OCR INSTALLATION GUIDE

1. Requirements
2. Installation
3. How to Run/ Inputs-Outputs

Prepared in Dec 4th, 09
1) REQUIREMENTS
   a) Any 32 bit operating system in below:
   b) Microsoft Office 2007
   c) .Net Framework 2.0
   d) Open Xml Format Sdk 1.0
   e) Ocr.exe

2) INSTALLATION
   a) Microsoft Office 2007 Installation :
      Choose Customize:
Make sure *Microsoft Office Word* is selected as “Run from My Computer.”

Under Office Tools make sure you choose *Microsoft Office Document Imaging* as “Run from My Computer.”

Ok → Next → Next → Finish
b) **.Net Framework 2.0 : dotnetfx.exe installation**

   It can be downloaded from [http://www2.cs.uh.edu/~taytun/Install/dotnetfx.exe](http://www2.cs.uh.edu/~taytun/Install/dotnetfx.exe)

   ![Microsoft .Net Framework 2.0 Setup](image.jpg)

   Next → Next → Finish

c) **Open Xml Format Sdk 1.0 installation**

   It can be downloaded from [http://www2.cs.uh.edu/~taytun/Install/OpenXMLSDK.msi](http://www2.cs.uh.edu/~taytun/Install/OpenXMLSDK.msi)

   ![Welcome to the Open XML Format SDK 1.0 Setup Wizard](image.jpg)

   The installer will guide you through the steps required to install Open XML Format SDK 1.0 on your computer.

   WARNING: This computer program is protected by copyright law and international treaties. Unauthorized duplication or distribution of this program, or any portion of it, may result in severe civil or criminal penalties, and will be prosecuted to the maximum extent possible under the law.

   ![Cancel](image.jpg) ![Back](image.jpg) ![Next](image.jpg)
3) **HOW TO RUN / INPUTS-OUTPUTS**

Ocr.exe can be downloaded [http://www2.cs.uh.edu/~taytun/Install/Ocr.exe](http://www2.cs.uh.edu/~taytun/Install/Ocr.exe)

It can be called in command line.

Ocr.exe has a config file that can be downloaded [http://www2.cs.uh.edu/~taytun/Install/Ocr.exe.config](http://www2.cs.uh.edu/~taytun/Install/Ocr.exe.config)

```xml
<configuration>
  <appSettings>
    <add key="InputFolder" value="F:\OCRLAST\TestVideos\Index_qtj\txtFiles" />
    <add key="TempFolder" value="F:\OCRLAST\TestVideos\Index_qtj\tempFolder" />
    <add key="XmlFolder" value="F:\OCRLAST\TestVideos\Index_qtj\xmlFolder" />
    <add key="LogFolder" value="F:\OCRLAST\TestVideos\Index_qtj\logFolder" />
    <add key="SleepTime" value="10000" />
  </appSettings>
</configuration>
```
“Ocr.exe “ will create some preprocessed images in the tempFolder and 2 xml files named as inputFolder read in txt file.

If the folder to be processed “ c:\pictures\DrSubhlok01” then
drsubhlok01_TP.xml file will be created for Transitions point.
http://www2.cs.uh.edu/~taytun/Install/drsubhlok01_TP.xml

drsubhlok01_IP.xml file will be created for Indexs point.
http://www2.cs.uh.edu/~taytun/Install/drsubhlok01_IP.xml

Transition points will include all transition points. Here is 73

------------------------------------
<?xml version="1.0" encoding="utf-8" ?>
<files>
 <InputFolder>f:\ocrlast\testvideos\index_qtj\drsubhlok01</InputFolder>
 <IndexCount>73</IndexCount>

 <indexFile>
   <id>1</id>
   <name>5.jpg</name>
   <text>1i4, 2004, 3rd, computer, 4ici, 4j1, adapted, add, after, and, any, approach, are, ask, available, booki, bown, can, chapter, cjt, classroom, content, copynght, copyriit, datalink, decks, delete, disconnected, edition, enjoy, faculty, featurf, file, fir, flak, flé, fll, following, for m, freely, from, help, ians, ibstantially, identical, including, instructor, internet, james, july, kei, iihw, keith, kurose, lans, layer, like, lir, liy, lot, lrsr, making, material, material, mention, modify, need, networking, nflu, note, obviously, of, on, one, only, our, part, people, perhaps, post, powerpoint, ppt, presenter, qk, readers, represent, reserved, return, rights, role, ross, site, slides, source, start, students, substantially, tcflic, thanks, that, their, these, they, this, tk, tools, top, tworkü, unaltered, use, view, wed, were, www, y1, you, your, </text>
 </indexFile>
 <indexFile>
   <id>2</id>......
--------------------------------------

Index points will include all transition points but merged in indexed points. Here is 10 indexed points:

<?xml version="1.0" encoding="utf-8" ?>
<files>
 <InputFolder>f:\ocrlast\testvideos\index_qtj\drsubhlok01</InputFolder>
 <IndexCount>10</IndexCount>

 <indexFile>
   <id>1</id>
   <name>207.jpg</name>
   <text>106, 1996, 1l4, 1o6, 1t, lv4ti, 2004, 3fr, 3multiple, 3rdedthono, 3rdedthong, 3x, 4?, 41445, 4526f 2a7f3, 4c, 4computer, 4ici, 4ides, 4j1, 4lfr4, 4lfr4, 41k4, 4s?, 4tr, 5ir, 5ri, 74fl, 802, 9ink, access, adapted, addison, addressing, administrator, after, agent, aidir2, ait, algorithm, along, analogy, any, approach, apter5, aqk, are, ask, atj, atm, available, bata, behind, booki, bown, broadcast, bt, byr, c4, can, channels, chapter, cjh, cjtj, classroom, clt, cna, communication, conferencexp, connect, content, context, control, copynght, copyriit, correction, current, classroom, datagram, datalink, datagram, db, dcn, decks, delete, detection, different, disconnected, dkv, done, dufl, elfl, each, edition, encapsulates, enjoy, erfr, erfv, error, ethernet, faculty, featurf, file, first, flak, flar, flatly, flé, fll, flow, flsfl, following, form, frame, freely, frg, from, frsflkw, fx, geneva, goals, has, help
Every actions/errors will be stored in OcrLog.txt file which will stored in logFolder user defined in config.file.