Information Retrieval

Design and Sources

Roslina Othman

IIUM PRESS
INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

INFORMATION RETRIEVAL: DESIGN AND SOURCES

Editor

Roslina Othman

IIUM Press

Published by:
IIUM Press
International Islamic University Malaysia

First Edition, 2011
©IIUM Press, IIUM

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without any prior written permission of the publisher.

Perpustakaan Negara Malaysia Cataloguing-in-Publication Data

Roslina Othman: Information Retrieval: Design and Sources

ISBN: 978-967-418-103-1

Member of Majlis Penerbitan Ilmiah Malaysia – MAPIM
(Malaysian Scholarly Publishing Council)

Printed by :
IIUM PRINTING SDN. BHD.
No. 1, Jalan Industri Batu Caves 1/3
Taman Perindustrian Batu Caves
Batu Caves Centre Point
68100 Batu Caves
Selangor Darul Ehsan

TABLE OF CONTENTS

1. IR MODELS AND APPROACHES: AN OVERVIEW 1
Roslina Othman

2. SIMILARITY-BASED, RELEVANCE-BASED AND INFERENCE-BASED MODELS 7
Suleiman Shammas Ali and Roslina Othman

3. COGNITIVE AND BEHAVIORAL APPROACHES IN INFORMATION RETRIEVAL.. 13
Zalifah Awang Zakaria and Roslina Othman

4. ANOMALOUS STATES OF KNOWLEDE IN INFORMATION RETRIEVAL 19
Zalifah Awang Zakaria and Roslina Othman

5. THE INFLUENCE OF ASK MODEL ON USER-SYSTEM INTERACTION 31
Zalifah Awang Zakaria and Roslina Othman

6. UNCERTAINTY AND ITS RELATIONSHIP WITH RELEVANCE 37
Zalifah Awang Zakaria and Roslina Othman

7. THE CURRENT TRENDS AND DEVELOPMENTS IN INFORMATION RETRIEVAL 45
Zalifah Awang Zakaria and Roslina Othman

8. NATURAL SYMMETRY BETWEEN QUERIES AND DOCUMENTS 51
Suleiman Shammas Ali and Roslina Othman

9. STRENGTHS AND WEAKNESSES OF IR EVALUATION TESTS 55
Roslina Othman and Zalifah Awang Zakaria

10. LABORATORY VERSUS OPERATIONAL SYSTEM TEST 61
Suleiman Shammas Ali and Roslina Othman

11. NON-ENGLISH MONOLINGUAL IR CASES: AN OVERVIEW 67
Roslina Othman

12. STEMMING FOR DOCUMENT IN NON-ENGLISH DOCUMENTS 73
Roslina Othman and Ouahiba Saoudi

13. STEMMING FOR DOCUMENT IN ENGLISH LANGUAGE 83
Roslina Othman and Ouahiba Saoudi

14. ADDAALL STEMMER FOR ARABIC NEWS OF AL-JAZEERA 91
Roslina Othman and Ouahiba Saoudi

15. MORPHOLOGICAL ANALYSIS FOR RETRIEVAL OF DOCUMENTS IN ARABIC LANGUAGE 99
Roslina Othman and Ouahiba Saoudi

16. *QURANIC* TEXTS IN MULTI-SCRIPT ENVIRONMENT: AN OVERVIEW TO THE REQUIREMENTS 107
Roslina Othman and Fauziah Abdul Wahid

