

منهج التشابه الدلالي متعدد اللغات للنصوص العربية - الإنجليزية

رسالة مقدمة لكلية الحاسبات والمعلومات جامعة الفيوم ، كجزء من متطلبات الحصول على
درجة الدكتوراه فى نظم المعلومات

إعداد

إسراء أحمد عبدالوهاب رسلان

قسم نظم المعلومات - كلية الحاسبات والمعلومات
جامعة الفيوم

إشراف

د. رشا محمد بدرى
مدرس بقسم نظم المعلومات،
كلية الحاسبات والمعلومات،
جامعة الفيوم

أ.د. محمد حلمى خفاجي
أستاذ دكتور بقسم علوم الحاسب،
عميد كلية الحاسبات والمعلومات،
جامعة الفيوم

المخلص

يعتبر تشابه النص الدلالي من أهم المشاكل في مجال التنقيب عن النصوص واسترجاع المعلومات. تأتي أهمية تشابه النصوص من قدرتها على توفير المعلومات الأكثر أهمية من نص كبير حيث يركز تشابه النص على استخراج أهم المعلومات من مجموعة من المستندات النصية.

تهدف رسالة الدكتوراة إلى ما يلي : القيام بتطوير منهج لقياس التشابه الدلالي باستخدام (LSA) Latent Analysis Semantic بين نص مترجم من اللغة العربية إلى الإنجليزية ونصاً آخر باللغة الإنجليزية وتطبيق هذا المنهج في مجال الزراعة. ذلك سيساعد في البحث عن حلولاً للمشاكل الزراعية التي يقوم المزارع بكتابتها عن طريق البحث عن مشاكل شبيهة في قاعدة البيانات لاسترجاع اقرب حل يناسب هذه المشكلة و التي تساعد المزارعين على حل مشكلاتهم.

يستخدم (LSA) للعثور على السمات الدلالية الرئيسية للكلمات المستخدمة في المشاكل الزراعية. فقد قدمت هذه الرسالة نظاماً جديداً يستخدم التشابه الدلالي (LSA) علاوة على ذلك ، استخدام النظام خوارزمية Support Vector Machine مع Hadoop MapReduce لتصنيف المشاكل الزراعية على حسب نوع المحصول مثل القمح و القطن و الارز او نوع المشكلة مثل مشاكل مرضية او إنتاجية او مشاكل خاصة بالرى. تظهر النتائج التجريبية التي تم الحصول عليها أن النظام قد تفوق على الانظمة الأخرى ذات الصلة التي تم اختبارها. فقد أظهرت نتائج عملنا أن نظام التصنيف الدلالي وطرق التصفية يمكن أن تحسن بشكل كبير دقة التشابه بين النصوص. لقد تفوق نظامنا المقترح على نماذج أخرى بدقة تبلغ ٨٧٪.

وتحتوى الرسالة على سته فصول كما يلي:

الفصل الاول: مقدمة.

الفصل الثانى: الخوارزمات المستخدمه.

الفصل الثالث: الأبحاث ذات صلة.

الفصل الرابع: النظام المقترح.

الفصل الخامس: النتائج والمقارنات.

الفصل السادس: الملخص و الأبحاث المستقبلية.