

Evaluation des systèmes de traitement de l'information



*sous la direction de
Stéphane Chaudiron*

IST 2757

Evaluation des systèmes de traitement de l'information

BIBLIOTHEQUE DU CERIST

7673

© LAVOISIER, 2004

LAVOISIER

11, rue Lavoisier

75008 Paris

Serveur web : www.hermes-science.com

ISBN 2-7462-0862-8

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, d'une part, que les "copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective" et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, "toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause, est illicite" (article L. 122-4). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

**Evaluation
des systèmes de traitement
de l'information**

sous la direction de
Stéphane Chaudiron

hermes
Science
— publications —

*Il a été tiré de cet ouvrage
30 exemplaires hors commerce réservés
aux membres du comité scientifique,
aux auteurs et à l'éditeur
numérotés de 1 à 30*

COMITÉ SCIENTIFIQUE DU
TRAITÉ DES SCIENCES ET TECHNIQUES DE L'INFORMATION

Jacques Rouault
Hubert Fondin
Brigitte Guyot

Le Traité des Sciences et Techniques de l'Information répond au besoin de disposer d'un fonds commun de connaissances dans les domaines où les technologies ont transformé la production et la gestion de l'information. Absorbant et transformant des disciplines traditionnelles, comme la documentation et la bibliothéconomie, les sciences de l'information occupent en effet un champ à part, où se rencontrent l'informatique et les sciences humaines et sociales.

Les ouvrages du Traité des Sciences et Techniques de l'Information analysent les différents aspects d'une discipline étendue et diverse : les documents et leur gestion, l'information dans la société, la communication homme-machine, la représentation et l'instrumentalisation des connaissances. Ils abordent les technologies de base, les institutions, les grandes options méthodologiques et les savoir-faire.

Chaque volume étudie aussi bien les aspects fondamentaux qu'expérimentaux. Une classification des différents chapitres contenus dans chacun, une bibliographie et un index détaillé orientent le lecteur vers ses points d'intérêt immédiats : celui-ci dispose ainsi d'un guide pour ses réflexions ou pour ses choix.

BIBLIOTHEQUE DU CERIST

Liste des auteurs

Frédéric BIMBOT
IRISA
CNRS - INRIA
Rennes

Nozha BOUJEMAA
INRIA
Rocquencourt

Stéphane CHAUDIRON
Centre de recherche en
information spécialisée
Université de Paris 10
Nanterre

Christophe D'ALESSANDRO
LIMSI
CNRS
Orsay

Laurence DEVILLERS
LIMSI
CNRS
Orsay

Christian FLUHR
CEA
LIST/LIC2M
Fontenay-aux-Roses

Brigitte GRAU
LIMSI
CNRS
Orsay

Guillaume GRAVIER
IRISA
CNRS - INRIA
Rennes

Anthony HARTLEY
Centre for Translation Studies
University of Leeds
Royaume-Uni

Madjid IHADJADENE
Centre de recherche en
information spécialisée
Université de Paris 10
Nanterre

Joëlle LE MAREC
Ecole Normale Supérieure Lettres
et Sciences Humaines
Lyon

Ivan MAGRIN CHAGNOLLEAU
Laboratoire Dynamique
du Langage
CNRS
Lyon

Jean-Luc MINEL
LALICC
CNRS

Université de Paris 4

Widad MUSTAFA EL HADI
CERSATES
Université de Lille III
Lille

Adeline NAZARENKO
Laboratoire d'Informatique de
Paris-Nord
Université de Paris 13
Villetaneuse

Patrick PAROUBEK
LIMSI
CNRS
Orsay

Thierry POIBEAU
Laboratoire d'Informatique
de Paris-Nord
Université Paris 13
Villetaneuse

Andrei POPESCU-BELIS
École de Traduction
et d'Interprétation
Université de Genève
Suisse

Table des matières

Introduction	17
Stéphane CHAUDIRON	
 PREMIÈRE PARTIE. LES SYSTÈMES DE RECHERCHE D'INFORMATIONS	25
 Chapitre 1. L'évaluation des systèmes de recherche d'informations textuelles	27
Christian FLUHR	
1.1. Introduction	27
1.2. Historique	28
1.3. Mesures de performance des systèmes booléens	29
1.4. Mesures de performance des systèmes hiérarchiques	30
1.5. Constitution des bases d'évaluation	33
1.6. Les campagnes d'évaluation	34
1.6.1. La constitution des corpus	34
1.6.2. La recherche rétrospective et le filtrage	35
1.6.3. Résultat de l'évaluation	35
1.7. Les pistes	37
1.7.1. La piste très grande collection	37
1.7.2. La piste systèmes interactifs	38
1.7.3. La piste grande précision de TREC	40
1.7.4. La piste interlingue et CLEF	41
1.7.5. L'interrogation en ligne d'Amaryllis	42
1.8. Discussion	43
1.9. Conclusion	44
1.10. Bibliographie	44

