|  |  |
| --- | --- |
| التعرف على اللغة في تحليل الوثائق (LIDA) | Research Title (Arabic) |
| Language Identification in Document  Analysis (LIDA) | Research Title (English) |
| التعرف على لغة النص العربي أو الإنجليزي | Research Topic (Arabic) |
| Language Identification of a text Arabic or English | Research Topic (English) |
| IASTED | Publisher (Arabic) |
| IASTED | Publisher (English) |
| 1424 | Publishing Year (Arabic) |
| 2004 | Publishing Year (English) |
| 0-88986-453-5 | ISBN |
| تحليل الوثائق، الذكاء الاصطناعي، تقنية الوكيل | Key Words (Arabic) |
| Document Analysis, AI, Agent Technology | Key Words (English ) |
| Proceeding of the IASTED International Conference Circuits, Signals, and Systems, Nov. 28- Dec. 1, 2004, Clear Water, Florida – USA. | Journal Name, or (Conference + place and date being held) |
| Pp 278-283 | Volume No. or Issue No. and the Number of Pages in case it has been published in a scientific journal |
| يناقش البحث عملية التمييز مابين النص العربي والإنجليزي. ويتم التعرف على مستوى الكلمة والسطر حيث تم استعراض خوارزم يقوم بعملية التمييز بين اللغتين. ويمكن استخدام الخوارزم في عدة تطبيقات منها عملية المعالجة للغة الطبيعية. | Research Abstract (Arabic) |
| This paper presents a technique that can be used to discriminate between texts written in Arabic script and texts written in Latin script. This technique addresses the language identification problem on the word level and on the text line level. This technique uses an algorithm for horizontal projection profiles. This paper presents a new algorithm of language identification to determine languages of a document. This approach may be used in identifying the language in many applications. These applications cover encoding of document pages, language specific web crawling, information retrieval, natural language processing, text mining, translation service bureau software, spell checking software, stemming or morphological analyzers, and knowledge management systems. | Research Abstract (English) |