جامعة حلوان
كلية الحاسبات والمعلومات

قسم نظم المعلومات

تلخيص النصوص ثنائية اللغة

رسالة مقدمة لإستكمال متطلبات الحصول على درجة دكتوراه الفلسفة فى
في الحاسبات والمعلومات- تخصص نظم المعلومات

اعداد الباحثه

رشا محمد بدرى سيد

بكالوريوس الحاسبات والمعلومات، نظم المعلومات، مايو ٢٠٠٣
ماجيستير الحاسبات والمعلومات، نظم المعلومات، يناير ٢٠٠٨

تحت اشراف

أ.د. أحمد شرف الدين أحمد
أستاذ بقسم نظم المعلومات
كلية الحاسبات والمعلومات
جامعة حلوان

أ.م.د. دعاء سعد الزنفلى
أستاذ مساعد بقسم نظم المعلومات
كلية الحاسبات والمعلومات
جامعة حلوان

٢٠١٥

ملخـــــص

تلخيص النصوص ثنائية اللغة

يواجه مستخدمى الحاسب الآلى كثيراً من النصوص الطويلة فى ظل تعاظم ثورة المعلومات وشبكة الويب العالمية . ويحتاج المستخدمون إلى قراءة جميع الوثائق المتاحة لتحديد النص الأكثر ارتباطا. بالإضافة إلى ذلك فهم يفضلون النصوص القصيرة .إن تلخيص الوثائق الكبيرة يدوياً عملية صعبة تستغرق وقتاً طويلاً ولذا من الضرورى استخدام التلخيص الآلى للنص لحل هذه المشكلة. وتعتبر نظم تلخيص النص هى واحدة من أهم مجالات البحث فى الوقت الحاضر. حيث أن التلخيص الآلى ينتج نصا قصيرا مع امكانية الإحتفاظ بالمعنى العام للنص.

وهناك عديد من الأساليب المختلفة التى تهدف إلى انشاء ملخصات جيدة منسقة . ومن أهم هذه الأساليب الشائعة هى التحليل الدلالى الكامن(LSA) . وهو ما تم التركيز عليه فى هذا البحث. يوجد مئات الملايين من المنتمين والمهتمين باللغة العربية. ولكن مع كل هذا تم عمل عدد قليل من الأبحاث فى هذا المجال . ولهذا السبب فإن هذه الدر اسة تعالج الوثائق العربية الإنجليزية . وتم تقديم وتصميم وتنفيذ وتقييم نظام تلخيص للنصوص عربى / انجليزى على أساس التحليل الدلالى الكامن (LSA) والطريقة التقليدية (خصائص الجملة)

ولذلك فإن النظام المقترح يجمع نوعين من الخصائص وهى خصائص الجملة و الخصائص الدلالية. ويتكون النظام المقترح من عدة خطوات. اولاً ، تم تمثيل وثيقة الإدخال على شكل مصفوفة. ثم تطبيق

المصدرية (stemming) وحذف كلمات التوقف لتحسين وتعزيز أداء النظام. ثانياً ، تم وضع قيم لخلايا المصفوفة وفقاُ لطريقة معينة. ثم يتم تطبيق بعض الحسابات (SVD) على المصفوفة المدخلة. وأخيراً ، يتم وضع درجات للجمل المستخدمةوفقاُ الى خصائص الجملة والخصائص الدلالية. ثم بعد ذلك يتم ترتيب الجمل واختيار عدد من هذه الجمل ذات الدرجات الأعلى لانشاء الملخص.

وتم اختبار النظام المقترح على مجموعة مكونه من ثلاث و ستين وثيقة.وهذه الوثائق هى وثائق عر بية وانجليزية و ثنائية اللغة (عربى / انجليزى). وقد قمنا بعملية الاختبار باستخدام مجموعه قياسية من الوثائق المتاحة على شبكة الانترنت. و لقد تم استخدام ROUGE لتقييم النظام . حيث أن ROUGE هى طريقة لقياس جودة الملخص باستخدام ملخص يدوى. علاوة على ذلك ، تمت مقارنة نظامنا للتخيص الآلى مع سبعة من نظم التلخيص التجارية.

وهناك تحسن كبير فى نظام التلخيص الآلى الذى قام به نظامنا بالمقارنة مع أنظمة تلخيص النص التجارية المتاحة للوثائق المكتوبة باللغة الإنجليزية. ويلاحظ أيضا تحسن مماثل عند مقارنة نظامنا مع أربعة من أنظمة التلخيص المتاحة التى تدعم اللغة العربية للوثائق المكتوبة باللغة العربية وأيضا للوثائق ثنائية اللغة (عربى / إنجليزى).