Chapitre 2. Evaluation des systèmes de recherche par le contenu visuel : pertinence et critères	47
Nozha BOUEMAA	
2.1. Introduction.	47
2.2. Position du problème	48
2.2.1. Principe de la recherche d'images par le contenu	48
2.2.2. Evaluation quantitative versus qualitative	49
2.3. Ingrédients de l'évaluation.	53
2.3.1. Fidélité des signatures	53
2.3.2. Mesures de similarité.	56
2.3.3. Paradigmes de recherche.	58
2.4. Critères d'évaluation quantitative.	60
2.4.1. Précision et rappel	61
2.4.2. Autres critères	66
2.4.3. Initiatives internationales d'évaluation	68
2.5. Conclusion	73
2.6. Bibliographie.	73
Chapitre 3. L'évaluation des systèmes de question-réponse	77
Brigitte GRAU	
3.1. Les différentes dimensions d'une tâche question-réponse	77
3.2. Les systèmes de question-réponse dans le temps.	80
3.3. Description d'un système de question-réponse	80
3.4. Evaluation de la production finale du système :	
les campagnes d'évaluation	82
3.4.1. Campagnes d'évaluation TREC	83
3.4.2. Campagne d'évaluation CLEF	84
3.4.3. Campagne d'évaluation NTCIR	85
3.4.4. Les questions	85
3.4.5. Les mesures	88
3.4.6. La méthode d'évaluation.	89
3.4.7. Les résultats	90
3.5. Facteurs affectant les performances d'un système de question-réponse.	93
3.6. Evaluation des composants d'un système de question-réponse.	94
3.7. Conclusion	95
3.8. Bibliographie.	96

DEUXIÈME PARTIE. LE TRAITEMENT DE L'INFORMATION TEXTUELLE 99**Chapitre 4. L'évaluation des systèmes d'analyse morphosyntaxique et syntaxique 101**

Patrick PAROUBEK

4.1. Introduction.	101
4.2. Influence des méthodes de segmentation	104
4.3. L'analyse morphosyntaxique	106
4.3.1. Objectif de l'analyse morphosyntaxique	106
4.3.2. Méthodologies et mesures d'évaluation de l'annotation morphosyntaxique.	109
4.3.3. Etat actuel de la technologie de l'annotation morphosyntaxique	113
4.4. L'analyse syntaxique	114
4.4.1. Objectifs de l'analyse syntaxique	114
4.4.2. Méthodologies et mesures d'évaluation de l'analyse syntaxique	117
4.4.3. Etat actuel de la technologie de l'analyse syntaxique automatique	119
4.5. Conclusion	120
4.6. Bibliographie.	121

Chapitre 5. L'évaluation des systèmes d'analyse et de compréhension de texte 127

Adeline NAZARENKO et Thierry POIBEAU

5.1. Introduction.	127
5.2. Evaluer ce qui est compris : un défi	128
5.2.1. Que signifie comprendre un texte ?	129
5.2.2. Qu'est-ce qu'un système de compréhension de texte ?	130
5.2.3. Pourquoi et comment évaluer la compréhension ?	132
5.3. Evaluation technologique	133
5.3.1. L'expérience des tests de lecture et d'écoute.	133
5.3.2. Les tests de compréhension	134
5.4. Evaluation des performances	137
5.4.1. Le principe des conférences d'évaluation MUC	138
5.4.2. Une évaluation objective.	138
5.4.3. L'évolution des conférences MUC.	140
5.5. Limites et points forts des expériences d'évaluation.	141
5.5.1. Une utilisabilité difficile à mesurer	142
5.5.2. Un puissant moteur pour la recherche	143
5.5.3. Quelle évaluation pour quelle recherche en compréhension ?	144
5.6. Conclusion	145
5.7. Bibliographie.	146

Chapitre 6. L'évaluation d'outils d'acquisition**de ressources terminologiques 149**

