****Cairo University

Faculty of Engineering

Electronics and Electrical Communications Engineering Dept.

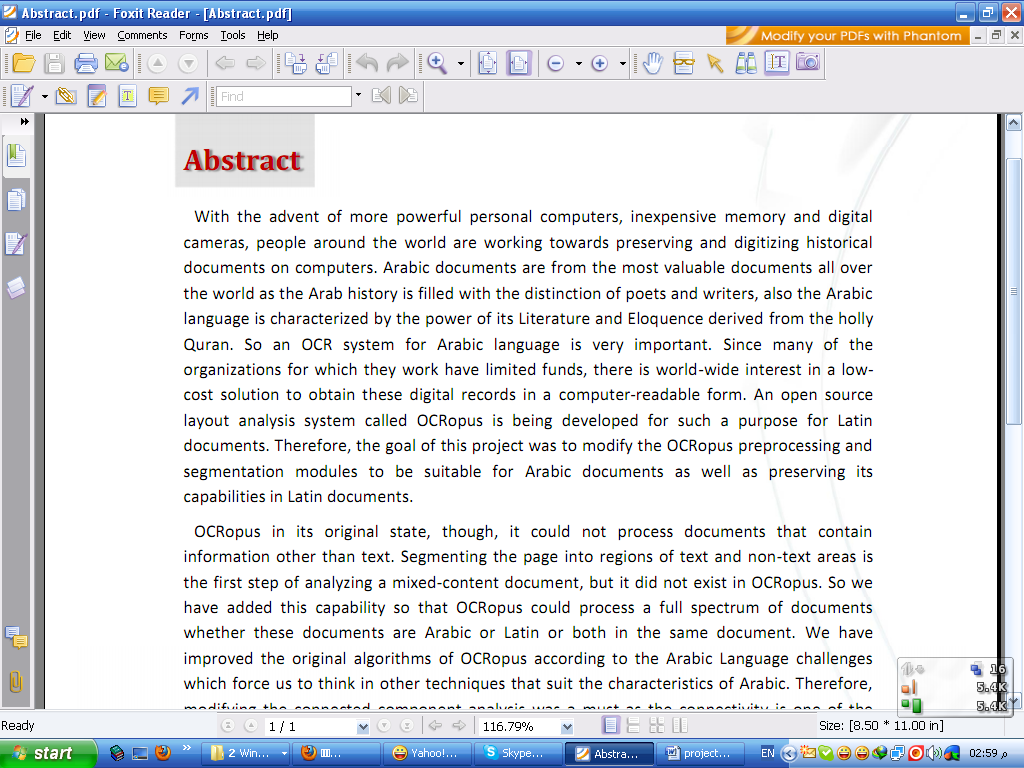
|  |
| --- |
|  |
| Optical Character Recongnition for Arabic |
|  |
| Prof. Dr. Mohsen Rashwan |

