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

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

11. NON-ENGLISH MONOLINGUAL IR CASES: AN OVERVIEW

Roslina Othman

.BSTRACT

'his chapter provides an overview on the non-English monolingual IR :ases. These works mostly focused on stemming and transliterations. Stemming needs to be integrated with a morphological analyzer, while transliterations with a certain standard. Stemming is aimed at improving recall, and transliteration to overcome terminology barrier. These cases revealed issues such as ambiguity and transformation of letters. One solution is to develop a comprehensive is to develop a comprehensive list of morphological variants for stemming and dictionary of transliterated words with origins of the words.

11.1 Introduction

This chapter introduces briefly the cases of IR non-English monolingual IR cases with special reference to stemming and transliterations. Arabic for example is a highly inflected language, and thus most works were directed towards stemming. In the field of Science and Technology, Arabic and other languages faced terminology barrier, and thus transliteration was adopted. Even for texts in non-Roman scripts such as Arabic that required translations into other languages also require transliteration.

Monolingual IR cases were evaluated as a track in Cross-Language Evaluation Forum (CLEF) since 2000, which includes French, German, Italian, Spanish, Arabic, and Chinese among others. In CLEF,