

الملخص

تعتبر عملية التلخيص التلقائي للنصوص من أهم المهام المتعلقة بمجال التنقيب عن البيانات واسترجاع المعلومات. تكمن أهمية التلخيص التلقائي للنصوص بأنها قادرة على استخراج أكثر المعلومات أهمية من مجموعة كبيرة من المستندات النصية. وبسبب تزايد كميات الوثائق بشكل ملحوظ في مجالات مختلفة مثل المقالات الاخبارية ورسائل البريد الالكتروني وتقارير الاعمال والتقارير الطبية وما الى ذلك وهذا بسبب التطور الكبير في مجال تكنولوجيا المعلومات والبيانات الكبيرة التي نتجت وانتشرت عن طريق وسائل التواصل الاجتماعي المختلفة وازدياد عدد المستخدمين لها . لذلك أصبحت عملية التلخيص التلقائي للنصوص المتعددة كبيرة الحجم عملية تحدي وذلك بسبب الزيادة المطردة والمتواصلة لأنشطة البشر والمصادر المختلفة للبيانات خاصة وأن الابحاث المتعلقة بتلخيص الوثائق باللغة العربية لا يزال في مراحلها المبكرة مقارنة مع اللغات الاخرى كالانجليزية مثلا . وبناء على ذلك يقترح هذا البحث نموذج لعمل التلخيص التلقائي لمجموعة من النصوص العربية كبيرة الحجم باستخدام خوارزمية التجميع وطريقة استخراج الخصائص.

ويركز النموذج المقترح في هذا البحث على تلخيص الوثائق العربية المتعددة بشكل فعال وصحيح من خلال فحص جودة الملخص اللغوية ، والاعتماد على خوارزمية تجميع خطية للحصول على ملخص سريع ومتناسك ومتسلسل في أفكاره ودقيق. كما ويستخدم هذا النظام أيضا إطار مفاهيمي يستند إلى بحوث منشورة تتناول تقنيات تلخيص النص بلغات مختلفة. وقد تم استخدام مجموعة بيانات مشهورة وكبيرة بالحجم وتتضمن مجالات مختلفة من المقالات الاخبارية ، مثل التعليم والرياضة والسياسة والفن والتكنولوجيا ، كما و تحتوي مجموعة البيانات هذه على نصوص مختلفة الأحجام وقد تم عمل التجارب لفحص النظام بالارتكاز فقط على المقالات الاخبارية . ويغطي النموذج المقترح النقص في أنظمة التلخيص الآلي العربية المعروفة في جودة الملخص النهائي من خلال اجراء عمليات فحص لغوية واستخراج ميزات خاصة للجمل الامر الذي أدى الى انتاج ملخص دقيق ومتناسك ليكون مشابهها تماما للملخصات البشرية . حيث أثبتت النتائج بأن هذا النظام أظهر وجود مؤشر جيد بالنسبة لمقاييس الاستدعاء والدقة (**Precision and Recall**)، وهذا يشير الى أن النظام قد قام باسترجاع الجمل المهمة من هذا العدد الكبير من النصوص وقد حقق نسبة دقة بلغت 69.4%.