



ساده‌ترین راه تجربه فرآیند OCR متون لاتین برای کتابداران

ابوذر رضانی دانشجوی کارشناسی ارشد
دانشگاه آزاد واحد تهران شمال
Aboozar@parsonline.net

روی منوی فایل باز شده، شما دستور زیر را انجام دهید.
Tools/Recognize Text using OCR...

بعد از چند لحظه شما دستور زیر را انجام دهید.

Tools/Send Text to Word ...

سپس در منوی که ظاهر می شود، سه حالت انتخابی که All pages شما کلیک نمایید و تنظیمات محل ذخیره را نیز تعیین نمایید، سپس روی OK را کلیک نمایید. من توصیه می کنم یک برگه اسکن شده از کتاب لاتین را تهیه نمایید و برای تمرین، مراحل بالا را انجام دهید.

شما این مراحل را می توانید بر روی فایل PDF نیز انجام دهید، دستور پرینت را از منوی فایل بدهید، در صفحه ای که ظاهر می شود، شما نام پرینتر خویش را از Microsoft Office Document Image Writer انتخاب و مراحل بالا پی در پی انجام دهید.

در بازار ایران نسخه ویژه OCR خارجی با قیمت‌های مختلف به فروش می‌رسد، یکی از معروفترین آنها، نام تجاری آن Readiris است که نسخه ویرایشی ۱۰ به بالای آن در بازار ایران موجود است.

علاوه بر نسخه‌های خارجی مهندسین ایرانی در راستای طرح تکفای اقدامات موثری انجام داده اند شما می‌توانید گزارشی از برخی تلاشها را در دو وب سایت‌های زیر مطالعه نمایید:

* <http://www.airayvarz.com/fa/persianocr.aspx>

* <http://www.farsiocr.ir/>

وقتی کتابداران جهت آشنایی با نحوه ایجاد کتابخانه دیجیتال مطالعه می کنند، اطلاعات زیادی پیرامون شیوه‌های مختلف دیجیتالی کردن اسناد از طریق اسکن و OCR آنها مطالعه خواهند کرد، حتماً علاقمند خواهند شد که یکبار دیجیتالی نمودن اسناد را تجربه نمایند در این راستا بر آن شدم تجربه خودم را در این زمینه برای متون انگلیسی به نگارش در آورم.

قبل شروع شما باید در سیستم رایانه خویش برنامه آفیس ۲۰۰۳ به طور کامل نصب نمایید و برای این کار به یک صفحه اسکن شده از مقاله یا کتاب لاتین با درجه تفکیک‌پذیری بالاتر از ۲۰۰ نیاز دارید یا در صورت نداشتن امکانات پویشرگ شما می‌توانید از یک فایل pdf که از یک مقاله تمام متن باشد، هم استفاده نمایید.

ابتدا بعد از اینکه شما یک مقاله یا بخشی از کتاب انگلیسی زبان را اسکن کردید، روی یکی از عکسها راست کلیک نمایید، از منوی Open With با انتخاب برنامه windows pictures and fax viewer تصویر اسکن شده را باز نمایید، سپس با فشار همزمان Ctrl + P دستور پرینت بدهید. منوی Photo Pringing Wizard ظاهر می شود، شما روی Next کلیک نمایید، در بخش Picture Selection برگه های اسکن شده مورد نظر خود را انتخاب می‌کنید، سپس روی Next کلیک نمایید، سپس در منوی Printing Options پرینتر خود را Office Document Image Writer Microsoft انتخاب نمایید (اگر شما ۲۰۰۳ Office را کامل نصب کرده باشید، چنین منوی ظاهراً باید وجود داشته باشد) بعد شما Next را کلیک نمایید، مجدداً Next را کلیک نمایید. بعد از آن از شما می خواهد فایل ایجاد شده را کجا ذخیره نماید، شما جای مورد نظر خود را انتخاب سپس روی save کلیک نمایید.

در پایان برای شما فایلی با فرمت Microsoft Document Imaging Format (.mdi) ذخیره می شود، که شما روی Finish کلیک نمایید.

۱. Optical Character Recognition(OCR).

۲. توجه: بعد از نصب کامل آفیس ۲۰۰۷ موجود در بازار ایران متوجه می‌شویم ظاهراً دیگر برنامه Microsoft Office Document Image Writer در سیستم رایانه نصب نمی‌شود، اما این مشکلی به وجود نخواهد آورد آفیس ۲۰۰۷ برای شما این امکان به وجود آورده که هم نسخه ۲۰۰۳ و هم نسخه ۲۰۰۷ را در سیستم خود نصب نمایید.