Dedication

It is my pleasure to dedicate this research to my beloved

Parents,

Husband,

Brothers,

Sisters,

And friends……

Acknowledgement

I wish to record my thanks and gratitude to my supervisor Prof. Dr. Izzeldin Mohammed Osman, I am grateful to him for entrusting me with the task of preparing this research .I hope his trust has not been misplaced.

I also sincerely appreciate the valuable help and obtained from Ustaz\ Hisham Abdallah Monsoor, Yassir Sedig, and also the assistance of all Computer Science and Information Technology College staff who supplied me with the research resources and gave me their advice.

I acknowledge also with grateful the care of my friends. Finally I wish to record my thanks to my family and husband for their encouragement.

Abstract

How to find needed information from the web is a critical issue in the Internet. Fortunately, search engines are useful tools to retrieve information from the Internet.

Although, Internet users speak different languages most of the resources are written and published in English. This research investigates multilingual search and ways to develop multilingual search engine models. The purpose of multilingual search is to help people finding useful contents stored in multiple languages. This research examines the case of English and Arabic languages.

The research addresses search engines and information retrieval topics such as ranking, crawling, indexing resources, query execution and results relevancy.

The model presented for multilingual search engine is illustrated by a practical example of the on-line site of Sudan University of Science and Technology journal.

مستخلص البحث

مما لا شك فيه ان شبكة الانترنت تعتبر من اكبر المصادر المعلوماتيه حيث تزخر بشتى ضروب المعلومات وتصنيفاتها ونجد ان محركات البحث ادوات مفيده جدا لمستخدمى الانترنت من طلاب وباحثين. وبالرغم من الكم الهائل من المعلومات وتنوع الجنسيات واللغات الا ان معظم معلومات الشبكة العنكبوتيه تكتب وتنشر باللغة الانجليزية.

من هنا تنبع اهميه وجود محركات بحث تتمتع بامكانيه البحث متعدد اللغات حيث توفر للمستخدم نتائج البيانات المخزونة بلغات متعددة . هذا البحث يتناول البحث بلغات متعددة ويحاول تطوير نموذج لمحرك بحث متعدد اللغات.والهدف الاساسى من البحث متعدد اللغات مساعدة مستخدمى محرك البحث الحصول على المعلومة المطلوبة بأكثر من لغة ويركز على البحث باللغتين العربية والانجليزية.

ولتحقيق ما سبق ذكره كان لا بد من الاهتمام بالموضوعات المتعلقة بمحركات البحث واسترجاع المعلومات مثل نظام الرتبه للمواقع، الزحف على الانترنت، فهرسة مصادر المعلومات، تنفيذ الاستعلامات وملاءمة المعلومات المسترجعة للإستعلامات .

ونعرض مثالاً عملياً لمحرك بحث باللغة العربية والإنجليزية يقوم ببحث موقع مجلة العلوم والتقانة التابع لجامعة السودان للعلوم والتكنولوجيا.

Table Of Contents

Chapter 1:Litreture Review

Chapter 2: Internet Search Engines

Chapter 3:Google Search Engine

Chapter 4: Intranet Search Engines

Chapter 5: Multilingual Search Engine Model

Chapter 6: Case Study

Table of Figures

Table of Tables

Introduction

A great number of search engines can be used to retrieve relevant information on the internet. This research focuses on multilingual information retrieval to help searchers find useful contents and documents.

The researcher developed model based on the characteristics of search engines and the architecture of the client-server system.

Chapter1: the research starts by giving a general overview of internet, search engines and then turns to the research problem, research overview, research goals and objectives, research methodology, research risks, and the related works.

Chapter2: In chapter two the research gives more details in search engines. It starts by defining them; gives their historical background, how they work and their components. It then goes to an important issue, which is considered as the corner stone of the search process, it is query processing and how it can be optimized particularly by caching of queries results using replacement algorithms. The step after processing a query is information retrieval and how to display the results to the user. Then it turns to describe the features, characteristics and the architectures of the search engines as well as the problems and challenges they face. Lastly this chapter makes a little focus on cloaking and spamming techniques.

Chapter3: this chapter talks about Google as a hypertextual search engine. It describes design goals, system features, anchor text, Google architecture and data structure, crawling the web, searching and ranking system, and finally performance including storage requirement, system and search performance.

Chapter4: chapter four describes the intranet search engines start with definition, how they work, collecting information to index it, search an intranet, intranet metadata, important features for intranet search and the process of developing a successful intranet. Then the chapter turns to a comparison between internet search engine and intranet search engine from axioms and structural points of views.

Chapter5: this chapter describes the proposed multilingual search engine model. The language used to build this model is UML (Unified Modeling Language). The main designed models are logical model (class diagram), use case model, and dynamic model (sequence and activity diagrams).

Chapter6: this chapter describes the multilingual search engine case study.

Chapter7: this chapter contains conclusion and recommendations for future researches and studies.

Lastly a list of references and an appendix containing the case study program are included.