

[4:51:24 PM] \*\*\* European Weekend Testers added Ajay Balamurugadas \*\*\*  
[4:51:54 PM] \*\*\* European Weekend Testers added Markus Gärtner, Michael Larsen, thomas.ponnet \*\*\*  
[4:53:27 PM] European Weekend Testers: Hi  
[4:53:35 PM] European Weekend Testers: Can you see this group?  
[4:54:18 PM] Ajay Balamurugadas: Hi All  
[4:54:19 PM] \*\*\* Ajay Balamurugadas added Balu \*\*\*  
[4:54:27 PM] Michael Larsen: I can :)  
[4:54:34 PM] European Weekend Testers: cool  
[4:55:22 PM] Ajay Balamurugadas: Hey Michael, Balu, Markus..  
[4:55:34 PM] European Weekend Testers: Hi all  
[4:55:38 PM] Michael Larsen: Evening, Ajay :).  
[4:55:42 PM] European Weekend Testers: Do we need to add someone else?  
[4:55:52 PM] \*\*\* Ajay Balamurugadas added madhusudhan \*\*\*  
[4:56:08 PM] Balu: Hello everyone  
[4:56:14 PM] \*\*\* Markus Gärtner has changed the conversation topic to "EWT39" \*\*\*  
[4:56:20 PM] Markus Gärtner: hi  
[4:56:24 PM] Michael Larsen: Anna was on the earlier list. Will she be attending today?  
[4:56:45 PM] European Weekend Testers: She wasn't sure. Let's add her to be sure :)  
[4:56:55 PM] \*\*\* European Weekend Testers added annabaik \*\*\*  
[4:57:57 PM] European Weekend Testers: Ok guys. We should start introducing ourselves  
[4:58:37 PM] \*\*\* Ajay Balamurugadas added Narasimha Reddy \*\*\*  
[4:59:09 PM] Markus Gärtner: Markus Gärtner, software tester, consultant, black belt in Miagi-Do, writing a book  
[4:59:38 PM] Markus Gärtner: blog: [shino.de/blog](http://shino.de/blog), twitter: @mgaertne  
[5:00:05 PM] Ajay Balamurugadas: Ajay Balamurugadas, based in Bangalore, blog: [www.enjoytesting.blogspot.com](http://www.enjoytesting.blogspot.com) twitter: @ajay184f  
excited to be attending EWT after a long time.  
[5:00:23 PM] Balu: Balaji, software tester from Hyderabad, India, twitter: @balajiponnada blog: <http://balajiponnada.wordpress.com/>  
[5:01:23 PM] Narasimha Reddy: Narasimha Reddy Pathagunta ,based in Chennai,India Twitter @narasimhareddy First time joining Euro weekend testing  
[5:01:25 PM] Michael Larsen: copy -> paste  
[5:01:27 PM] Michael Larsen: I'm Michael Larsen. software tester in San Francisco, CA. brown belt in the Miagi-Do School of Software Testing, podcast producer for the "This Week in Software Testing" Podcast, co-founder of the Americas Chapter of Weekend Testing, and instructor with the Association for Software Testing's Black Box Software Testing series of classes. I blog at <http://mkl-testhead.blogspot.com/> and can be found on Twitter at @mkltesthead and on Skype as "mkltesthead".  
[5:02:05 PM] European Weekend Testers: Ok guys, so here is the mission  
[5:02:07 PM] European Weekend Testers: Compare the following image to text converters:

<http://www.free-ocr.com/>

<http://www.onlineocr.net/default.aspx>

<https://docs.google.com/DocAction?action=updoc&hl=en>

by:

1)file size capability

2)speed of conversion

3)quality of conversion (defined by the tester)

4)find other aspects that can be a relevant metric and point out the observations

[5:03:17 PM] Ajay Balamurugadas: INTERESTING...

[5:03:28 PM] Ajay Balamurugadas: sorry caps lock was on by mistake

[5:03:44 PM] Michael Larsen: OK, before we go scurrying off :).

