限級数表記 .....	211
対角化可能行列のスペクトル定理 .....	212
0 への収束とノイマン (Neumann) 級数 .....	213
ベキ乗の極限 .....	213
総和可能性 .....	214
ベキ乗法 .....	215
線形定常反復 .....	216
漸近的な収束の速さ .....	218
3 つの古典的分割法 .....	218
M-行列 .....	219
正行列に関するペロンの定理 .....	220
非負行列に対するペロンの定理 .....	221
連結性とランク .....	223
既約性と連結性 .....	225
ペロン-フロベニウスの定理 .....	226
原始行列 .....	227
極限と原始性 .....	227
原始性判定 .....	228
非原始性指数 .....	229
フロベニウス形 .....	229
一時的な挙動 .....	235
既約マルコフ連鎖 .....	238
単位固有値 .....	240
すべての確率行列は総和可能である .....	240
可約なマルコフ連鎖 .....	242
吸収確率と吸収までの時間 .....	245
ペロン補完 .....	246
既約性再び .....	247
主小行列 .....	248
受け継いでいるペロンの性質 .....	249
結合定理 .....	252
打ち切りマルコフ連鎖 .....	256
打ち切り確率分布 .....	258
マルコフ連鎖の凝集定理 .....	260
マルコフ連鎖の逆凝集 .....	262