

راهنمای تصویری

روش تبدیل تصویر به متن (OCR) از طریق آفیس ۲۰۰۳

OCR مخفف Optical Character Recognition و یا Optical Character Reader می باشد و آن در واقع نرم افزار متن خوانی است که برای تشخیص و بازیابی الفبا و ... نوشته های دست نویس یا تایپ شده طراحی شده است.

OCR چیست؟

فرض کنید که ما متنی را روی کاغذ داریم و می خواهیم آن را وارد رایانه کنیم. اولین روشی که به ذهن می رسد این است که متن را به تایپیست بدهیم تا با کامپیوتر تایپ کند. اما آیا می شود عین همان متن را وارد رایانه بکنیم تا نیازی به تایپ نباشد؟ البته دستگاه «اسکنر» می تواند تصویری از آن متن را وارد رایانه کند، تا اینجا بخشی از مشکل ما حل شده است. اما تصویر وارد شده قابل ویرایش نیست یعنی حروف و کلمات از هم تشخیص داده نمی شوند و پردازشی روی آنها صورت نمی گیرد.

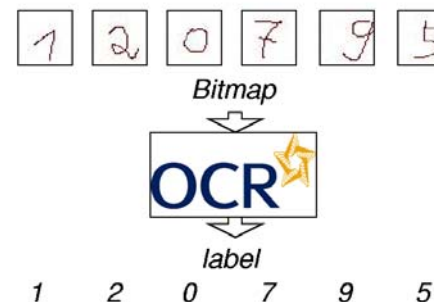
پس با Document Imaging ۲۰۰۳ می توان متن های دست نویس یا متونی را که قبلا با ماشین تحریر تایپ شده و یا نوشته های چاپ شده را به صورت خودکار به متن های کامپیوتری تبدیل نموده و در آن ذخیره کرد. یعنی تبدیل تصاویر کلمات به کدهای کلمات (قابل ذخیره در داخل کامپیوتر) و می دانیم هر متنی که در کامپیوتر ذخیره شد می توان روی آن پردازش های دلخواه انجام داد.

با توجه به اینکه تقریباً تمام کاربران از آفیس نسخه ۲۰۰۳ آشنایی دارند و از آن استفاده می کنند در این راهنما مبنای ما نرم افزار است که همراه مجموعه office 2003 نصب می شود.

قبل از استفاده توجه به نکات ذیل ضروری است:



روش تبدیل تصویر به متن (OCR) از طریق آفیس ۲۰۰۳



تهیه شده در: سمعی و بصری بیمارستان قائم^(ع)

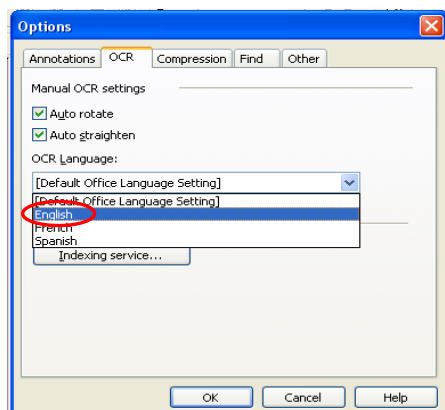
<http://www.mums.ac.ir/quuem/fa/auv>

جهت دریافت مطالب آموزشی سمعی و بصری به سایت فوق مراجعه نمایند

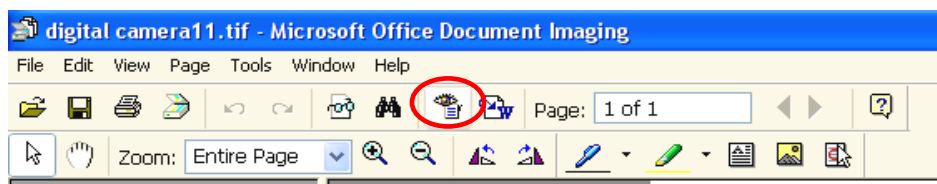
Email: eidih1@mums.ac.ir

تهیه کننده: حسن عیدی بایگی

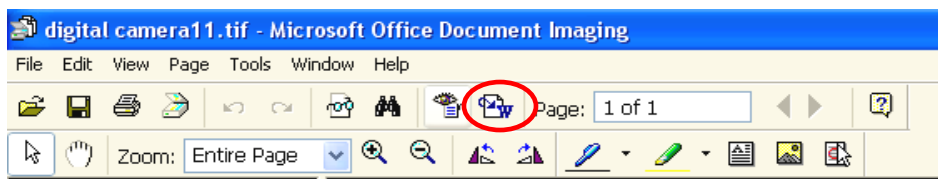
۳) تغییر زبان به انگلیسی از منوی tools گزینه options



۴) انتخاب ابزار تشخیص متن



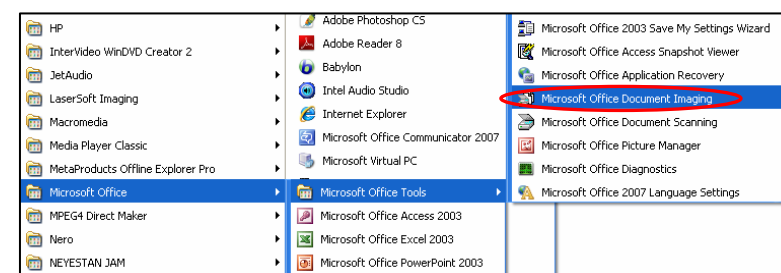
۵) ذخیره به صورت فایل Word



در این مرحله می توانیم به جای ایجاد یک فایل جدید قسمتی از متن را کپی و در یک فایل متنی دیگر برگردانیم.

- ۱- این نرم افزار و نرم افزارهای مشابه از زبان فارسی و عربی پشتیبانی نمی کنند. (به دلیل شباهت حروف فارسی و رسم الخط مشکل حروف پیوسته)
- ۲- نرم افزارهایی که به این منظور طراحی شده اند کامل نیستند و خطاهای زیادی دارند چنانکه نرم افزار OCR گسسته فارسی در ایران مدتی است که طراحی و نوشته شده و در بعضی جاها استفاده می شود. اما مدت زمان زیادی طول خواهد کشید تا نرم افزار OCR پیوسته برای خط فارسی نوشته شود.
- ۳- اسکنرهای مانند HP و CANON همراه خود نرم افزارهایی دارند که به صورت خودکار در هنگام اسکن تصاویر این کار را انجام می دهند.
- ۴- فرمت (قالب) متن در فایل word با فرمت اصلی (عکس) متفاوت است که نیاز به ویرایش دارد
- ۵- عکسهایی که توسط دوربین گرفته می شوند باید دارای فوکوس خوب، کنتراست مناسب و زمینه روشن باشند. عکس هایی که با گوشی های موبایل گرفته شده باشند کمتر دارای شرایط فوق می باشند.

۱) اجرای برنامه از منوی start



۲) بازکردن (import) عکس با فرمت tiff

جهت باز کردن عکسها از منوی file گزینه های open یا import را انتخاب کنید. تصاویر باید با فرمت tiff باشند.

جهت تبدیل فرمت عکس از نرم افزار های ویرایش عکس مثل ACDSee ، PHOTO SHOP و یا نرم افزار Picture Manager که همراه نرم افزار آفیس نصب می شود استفاده نمائید.