[5:04:11 PM] Ajay Balamurugadas: 1 hr testing?

[5:04:29 PM] European Weekend Testers: we test for 1h then the next hour is for discussion

[5:04:34 PM] Ajay Balamurugadas: ok

[5:04:55 PM] Michael Larsen: So our charter is to compare and contrast the different services.

[5:05:06 PM] European Weekend Testers: yes

[5:05:16 PM] Ajay Balamurugadas: any test report needed?

[5:05:18 PM] Michael Larsen: Do we have a section for bug reporting? Should we use the chat to do so, such as an #issue hashtag?

[5:05:35 PM] European Weekend Testers: but also give reasons why you consider one better than the other

[5:05:48 PM] European Weekend Testers: I think we should use hash

[5:05:54 PM] European Weekend Testers: because its easier

[5:06:10 PM] Michael Larsen: OK, will do.

[5:06:21 PM] European Weekend Testers: and not too many look to the separate bugs

[5:06:56 PM] Michael Larsen: Small group today, should we all have a go at it here? each of us pick one and then rotate after about 15 minutes?

[5:07:26 PM] Ajay Balamurugadas: would be good that way, somehow reminded of Dr. Bono's hats here...

[5:07:29 PM | Edited 5:08:06 PM] European Weekend Testers: well we need to compare, so each should try at least once all of the converters

[5:08:06 PM] Michael Larsen: OK, I'll take on <http://www.free-ocr.com/> this first bit.

[5:08:36 PM] Ajay Balamurugadas: oh already 8 mins over, I will take the last one - Google

[5:09:02 PM] Balu: Are we supposed to take one at a time?

[5:09:11 PM] Narasimha Reddy: i will take <http://www.onlineocr.net/default.aspx>

[5:09:11 PM] European Weekend Testers: we need to compare somehow

[5:09:22 PM] European Weekend Testers: so its your choice

[5:09:56 PM] Michael Larsen: I'd suggest spening about 15 minutes with each, so that you can get a good view of each approach and quality.

[5:10:10 PM] \*\*\* European Weekend Testers added vamshi \*\*\*

[5:10:39 PM] Markus Gärtner: just in case: abcdefghijklmnopqrstuvwxyzäöüß

[5:11:18 PM] Balu: ok @Michael

[5:11:45 PM] vamshi: Hi all

[5:11:54 PM] Michael Larsen: Hello Vamshi.

