

استمارة مستخلصات رسائل وأطاريح الماجستير والدكتوراه في جامعة البصرة

الكلية : العلوم
القسم : حاسبات
التخصص : لغات طبيعية
عنوان الرسالة أو الاطروحة :
اسم الطالب : خديجة علك حسن
اسم المشرف : د. زينب علي خلف
الشهادة : ماجستير

تصنيف الاخبار العربية باستخدام نظام الدمج وطريقة اختيار الصفات

ملخص الرسالة أو الاطروحة :

الهدف من هذا البحث هو تصميم نظام آلي لتصنيف المستندات وتحسين انجازية النظام باستخدام عدد اقل من الصفات . في هذا البحث تم اقتراح نهجين هما تصفية الصفات باستخدام أنواع ن-غرام و دمج الأنظمة. هذه الأنظمة طبقت على مدونة باللغة العربية .

نهج التصفية قلل اعداد الصفات من (5,799) و (12,474) بالنسبة لطريقة التشابه باستخدام جيب تمام الزاوية وخوارزمية الشعاع المدعم الى (3,404) ، (3,466) و (3,400) باستخدام نظام التصفية المقترحة باعتماد انواع مختلفة من ن-غرام ، حيث قلل النظام المقترح الصفات من دون التأثير على انجازية نظام التصنيف . كذلك تم تحسين انجازية تصنيف النصوص من (81,6) و (91.2) بالنسبة لطريقة التشابه باستخدام جيب تمام الزاوية وخوارزمية الشعاع المدعم الى (94.2) بالنسبة لنهج دمج الأنظمة .

College: Science

Name of student: Khadeija Alak Hassan

Dept: Computer Science

Name of supervisor: Dr. Zainab Ali Khalaf

Specialization: Natural Language

Certificate: Master

Title of Thesis:

System Combination and Filtering approach for Arabic news classification

Abstracts of Thesis:

The objective of this thesis is to design an automatic text classification system and improve its performance with less number of features. Two approaches are proposed in this research. They are filtering features of N-gram types and system combination. The systems are applied on Arabic corpus.

The filtering approach reduced the number of features from (5,799) and (12,474) for cosine and SVM systems to (3,404), (3,466) and (3,400) for N-gram types representation without reduce the performance. The performance of text classification is enhanced from (81.6) and (91.2) for cosine and SVM respectively to (94.2) for system combination.