

Overview

ShutterFlex Text Mining tools provide powerful text processing capabilities to derive high quality information from structured and non-structured Arabic text content and databases. The Text Mining platform includes a server engine and SDK. Powered by ShutterFlex industry-leading Arabic natural language processing (NLP) technology, Text Mining finds patterns and trends through means such as statistical pattern learning. The Text Mining process includes structuring input text, deriving patterns within the structured data, and evaluation and interpretation of the output.

Capabilities

Keyword Extraction

Extracts keywords and expressions with special significance to the text subject. Also extracts named entities including persons' names, places, named organizations, famous expressions, and others. Keyword Extractor saves time by extracting the most important information within high volumes of content.

Summarization

Identifies the most relevant sentences within a text and presents a short text summary. Reduces reading and processing time. Summarizer extracts key significant sentences from the input text based on linguistic and statistical rules such as writing style, sentence position in the text, included keywords, and keyword types. Users are given the ability to specify a level of summarization – for example, as a percentage of the input text, or a fixed number of phrases or words.

Categorization

Accurately and efficiently organizes valuable information into a logical topic tree. The categorization tree involved in this process is made up of general categories that cover the most common areas, such as Politics, Economy, Religion, Sports, and others. The categorization engine employs ShutterFlex Corrector to automatically correct Arabic input text for common mistakes.



SDK availability

The SDK lets developers integrate Text Mining with other applications.



التصنيف

تقوم هذه الأداة بتنظيم المعلومات القيمة بدقة وكفاءة على شكل شجرة موضوعات منطقية. وتتكون هذه الشجرة من تصنيفات عامة تغطي أعم وأشمل المجالات مثل السياسة والاقتصاد والدين والرياضة وغيرها. ويوظف محرك التصنيف هذا المصحح الآلي من شاتر لتصحيح الأخطاء الشائعة الموجودة بالنص العربي المدخل تلقائياً



توافر مجموعة تطوير البرامج

للمطورين دمج أدوات (SDK) تتيح مجموعة تطوير البرامج تحليل النصوص مع التطبيقات الأخرى



نبذة عن البرنامج

توفر أدوات تحليل النصوص من شاتر فلكس إمكانيات فائقة لمعالجة النصوص لتقدم للمستخدم معلومات عالية القيمة مستقاة من قواعد بيانات ومصادر مستندات مختلفة تحتوي على نص عربي سواء أكانت هذه المصادر مهياً (رقمية) أو غير مهياً (يحتوي النظام الأساسي لأدوات تحليل النص على محرك يعمل (SDK) عن طريق خادم مع مجموعة تطوير البرامج جاءت أدوات تحليل النصوص نتاجاً للأبحاث المتواصلة، وتقوم شاتر فلكس في مجال المعالجة الآلية للغة العربية أدوات تحليل النصوص (NLP) باستخلاص الأنماط والنماذج من خلال بعض الطرق مثل دراسة النموذج الإحصائي. وتشمل عملية تحليل النصوص تهيئة النص المدخل، وتتبع الأنماط داخل البيانات المهياً ثم تقييم النتائج و تفسيرها

الإمكانيات

مستخلص الكلمات المفتاحية

يقوم باستخلاص الكلمات المفتاحية والتراكيب ذات الدلالات الخاصة المرتبطة بموضوع النص، بالإضافة إلى استخلاصه للأعلام الموجودة بالنص سواء أكانت لأشخاص أو أماكن أو مؤسسات أو عبارات شائعة... إلخ. كما يوفر مستخلص الكلمات الأساسية الوقت عن طريق استخلاص أهم المعلومات الموجودة في البيانات الضخمة.

التلخيص

تتعرف هذه الأداة على أكثر الجمل ذات الصلة داخل النص وتعرضها على شكل ملخص للنص مما يقلل بشكل كبير الوقت الذي نستغرقه في القراءة ومعالجة النصوص. ويستخرج الملخص الجمل الأساسية من النص المدخل على أساس بعض القواعد الإحصائية واللغوية مثل: أسلوب الكتابة، موضع الجملة في النص، والكلمات الأساسية المتضمنة وأنواعها، مما يعطي المستخدم القدرة على وضع حد معين للتلخيص - على سبيل المثال، يحدد نسبة معينة من النص المدخل أو عدد معين من الجمل أو الكلمات