[5:12:26 PM] vamshi: what exactly is ocr?  
[5:12:48 PM] Michael Larsen: Optical Character Recognition.  
[5:13:02 PM] Ajay Balamurugadas: recognizing text from an image...  
[5:13:03 PM] Michael Larsen: But in this case, we're using online services in stead of an optical scanner  
[5:13:15 PM] Michael Larsen: or yeah, what Ajay said ;).  
[5:13:19 PM] Ajay Balamurugadas: a software which will scan the image and give us the text in the image  
[5:13:20 PM] \*\*\* European Weekend Testers added eusebiu blindu \*\*\*  
[5:13:29 PM] \*\*\* eusebiu blindu can't be added until they accept your contact request. \*\*\*  
[5:14:05 PM] vamshi: ok  
[5:14:36 PM] Michael Larsen: #issue, PDF weith table causes confuision for character recognition in Free-ocr.  
[5:15:10 PM] \*\*\* European Weekend Testers added eusebiu blindu \*\*\*  
[5:15:15 PM] Michael Larsen: Thought... would it make sense to include samples of the files we tried to convert?  
[5:15:45 PM] eusebiu blindu: sure!  
[5:16:05 PM | Edited 5:16:15 PM] Michael Larsen: OK, I'll send them to you :).  
[5:16:40 PM] \*\*\* Michael Larsen sent MeritBadges.pdf \*\*\*  
[5:16:57 PM] Ajay Balamurugadas: see you all in 40 mins...  
[5:18:09 PM] Michael Larsen: Based on the file I uploaded, here is the text that was displayed:  
[5:18:20 PM] Michael Larsen: SAN FRANCISCO BAY AREA COUNCIL BOY SCOUTS OF AMERICA  
MERIT BADGE INFORMATION  
be able to pull back a 25 lb bow; \$5 fee covers cost of arrow and bowstring kits  
a blue card; Will require observation time outs1de of merit badge session  
Will be done on an individual basis onl if all re uirements are done.  
Communit cam \*g must be at least l4 ears old and a Star or Life  
Star or Life  
or older and a Star or Life  
budget; Req. 8 not covered at can1p\*; must be l4 or older and a Star or Life  
Pottery 4 . . . \_ . . . \_ .  
meeting after flung, sess1onl1m1t 20, Pre S1gn~  
--  
18  
[5:18:36 PM] Michael Larsen: That was it.  
[5:18:36 PM] European Weekend Testers: what converter was that?  
[5:18:44 PM] Michael Larsen: Free-ocr.  
[5:19:08 PM] Michael Larsen: So it does convert the image text, but it has limitations.  
[5:19:48 PM] Michael Larsen: It starts with the first section of the document, captures the header and the first line, but the table format throws it off.  
[5:20:13 PM] Michael Larsen: I'll try a more linear document for a comparison.  
[5:20:42 PM] European Weekend Testers: well is actually made for text in particular  
[5:21:05 PM] European Weekend Testers: images and tables will be a great impediment in

conversion

[5:21:55 PM] Michael Larsen: #issue free-ocr. Captcha failure, have to reload the document again.

[5:22:19 PM] Michael Larsen: Sure, may be out of scope (re tables) but I figured I'd mention it ;).

[5:23:47 PM] Ajay Balamurugadas: Who is the customer here? EWT 39?

[5:24:08 PM] Ajay Balamurugadas: Any samples you have already tested with?

[5:24:20 PM] Ajay Balamurugadas: Any specific kinds of files you want me to test with?

[5:24:22 PM] Michael Larsen: OK, with free-ocr, I uploaded a document in a checklist format. this one had no trouble getting all the text. It converted the squares into the tex characters 'EI', which I thought was interesting.

[5:25:05 PM] Ajay Balamurugadas: Is that ok if I test based on my assumptions & state my assumptions (regarding files, browsers, network) & then let you know in a report?

[5:25:09 PM] Markus Gärtner: are we discussing results here already?

[5:25:22 PM] Markus Gärtner: just wondering :)

[5:25:45 PM] Ajay Balamurugadas: Stating my assumptions & moving forward :)

[5:26:20 PM] Michael Larsen: Just reporting what I'm finding so that others may be able to compare.

[5:27:20 PM | Edited 5:27:31 PM] Michael Larsen: Follow-up. I used a file in a table format but with no lines drawn (a phone list). though it separated everything out into its constituent columns and printed them as one list, it did get all of the characters in the file.

[5:27:24 PM] European Weekend Testers: that is actually a good idea - for limited time we have

[5:28:21 PM] Michael Larsen: OK, I think I have a clear enough view of open-ocr. I'm going to move over to <http://www.onlineocr.net/default.aspx>

[5:28:35 PM] Ajay Balamurugadas: [5:27 PM] European Weekend Testers:

<<< that is actually a good idea Which one?

[5:29:22 PM] Michael Larsen: I think EWT means my popping my findings in on the main chat. consider it a session report in real time (LOL!).

[5:29:30 PM] European Weekend Testers: @Ajay [5:26:20 PM] Michael Larsen: Just reporting what I'm finding so that others may be able to compare.

[5:29:49 PM] European Weekend Testers: [5:26 PM] Michael Larsen:

<<< Just reporting what I'm finding so that others may be able to compare.that i though is a good idea

[5:30:00 PM | Edited 5:30:14 PM] Michael Larsen: If it's distracting, though, I can do it in a separate file and load it during debrief.