17. MALAY TRANSLATION OF THE *QURANIC* TEXTS IN INFORMATION RETRIEVAL 113
Roslina Othman and Fauziah Abdul Wahid

18. PROCEDURES FOR EVALUATING *QURANIC* TEXTS IN MULTI-SCRIPT ENVIRONMENT 119
Roslina Othman and Fauziah Abdul Wahid

19. RETRIEVAL PERFORMANCE OF QURANIC TEXTS IN MULTI-SCRIPT ENVIRONMENT 127
Roslina Othman and Fauziah Abdul Wahid

20. INDEXING MODELS AND SEARCHING FEATURES: AN OVERVIEW 135
Roslina Othman

21. PRECISION-BASED INDEXING 139
Roslina Othman and Siti Fatimah Mohd Tawil

22. ONTOLOGY-BASED INDEXING 147
Roslina Othman and Siti Fatimah Mohd Tawil

23. A REVIEW OF THE PROBABILISTIC INDEXING 157
Suleiman Shammas Ali and Roslina Othman

24. INTERACTIVE TOPIC DETECTION AND TRACKING 161
Masnizah Mohd

25. RETRIVAL FEATURES OF SCIENCEDIRECT 173
Nur Leyni Nilam Putri Junurham, Nurul Hasni Abu Hassan and Roslina Othman

26. RETRIEVAL FEATURES OF EMERALD 179
Roslina Othman, Nor Sa'adah Md. Nor and Nik Roslina Raja Ismail

27. RETRIEVAL FEATURES OF EBSCO HOST 185
Muhammad Alif Ismail, Mohamad Hafizuddin Mohamed Najid and Roslina Othman

28. RETRIEVAL FEATURES OF IEEE XPLORE 191
Muslim Ismail @ Ahmad, Syafrizal Hj Saulan and Roslina Othman

29. CLASSIFCATION-BASED NOVELTY SEARCH FOR PATENT APLICATION IN USPTO 197
Noorfatin Muhamad Sharhabil and Roslina Othman

30. SEARCHING FOR MALAYSIAN GOVERNMENT DOCUMENTS 203
Ruzaimah Mohammed Zain, Wan Ali Wan Mamat and Roslina Othman

31. BORN-DIGITAL MATERIALS: RETRIEVABLE OR IRRETRIEVABLE? 209
Nik Roslina Raja Ismail and Roslina Othman

32. CORPUS AND SOURCES: AN OVERVIEW 215
Roslina Othman

33. ISSUES OF CORPUS ON SUFISM 221
Muslim Ismail, Murshaidi Hazlin Mohd, Muhammad Arif Osman and Roslina Othman

34. Sufism Literature in Online Databases 227
Muslim Ismail, Murshaidi Hazlin Mohd, Muhammad Arif Osman and Roslina Othman

35. Common Issues of Islamic Manuscripts 231
Ruzaimah Mohammed Zain, Nik Roslina Raja Ismail, Haniza Adnan and Roslina Othman

36. Possible Corpus for Islamic Manuscripts 237
Ruzaimah Mohammed Zain, Nik Roslina Raja Ismail, Haniza Adnan and Roslina Othman

37. Online Access to Islamic Manuscripts 243
Ruzaimah Mohammed Zain, Nik Roslina Raja Ismail, Haniza Adnan and Roslina Othman

21. PRECISION-BASED INDEXING

Roslina Othman and Siti Fatimah Mohd Tawil

ABSTRACT

This chapter reports on precision-based indexing aimed to assists users in targeting a high precision value in their search results even when items were ranked due to reasons such as information overload and too many similar images. Precision-based indexing took the form of specificity in vocabularies, uniqueness in properties, inclusion of meanings, specificities within a domain, hierarchical structure within a tree giving the selection of narrowness, inferences unique to a collection, and specificity within users' query. The target in all these forms was high precision values even among ranked items. However, limitations in some of these techniques included time spent on crafting pre-defined templates and adoption in real setting involving large size collections. All techniques were tested and significantly improved precision as targeted.

21.1 Introduction

Getting as many relevant items as possible is no longer worthy when achieved at high cost of precision, and especially when recall is not guaranteed. The situation of information overload though claimed could be solved through the use of precise search terms from the thesaurus produced results that still did not meet user's expectation in meeting his/her information needs. In meeting their expectation, users are targeting a certain level of precision values that depends on their tolerance in browsing for relevant items.

Precision in general indicates the fraction of relevant items in a search to all items retrieved for the question submitted by users. System relevance would give a range of 50% to 100% precision; however, would