المستخلص

الثورة العظيمة في صفحات الإنترنت أدت الي زيادة حجم النصوص في صفحات الإنترنت والتي قادت الي طرح الاسئلة حول كيفية التعامل مع هذا الكم الهائل من المعلومات . لذلك نحتاج الي طرق تمكننا من ايجاد البيانات بصوره مناسبه ومبسطه لكسب الزمن مع مراعاة عدم فقدان البيانات المهمه .

في هذا البحث تمت مناقشة موضوع التلخيص الالي للنصوص العربيه .نظراً لما تتميز به اللغة العربية من مميزات خاصه جعلتها من اللغات التي تحتاج الي فهم دقيق لمكوناتها ومميزاتها والصعوبات التي تواجهك في ضبط تشكيل الحروف وبناء الجملة السليمة .

تم التطرق الي كيفية بناء الجمله في اللغة العربية بالرجوع الي جذور الكلمة . ايضا تم التطرق الي مفهوم التلخيص الالي من مختلف النظريات والتعاريف مع ذكر امثلة الي مختلف البرامج التي تدع التلخيص الالي للنصوص سواء كانت باللغة الانجليزيه او باللغة العربية .

تم اخذ برنامج SARA© كدراسة حالة لبرامج التلخيص التي تعمل على النصوص العربية والذي قامت بتصميمه شركة RDI المصرية والتي تخصصت في البرامج التي تخدم اللغة العربية .

قدم هذا البحث طريقة لتقييم النصوص العربيه المستخلصة من النص الاصلي بحساب واحدات تدعي Recall, precision and F-measure . يتم تجريب النظام بعدة نصوص ومقارنة هذه النصوص مع تلخيص البشر المختص في هذا المجال . كما تم الحكم علي نتائج النظام من قبل المختصين . حيثم تم إختيار عدة نصوص عربية مختلفة من مواقع علي الانترنت مع مراعاة الاختلاف في مضمون النص مثلا النص التاريخي والنص العلمي والنص الاخباري ... الخ .

لخصت هذه النصوص عن طريق برنامج SARA© وايضا تم تلخيصها عن طريق اساتذة متخصصون في اللغة العربية وتمت مقارنة تلخيص SARA© مع تلخيص الاساتذة وقد تحصلنا علي نتائج جيدة ومشجعه تثبت ان النظام يمكن الاعتماد عليه .

Abstract

The evolution of the World Wide Web led to increase the size of Arabic texts on the Internet, which led to ask many questions about how to behave in this huge and growing number of these texts. And therefore need sure has become more than ever to provide the necessary means for rapid browsing of the texts, in order to enable the user to estimate the degree of relevance of the document to the required information.

The search has been subjected to debate automatic summarization for Arabic texts. Because Arabic Language has special features made it one of the languages that need to be carefully to understand the components and the characteristics and difficulties they experience in controlling the formation of letters and syntax sound.

Were mentioned how the syntax of the Arabic language by reference to the word roots. Also we mentioned the concept of automatic summarization of the various theories and definitions with examples to the various programs that support automatic summarization of texts, whether in Arabic or English We taken SARA © program as a case study for the programs that support the summary Arabic texts which designed by RDI Egyptian company which specialized in programs that serve Arabic.

This research introduces evaluation methods for an Arabic extractive text summarization system by calculating the Recall, precision and F-measure. The system is trainable and uses manually annotated corpus. We have introduced methods for evaluating the summary against other human summaries. Moreover, we used human judgment for system output, and finally we tested the system against a commercial Arabic summarization system. We selected several different Arab texts from Web sites, taking into account the difference in the content of the text, for example, the historical text and text and text scientific news ... etc.

The texts were Summarized by SARA © program and also summarized by a human specializing in the Arabic language and were compared between them. We have acquired a good encouraging results prove that the system can be relied upon it.

Acknowledgments

First of all, I thank Allah for everything and in any situation. All thanks go to Allah for all the knowledge and education I gain which leads to the achievement of this thesis.

Second, I would like thank my parents, brothers and sisters for their encouragement, support and advices in all my life stages;.

It is my pleasure to be supervised by *prof. Izz Eldin Mohammed Osman*. Their patience was unexpected; he gave my invaluable advices and help me in any time I ask him .

I want to thank group of teachers : Mohammed Abd-Elrahman(university of Africa – Africa center). Dr Altiraify AbdAllah- future university, Ahmed Dafaa-allah (Quran Kraeem University – Madami branch) ,DR. Khalid Abuzaid – Africa university – Arabic language institute ,Dr. Kamal Mohammed Jah Allah –Africa university – College of Arts , Dr. Mubarak Mohammed – Yusuf Alkhalifa Institute for their cooperation and they help me to fulfill this work.

Thanks to all member of the faculty of computer science in Sudan University for science and technology .Thanks for all my colleague and friends

Dedication

There are a number of people without whom this thesis might not have been written,

And to whom I am greatly indebted.

To who taught me patience and the success , To who I missed in facing difficulties ,to who he left without fed of his support and encourage , To my brother soul (Abd-Albagi)

To light that illuminates the path of success to me my father, to who taught me withstand whatever conditions changed my mother.

To those who were light my way And support me , my brothers and sisters , to my wife .

To all my friends and colleague .

List of Tables

List of Figures

Contents