[5:31:42 PM] eusebiu blindu: I got "Googlg" in Google docs for google logo

[5:32:03 PM] vamshi: what type of text result we should expect after uploading the image?

[5:32:35 PM] European Weekend Testers: the text withing the image should be as accurate as possible

[5:33:02 PM] European Weekend Testers: As a realistic example I will think of a book page that needs to be scanned then transformed into text

[5:33:11 PM] Michael Larsen: #comment, can anyone who has used <http://www.onlineocr.net/default.aspx> get the recognize button to respond? I'm using IE8 and it's not reacting.

[5:33:22 PM] Michael Larsen: Maybe the page doesn't play well with IE8.

[5:33:32 PM] Michael Larsen: Trying Firefox 4.

[5:34:15 PM] Michael Larsen: Whoops, nevermind.

[5:34:27 PM] European Weekend Testers: Did you know ((\*)) That captcha is used to actually use humans to convert book text

That's why you have 2 words usually and one doesn't matter when you enter it usually

[5:34:45 PM] Michael Larsen: Different rules. Expected I'd be able to convert a PDF, but this one requires an image file.

[5:35:30 PM | Edited 5:35:42 PM] Michael Larsen: Chalk that up to "Consistent with other products" or "tester didn't do a very close read" (LOL!).

[5:35:33 PM] European Weekend Testers: Google Docs does all for example

[5:35:48 PM] European Weekend Testers: but the focus was for images

[5:36:00 PM] European Weekend Testers: any good observation is welcomed though

[5:38:21 PM] Narasimha Reddy: google docs does not convert numbers

[5:38:31 PM] vamshi: online ocr didn't returned text while freeonline ocr does for .jpeg image

[5:38:59 PM] Narasimha Reddy: sorry converting

[5:41:48 PM] Michael Larsen: #comment OnlineOCR.net was able to take a JPEG of my original table PDF, and when I submitted it, it was able to read the text, but the results were returned in a jumbled format.

[5:42:13 PM] Michael Larsen: Also, many of the words were misspelled (not uncommon behavior with OCR software).

[5:45:06 PM] \*\*\* European Weekend Testers added David O'Dowd \*\*\*

[5:45:34 PM] Michael Larsen: the image file I created from my second document (checklist) contained all of the data and fewer misspelled words (OnlineOCR.net).

[5:49:31 PM] Markus Gärtner: how much time do we have left?

[5:49:44 PM] Michael Larsen: 11 minutes according to my clock :).

[5:49:45 PM] European Weekend Testers: 10 min

[5:49:52 PM] Michael Larsen: or ten ;)

[5:49:53 PM | Edited 5:50:01 PM] European Weekend Testers: +/-100 :)

[5:51:11 PM] Michael Larsen: #comment OnlineOCR converted my table format document without lines (phone list) without issue, and did all of the sections as a straight line with data. No tabs or new lines, but again, that is within expected behavior of OCR software I have used.

[5:51:20 PM] Michael Larsen: no drops or misspellings.

[5:55:37 PM] Michael Larsen: #issue, google tool, uploading pdf file to convert, tried to create a presentation file, opening presentation file stops midway, never completes loading (seen on IE8)

[5:55:38 PM] European Weekend Testers: 5 min

[5:57:29 PM] Michael Larsen: Interesting. When I did the same test with Google Chrome it handled it just fine ;).

[5:58:21 PM] David O'Dowd: #OnlineOCR.net I gave OnlineOCR.net it's own Captcha image to attempt to translate the image. Gave me this as a result %&2sk"ii Which was obviously not correct. Might be nice if it could give a confidence level for each letter it translates.

[5:59:37 PM] Michael Larsen: Actually, Google Chrome browser did a clean transfer of all my PDF files to google doc format, including the table with lines.

