استمارة مستخلصات رسائل واطاريح الماجستير والدكتوراه في جامعة البصرة

| | |
|---|---|
| الكلية: كلية العلوم | اسم الطالب: فراس حمودي نعمة عبود |
| القسم: علوم الحاسبات | اسم المشرف: أ.م.د. سلمى عبد الباقي محمود |
| التخصص: تنقيب البيانات | الشهادة: ماجستير |

عنوان الرسالة او الاطروحة:

| |
|---|
| تنقيب النصوص العربية باستعمال العنقدة |

ملخص الرسالة او الاطروحة:


تركز هذه الدراسة على بناء نظام تنقيب النصوص العربية باستعمال العنقدة وباتباع طريقتين مختلفين: تتضمن الطريقة الأولى تطبيق تقنية عنقدة النصوص العربية والتي تم تطبيقها على ٢٠٠ مستند من النصوص الرياضية وذلك باستعمال خوارزميتي k-mediods وk-means ومقياسي التشابه (Euclidean، Cosine) . مشكلة هذه الطريقة هي ضخامة الخصائص الناتجة من النصوص والداخلة في عملية العنقدة والتي قد تؤثر على كفاءة النظام وفي محاولة لاستخلاص المواضيع الأساسية في النصوص العربية في المدونة الرياضية ولتقليل خصائص هذه النصوص تم اقتراح طريقة اخرى ألا وهي استعمال تقنية تحليل الإحالة المشتركة Coreference Resolution وتطبيقها قبل تطبيق عملية العنقدة على النصوص العربية، وباستعمال خوارزميتي k-mediods وk-means ايضاً وبمقياسي التشابه نفسهما. واستعملت ثلاثة مقاييس لقياس كفاءة الطريقتين وهي: Precision، Recall، F-Measure.


| | |
|---|---|
| College: College Of Science | Name: Firas Hammoodi Neamah |
| Dept.: Computer Science Dept. | Name Of Supervisor: Dr. Salma A. Mahmood |
| Specialization: Data Mining | Certificate: Master In Computer Science |

Title of Thesis:

| |
|---|
| Arabic Text Mining Using Clustering |

Abstract of Thesis:


This study focuses on the construction of exploration Arabic texts system using clustering and following different two methods. The first method includes clustering Arabic texts technology, which is applied on the 200 sports text and using algorithms (k-mediods, k-means) with similarity metrics (Euclidean, Cosine). The problem this method is the huge features to the documents and may affect the efficiency of the system. In order to attempt extracting the basic topics in sports corpus and to reduce the features .the new method proposed the using coreference resolution which is used before applying clustering process on Arabic documents, also using algorithms (k-mediods , k-means) with the same similarity metrics. This study is used three metrics to measure the efficiency of the methods which Precision, Recall.