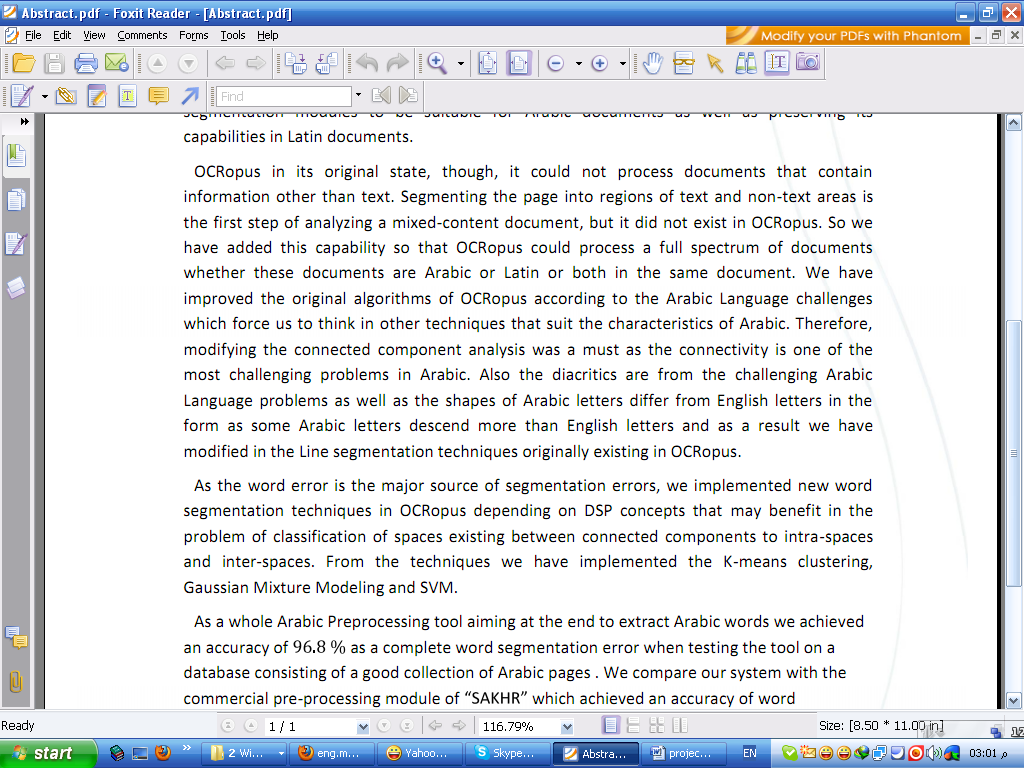
mRashwan@rdi-eg.com

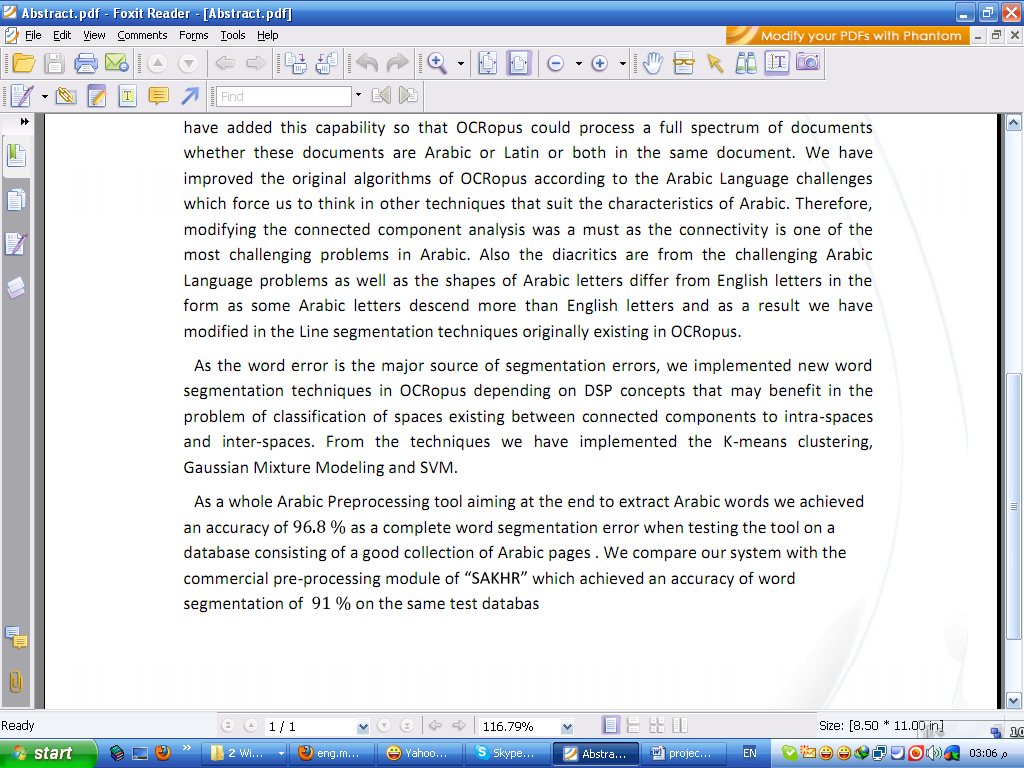
|  |  |
| --- | --- |
| Student Names | Student E-mail Addresses |
| Hossam Ahmed Nour El Deen | imprator\_hoss@hotmail.com |
| Hossam Samy Marzouk | fear\_comm@yahoo.com |
| Khaled Qorany Abd El Fadeel | alfares\_anygroup@yahoo.com |
| Ramy Ahmed Ali Mohamed | ramy\_ahmed\_ali@hotmail.com |
| Ramy Mohamed Nabil Tosson | ramy\_tosson@hotmail.com |
| Mostafa Mohamed Abd El Shafi | mostafa\_hawk44@hotmail.com |

Optical Character Recongnition for Arabic

Prof. Dr. Mohsen Rashwan

**Abstract:**

****

****

**CONCLUSION:**

In this paper we presented a document analysis system for Arabic documents. This system had shown to work well on Roman script so it was adapted to Arabic documents. The described system was tested on 125 scanned images from different sources like books, magazines, and first copy from magazines.

The algorithm achieved an accuracy of above 95% in terms of text-word detection or books and of above 96 % in magazine images and first copy images. We compared with a commercial tool ”Sakhr” on the same database that give an accuracy of 91%. Newspaper documents proved to be the toughest class presenting several challenges like many font sizes within the same image, small inter-line spacing, inverted text, and poor quality of page resulting in lot of noise.