[6:01:43 PM] European Weekend Testers: Ok guys, its time for debriefing!

[6:01:58 PM] Michael Larsen: all right, I think I'm done.  
[6:02:20 PM] European Weekend Testers: How many had the chance of actually uploading and checking each of the three OCRs?  
[6:02:20 PM] Markus Gärtner: I uploaded my results to <http://www.shino.de/images/EWT39.png>  
[6:03:03 PM] Michael Larsen: I did.  
[6:03:50 PM] Markus Gärtner: I did all three, but just had success with on converting my image  
[6:04:03 PM] European Weekend Testers: Ok so Michael and Markus..anyone else?  
[6:04:20 PM] Balu: I did, but not all tests  
[6:04:45 PM] Balu: didn't upload all different types of files to all three OCR tools  
[6:04:50 PM] Narasimha Reddy: i did on JPEG, PDF files  
[6:05:21 PM] Markus Gärtner: jpg, bmp and pdf here  
[6:05:31 PM] Michael Larsen: I did the same. I limited my testing to PDF and JPEG files, but used the same three files consistently.  
[6:05:51 PM] vamshi: I did on .jpeg and pdf file. Didn't try it on google docs  
[6:06:12 PM | Edited 6:06:15 PM] Michael Larsen: Umm, i guess that shouldbe six files consistently :)  
[6:06:26 PM] European Weekend Testers: Anyone tested the size limit ? I see Markus has in this results  
[6:07:01 PM] Balu: I didn't but i found the size limitations from the sites of three tools  
[6:07:33 PM] Ajay Balamurugadas: [6:02 PM] European Weekend Testers:

<<< How many had the chance of actually uploading and checking each of the three OCRs? I did

[6:07:50 PM] Michael Larsen: Did not on my end.  
[6:07:54 PM] Ajay Balamurugadas: [6:05 PM] Michael Larsen:

<<< I did the same. I limited my testing to PDF and JPEG files, but used the same three files consistently.Me too. :)

[6:08:10 PM] European Weekend Testers: I didn't but i found the size limitations from the sites of three toolsSo you mean you trusted the claims on those websites, not really trying?  
[6:08:11 PM] Ajay Balamurugadas: but I used PNG (2 files) & one BMP file  
[6:08:58 PM] Balu: @European - I couldn-t do it within the time i have.  
[6:09:18 PM] Michael Larsen: Had I more time I would have then repeated my tests with sizes of files that would push the boundaries. as it was, I was using significant file variation that I uncovered whast I felt was unusual behavior.  
[6:09:23 PM] Michael Larsen: there's always more :).  
[6:09:48 PM] Balu: It took quite some time to understand the functionality for me. I couldn't test much.  
[6:10:06 PM] Ajay Balamurugadas: You tested Balu , learning is also included in testing :)  
[6:10:39 PM] Ajay Balamurugadas: and the 'info that it took some time to learn' also reveals good information about the products.  
[6:10:58 PM | Edited 6:11:19 PM] Michael Larsen: True, every time we explore the product, we get a better understanding of how it should work, or how we feel it should work. that's relevant test data.

[6:11:34 PM] Balu: reading about three sites and how they work is what actually took time for me

[6:11:50 PM] Balu: read about, FAQs section

[6:12:07 PM] Balu: and learnt about limitations. Result - couldn't test much :)

[6:12:40 PM] European Weekend Testers: I am curious if someone used more Google Docs in report with the other tools or less, and why?

[6:12:53 PM] Michael Larsen: @balu So that is an approach. You were looking to see what the claims of the product were, so that you could be prepared to test.

[6:13:17 PM] Ajay Balamurugadas: @Michael, I prepared lots of test files :|

[6:13:32 PM] Michael Larsen: @EWT, my take on the google docs was that, a jpg was rendered as a jpg, and a pdf was rendered as a PDF.

[6:13:54 PM] Michael Larsen: In other words, I'm not entirely sure that Google docs is doing an real OCR here.

