

الخلاصة

تعتبر عملية تلخيص النصوص من المواضيع البحثية المهمة خصوصا بعد النمو الهائل للمعلومات في السنوات الاخيرة. يمكن تحديد عملية تلخيص النصوص بأنها عبارة عن برامج تقوم بتقليص حجم النص وذلك بأزالة المعلومات المكررة من النص الاصلي. النص المخرج يكون مستخلص من النص الاصلي.

تلخيص النصوص مفيد جدا بالاشتراك مع محرك البحث على الويب وكذلك يمكن ان يستخدم لتلخيص النص قبل ان يقرأ من قبل مؤلف الكلام الاوتوماتيكي, وبذلك يقل الوقت الذي يحتاجه لتحديد الاجزاء الرئيسية من النص.

النظام المقترح لتلخيص النصوص يتألف من ثلاث مراحل تولد نصا " أقصر ولا يحوي على تكرار في المعلومات وبدقة متناهية وهي (الطرق الاحصائية – التقنيات اللغوية باستخدام معالجة اللغات الطبيعية – تعلم الحاسبة باستخدام قواعد الارتباط) كما يستخدم النظام للتكامل بين تلخيص النصوص والبيانات. تلخيص النصوص يفيد تنقيب البيانات وذلك باستخلاص البيانات المهيكلية من الوثائق النصية, والتي يمكن ان تنقب بعد ذلك باستخدام الطرق التقليدية.

النظام المقترح يقوم بمهمتين في الوقت نفسه, المهمة الاولى هي بناء قاعدة بيانات لكلمات ذات العلاقة , وذلك من خلال تلخيص اكثر من نص بالاعتماد على الطرق الاحصائية ومعالجة اللغات الطبيعية في قياس وزن كل كلمة ومن ثم تحديد اي منها سيكون بالنص الملخص وفق هذه الطرق, ومن خلال هذه النصوص سيتعلم النظام كلمات ذات العلاقة ومن ثم يستخدمها في اعادة حساب وزن كلمات النص وتحديد اهمية كل كلمة واي منها سيتم استخدامه في النص الملخص, والمهمة الثانية هي توليد النص الملخص.

تمثل مخرجات النظام المقترح نصوص مستخلصة ومحافظة على المعنى العام للنص الاصلي. عدد الجمل المستخلصة يمثل نسبة (٢٥-٤٠) % من عدد جمل النص الاصلي ومن التجارب التي طبق عليها النظام تم الحصول على نتائج جيدة تصل الى ٩٥ % من النص المستخلص بواسطة الانسان.

Abstract

Text summarization is a computer program that summarizes a text. The summarizer removes redundant information from the input text and produces a shorter non-redundant output text. The output is an extract from the original text. Automatic text summarization is extremely useful in combination with a search engine on the Web. Text summarization can also be used to summarize a text before an automatic speech synthesizer reads it; thus reducing the time needed to absorb the essential parts of a document.

The proposed system is composed of several stages to generate summary, it can be abstracted in three main disciplines that are integrated to give accurate results: Statistical methods, linguistic approaches using natural language processing and machine learning using association rule. This system used for integrating text summarization and data mining, text summarization benefits DM by extracting structured data from textual documents, which can then be mined using traditional method.

The proposed system does two tasks at the same time, the first task is to build the database of relational words by learning them from the several input texts, the relations of each word effect in its weight that the system depends on to select the token which must be in summary text, and the second task is to generate a summary text. The output of the proposed system is an extract from the original text and it saves the meaning of the original text. The number of sentences of output text is (25-40) % from the number of sentences of the original text. The proposed system experiments show good results are obtained of about 95% of text summarization compared with text summarized manually by human.