

تلخيص نصوص اللغة العربية باستخدام القوالب التفاعلية

إعداد

محمد سامي علاونه

إشراف

الدكتور محمد نصار

الملخص

المعلومات عبر الإنترنت المتاحة على شبكة الإنترنت العالمية ذات حجم هائل. تم تطوير محركات البحث مثل جوجل ، وياهو لاسترداد المعلومات من قواعد البيانات. ولكن لم يتم الحصول على النتائج الفعلية لأن المعلومات الإلكترونية تتزايد يوماً بعد يوم؛ لذا جاء تلخيص النصوص التلقائي لیساعد في حل هذه المشكلة. ویلخص التلخیص التلقائي العديد من الملفات الرقمية ویوفر إصداراً مختصراً قصيراً كنتاج غني بالمعلومات، لا لیس فيه، ویوفر وقتاً قیماً. یعد تلخیص نص الوثيقة الواحدة أحد أنواع التلخیص التي لها القدرة على التمييز بین الجزء الأكثر إفادة أو الأقل أهمية من النص في مستند واحد. تهدف هذه الدراسة إلى اقتراح آلية جديدة لتلخیص النص باللغة العربية على أساس قالب ديناميكي يتم تطبيقه في نص الوثيقة الواحدة. تهدف هذه الدراسة إلى اقتراح آلية جديدة لتلخیص النصوص باللغة العربية على أساس القوالب التفاعلية يتم تطبيقه في نص الوثيقة الواحدة. تم تقييم الآلية المقترحة بتطبيقها على مجموعة نصوص من CNN Arabic corpus ثم مقارنة النتائج مع

نموذج مقترح آخر (آلية الزعبي). التقييم قام تبعاً لمقاييس تقييم مختلفة ، بما في ذلك: جودة النص ،

عدم التكرار ، وضوح المرجع ، والتماسك ، والدقة ، والاستدعاء ، والدقة.

وأشارت النتائج إلى أن الآلية المقترحة أكثر كفاءة من آلية الزعبي. بالتفصيل ، فإن الملخص

النتائج باستخدام الآلية المقترحة له مستوى جودة نص إجمالي قوي جداً (4.775 من 5) ، ومستوى

عام قوي جداً من عدم التكرار (4.84 من 5) ، ومستوى وضوح مرجعي إجمالي قوي جداً (4.775

من 5) ، مستوى التماسك الكلي قوي جداً (4.775 من 5). بالإضافة إلى ذلك ، فإنه يحقق مستوى

عالٍ جداً من (0.992) Accuracy (0.98) Precision (0.96) Recall.

Arabic Text Summarization Using Dynamic Template

Prepared by:

Mohammed Alawneh

Supervised by:

Dr. Mohammed Nasrar

Abstract

The online information available on World Wide Web is in enormous amount. Search engines like Google, Yahoo were developed to retrieve information from the databases. But actual results were not obtained as the electronic information is increasing day by day. Thus automatic summarization came into demand. Automatic summarization gathers several documents as input and provides the shorter summarized version as output which is informative, unambiguous, save valuable time. Single document text summarization is one of the summarization type that has the ability to distinguish the most informative or the less informative portion of the text in one document. This study is to propose a new mechanism for Arabic text summarization based on dynamic template to be applied in single document text. The proposed mechanism is evaluated using CNN Arabic corpus. Then, it compared with other existing mechanism (i.e. Al-Zoubi's mechanism)

based on different evaluation metrics, including: text quality, non-redundancy, reference clarity, coherence, precision, recall, accuracy.

The results indicated that proposed mechanism is more efficient than Al-Zoubi's mechanism. In detail, the summary resulted from the proposed policy has very strong overall text quality level (4.775 out of 5), very strong overall non-redundancy level (4.84 out of 5), very strong overall reference clarity level (4.775 out of 5), very strong overall coherence level (4.775 out of 5). In addition, it achieves very high level of accuracy (0.992), precision (0.98) and recall (0.96).