[6:14:15 PM | Edited 6:14:26 PM] European Weekend Testers: @Michael yeah but in the bottom it should, if you choose the right option during upload

[6:14:20 PM] Ajay Balamurugadas: I was left wondering for few mins, where my images are in GDocs & what abt OCR

[6:14:45 PM] Michael Larsen: @Ajay, yep, I spent some time preparing docs as well, including the "problem child" documents that I picked.

[6:15:17 PM] European Weekend Testers: "Convert text from PDF or image files to Google Docs documents." it should be checked at upload

[6:15:31 PM] Balu: @Ajay - that is what confused me too. But, i've found text

[6:16:16 PM] Ajay Balamurugadas: In GDocs, the file was uploaded & then, below the file , the text was displayed

[6:16:22 PM] Michael Larsen: @EWT, yeah, I did that, but the returned format was my PDF file (at least in Chrome, it crashed my IE8 browser.)

[6:16:31 PM] Michael Larsen: Convert documents, presentations, spreadsheets, and drawings to the corresponding Google Docs formats

Convert text from PDF or image files to Google Docs documents. How does this work?

[6:16:40 PM] European Weekend Testers: google docs does not coverting numbersWas this claim proven right?

[6:16:52 PM] Ajay Balamurugadas: not checked.

[6:16:57 PM] Markus Gärtner: google docs didn't find any text in the picture I uploaded

[6:17:18 PM] Ajay Balamurugadas: was there text in image @Markus? ;)

[6:17:20 PM] Markus Gärtner: I prepared one picture with text in printed form as well as hand drawn by mouse

[6:17:28 PM] Ajay Balamurugadas: ok

[6:17:31 PM] Markus Gärtner: the only engine that recognized it was the first one

[6:17:46 PM] Ajay Balamurugadas: GDocs recommends Arial or Times New Roman

[6:17:54 PM] Markus Gärtner: the drawback of the first converter was, that it would convert just the first page from a pdf as well

[6:17:57 PM] Narasimha Reddy: hand drawn images does not recognizing 3 tools

[6:18:10 PM] European Weekend Testers: @Markus can you give us your sample that didn't work in GoogleDocs but was working in other OCR?

[6:18:15 PM] Markus Gärtner: the first one recognized the text I had printed

[6:18:26 PM] Markus Gärtner: I had one image with both, printed and hand-drawn letters

[6:18:41 PM] Markus Gärtner: just one succeeded at converting it

[6:18:45 PM] Michael Larsen: My examples were to take documents that might be found in a normal course of work (a table with boundaries, a check list and a phone list, which is a table but without hard lines.

[6:19:06 PM] Ajay Balamurugadas: image files...

[6:19:53 PM] Ajay Balamurugadas: is different from what you mentioned Michael?

[6:19:56 PM] Michael Larsen: OK, so i went back and rechecked the option with google docs, and yes, it does convert jpg images to text.

[6:20:12 PM] Ajay Balamurugadas: cool

[6:20:25 PM | Edited 6:20:46 PM] Michael Larsen: @ajay, the how is this different was part of the text that I copied and pasted, sorry that that was confusing.

[6:20:48 PM] eusebiu blindu: anyone tried rotating text at certain degrees?

[6:20:59 PM] Ajay Balamurugadas: Here comes Sebi... too late :)

[6:21:08 PM] eusebiu blindu: wondering if its still to handle it

[6:21:22 PM] Markus Gärtner: I had text and image mixed up with google docs, and it didn't convert anything there at all

[6:21:33 PM] Markus Gärtner: it just created a word document with my image in it

[6:21:41 PM] Ajay Balamurugadas: Free Online OCR rotated my image by 180 degrees even though i never asked it to

[6:21:48 PM] Markus Gärtner: rotating is a nice idea

[6:21:53 PM] Markus Gärtner: I didn't consider that one

[6:22:04 PM] European Weekend Testers: cool

[6:22:10 PM] Markus Gärtner: but I was disappointed with the conversion already enough to not dare trying that one :)

[6:22:31 PM] Michael Larsen: OK, I checked the table document with hard lines on google docs again, and this time it had the same problem as the other OCR program did, in that it jumped up the sections and order of things.