Widad MUSTAFA EL HADI

6.1. Introduction.	149
6.2. Etat de l'art de l'acquisition de ressources terminologiques	150
6.2.1. Acquisition automatique de termes	150
6.2.2. Spécificités des outils de structuration de terminologie.	151
6.2.3. Aperçu des approches de structuration de relations sémantiques	152
6.2.4. Comment les outils d'ART peuvent-elles servir à l'élaboration des ontologies ?	154
6.3. Protocole d'évaluation	155
6.4. Evaluation d'outils d'acquisition de ressources terminologiques	157
6.4.1. Evaluation des extracteurs de termes	157
6.4.2. Evaluation des extracteurs de relations sémantiques : éléments de réflexion	158
6.5. Les problèmes posés par l'évaluation d'outils d'acquisition de ressources terminologiques	160
6.5.1. Perspectives	162
6.5.2. La prise en compte de la dimension de l'usage	162
6.5.3. Définition d'un protocole par une catégorie d'outils	163
6.5.4. Des modèles à tester	163
6.6. Conclusion	165
6.7. Bibliographie.	165

Chapitre 7. L'évaluation des systèmes de résumé automatique. 171

Jean-Luc MINEL

7.1. Introduction.	171
7.2. Evaluation par calcul	172
7.3. Evaluation par jury	175
7.3.1. Les protocoles du système SERAPHIN.	175
7.3.2. Le programme TIPSTER et son évaluation.	179
7.3.3. Le programme DUC	180
7.4. Synthèse critique sur l'évaluation.	183
7.5. Bibliographie.	184

TROISIÈME PARTIE. LES SYSTÈMES DE TRAITEMENT DE LA PAROLE. 187**Chapitre 8. L'évaluation des systèmes de reconnaissance de parole. 189**

Frédéric BIMBOT et Guillaume GRAVIER

8.1. Présentation.	189
8.2. Terminologie et typologie	190

8.2.1. Les tâches de reconnaissance de parole	190
8.2.2. Les facteurs de variabilité	191
8.2.3. Profil des systèmes de reconnaissance.	191
8.2.4. Fonctionnement d'un système de reconnaissance vocale.	196
8.3. Mesures de performance	197
8.3.1. Reconnaissance en ensemble fermé	197
8.3.2. Reconnaissance en ensemble ouvert.	199
8.3.3. Caractéristique de fonctionnement et fonction de coût	202
8.3.4. Comparaison de transcriptions	203
8.3.5. Evaluation de la mesure de confiance	205
8.3.6. Limites des métriques actuelles.	206
8.4. Campagnes d'évaluation	207
8.4.1. Protocole d'évaluation	207
8.4.2. Quelques grandes campagnes d'évaluation.	208
8.5. Conclusions.	212
8.6. Bibliographie.	212
8.6.1. Evaluation	212
8.6.2. Mesures de performance.	212
8.6.3. Courbes DET	212
8.6.4. Campagne NIST.	212
8.6.5. Campagne AUPELF-B1	213
8.6.6. Campagne ESTER	213

Chapitre 9. L'évaluation des systèmes de synthèse de la parole 215

Christophe D'ALESSANDRO

9.1. Introduction.	215
9.1.1. Typologie des systèmes de synthèse.	215
9.1.2. Composantes d'un système de synthèse.	216
9.1.3. Différents types d'évaluation	217
9.1.4. Typologies de tests, d'utilisateurs et d'applications.	219
9.1.5. Contraintes techniques des systèmes.	220
9.2. Evaluation globale des systèmes	221
9.2.1. Netteté d'articulation ou intelligibilité segmentale	221
9.2.2. Compréhension au niveau de la phrase	222
9.2.3. Évaluation de qualité globale : appréciations explicites	224
9.2.4. Evaluation de la charge cognitive	228
9.2.5. Evaluation de la qualité des voix	229
9.3. Evaluation analytique	231
9.3.1. Transcription graphème-phonème	231
9.3.2. Intelligibilité des noms propres	232
9.3.3. Evaluation de la prosodie	232
9.3.4. Evaluation du synthétiseur acoustique.	234

14 Traitement de l'information

9.3.5. Evaluation des discontinuités de sélection	235
9.4. Conclusion	235
9.5. Bibliographie	236

Chapitre 10. L'évaluation des systèmes de vérification du locuteur 241

Ivan MAGRIN-CHAGNOLLEAU

10.1. Introduction : quelques définitions	241
10.1.1. La reconnaissance du locuteur	241
10.1.2. Les différentes tâches et leurs erreurs	242
10.1.3. Dépendance au texte	247
10.2. Les composantes d'un système de vérification du locuteur	248
10.2.1. Architecture globale d'un système de vérification du locuteur	248
10.2.2. La paramétrisation du signal	248
10.2.3. La modélisation du locuteur	249
10.2.4. La phase de vérification – l'obtention d'un score	249
10.2.5. Les normalisations de scores	249
10.3. Les applications de la vérification du locuteur	249
10.3.1. Les applications sur site	250
10.3.2. Les applications à distance	250
10.3.3. Les applications judiciaires	251
10.3.4. L'indexation sonore	252
10.3.5. Les jeux	252
10.3.6. La personnalisation des services	252
10.4. L'évaluation des performances	253
10.4.1. Les deux taux d'erreur	253
10.4.2. La fonction de coût	253
10.4.3. HTER : une fonction de coût particulière	254
10.4.4. Le taux d'égale erreur	254
10.4.5. La courbe DET	254
10.4.6. Les EPC	255
10.5. Les évaluations à grande échelle en vérification du locuteur	256
10.5.1. Les évaluations NIST	256
10.5.2. L'évaluation ESTER	257
10.6. Bibliographie	257

Chapitre 11. L'évaluation des systèmes de dialogue oral 259

Laurence DEVILLERS

11.1. Introduction	259
11.2. Systèmes de dialogue oral homme-machine	261
11.3. Les enjeux et les objectifs de l'évaluation	263
11.3.1. Les enjeux de l'évaluation	263

11.3.2. Les objectifs de l'évaluation	263
11.4. Etat de l'art des méthodologies d'évaluation	265
11.4.1. Cadres d'évaluation	265
11.4.2. Axes de l'évaluation	267
11.4.3. Méthodologies d'évaluation	270
11.5. Discussion et conclusion	280
11.6. Bibliographie	281
QUATRIÈME PARTIE. L'APPROCHE USAGE DANS L'ÉVALUATION	285
Chapitre 12 La place de l'utilisateur dans l'évaluation des systèmes de recherche d'informations	287
Stéphane CHAUDIRON	
12.1. Introduction	287
12.2. L'évaluation des SRI dans le paradigme système.	288
12.2.1. Les caractéristiques de l'évaluation selon le paradigme « système »	290
12.2.2. Les limites du paradigme technique	292
12.3. L'évaluation dans le paradigme usager.	295
12.3.1. La question de la pertinence	296
12.3.2. L'approche cognitive	299
12.3.3. L'approche « processus »	301
12.3.4. L'approche holistique.	303
12.4. Conclusion	305
12.5. Bibliographie	306
Chapitre 13 L'évaluation des systèmes de traduction automatique	311
Anthony HARTLEY et Andrei POPESCU-BELIS	
13.1. Spécificités de la traduction automatique et de son évaluation	311
13.2. Le cadre théorique de l'évaluation de la TA.	313
13.2.1. Exploitation des normes ISO/IEC pour l'évaluation des logiciels	313
13.2.2. Evaluation contextuelle de la TA : le cadre FEMTI	314
13.2.3. Formalisation de l'évaluation par étapes.	317
13.3. Métriques visant la qualité du texte produit	319
13.3.1. Métriques nécessitant des juges humains	319
13.3.2. Métriques automatisées.	323
13.4. Analyse des campagnes d'évaluation en TA	325
13.4.1. Campagne ALPAC, 1966	326
13.4.2. Campagnes initiées par la DARPA, années 1990	326
13.4.3. Campagnes initiées par le NIST, 2002-2003	327

13.4.4. Autres spécifications d'évaluations : JEIDA, ATA, SAE, CESTA	327
13.5. Métaévaluation des métriques	330
13.5.1. Critères d'évaluation des métriques	330
13.5.2. Comparaison des métriques sur des traductions humaines et automatiques	331
13.6. Perspectives	333
13.7. Bibliographie	333

Chapitre 14. Usages des moteurs de recherche 337

Madjid IHADJADENE

14.1. Introduction	337
14.2. Pratiques informationnelles de l'Internet.	338
14.3. Évaluation des performances des moteurs de recherche	339
14.4. Analyse des termes des usagers et formulation de la requête	341
14.5. Tactique et stratégie de recherche	343
14.6. Reformulation et bouclage de pertinence	344
14.7. Surcharge d'informations et moteurs de recherche	345
14.8. Aspects cognitifs de la recherche d'information	347
14.9. Conclusion	348
14.10. Bibliographie	349

Chapitre 15. Les études d'usage et leur prise en compte dans le champ culturel 353

Joëlle LE MAREC

15.1. Introduction	353
15.2. La relation entre conception et études d'usages.	354
15.3. L'articulation entre études d'usages et enjeux de développement : le cas du milieu muséal	357
15.4. Les recherches centrées sur les usages : synthèses et évolution de la notion.	364
15.5. L'informatisation et les réseaux sur les terrains des sciences sociales	366
15.6. Entre le besoin d'un savoir sur les usages et l'évolution de la notion : un hiatus important	368
15.7. Conclusion : affronter les paradoxes	369
15.8. Bibliographie	370

Index 373