[6:22:43 PM] Markus Gärtner: I found it interesting to find the same webpage title for the first two

[6:22:49 PM] Markus Gärtner: at least the beginning of it

[6:23:23 PM] European Weekend Testers: Language support was an issue at any point? Anyone tried to do different language? I wonder if Japanese or Mandarin was working fine

[6:23:28 PM | Edited 6:24:15 PM] Michael Larsen: again, I will admit that the table example could be argued as out of scope, but I could equally argue it's a valid document that I'd want to convert for various reasons, and the inability to at least have the ordering be consistent could be seen as a valid pain point.

[6:23:32 PM] Ajay Balamurugadas: so many options

[6:23:37 PM] Ajay Balamurugadas: to test

[6:23:39 PM] Ajay Balamurugadas: ...

[6:23:45 PM] Ajay Balamurugadas: \Good Mission :)

[6:24:11 PM] European Weekend Testers: the number of OCR was maybe too high

[6:24:24 PM] European Weekend Testers: and one would have shown more test ideas

[6:24:58 PM] Markus Gärtner: oh, I found it perfect, as just one of them worked for me at all!

[6:24:59 PM] Markus Gärtner: :)

[6:25:04 PM] vamshi: The image should be again uploaded in freeonline ocr to generate text in other downloadable format

[6:25:07 PM] Michael Larsen: Perhaps, but the compare and contrast is a good tool as well. It allows us to perform A/B (or a/b/c) testing and see if similar tests would yield similar results.

[6:25:12 PM] Michael Larsen: We proved that it doesn't.

[6:25:13 PM] Ajay Balamurugadas: @Michael, i am wondering how a physical scanner would scan the table doc

[6:25:40 PM] Michael Larsen: Truth be told Ajay, I don't know.

[6:25:51 PM] Markus Gärtner: did the second one work for anyone? Was there a special trick? I didn't get any result, just a security code

[6:26:07 PM] Ajay Balamurugadas: do testers ever lie? (out of scope)

[6:26:13 PM] Michael Larsen: I was able to get results from all three programs.

[6:26:35 PM] European Weekend Testers: I was thinking that comparing the tools will not be very relevant from the perspective that Google usually tries to improve constantly its products. Thus eventually, Google's will be anyway the best to use in a real necessity

[6:26:37 PM] Michael Larsen: Yes, testers lie all the time. well, hang on, let me rephrase that.

[6:26:40 PM] Ajay Balamurugadas: [6:26 PM] Michael Larsen:

<<< I was able to get results from all three programs. Don't ask about the quality though

[6:26:42 PM] Balu: @Markus, you have to upload the file first and then click 'Recognize'

[6:26:49 PM] Michael Larsen: testers tell the truth as they understand it if they are honorable testers.

[6:27:04 PM] Ajay Balamurugadas: enter captcha & then hit recognize?

[6:27:21 PM] Balu: @Ajay- true , i missed captch

[6:27:29 PM] Michael Larsen: I'll leave out the option of a dishonorable tester who would deliberately fudge or manipulate data.

[6:27:41 PM] Markus Gärtner: I though "recognize" was to select the language

[6:27:50 PM] vamshi: [6:27 PM] Ajay Balamurugadas:

<<< enter captcha & then hit recognize?this got more options from the first one

[6:27:53 PM | Edited 6:28:05 PM] Michael Larsen: even with a "truthful and honorable tester", we don't always have the whole story, or we have gaps in our knowledge or perception.

[6:27:59 PM] Ajay Balamurugadas: which was easy to learn/use?

[6:28:07 PM] Markus Gärtner: free ocr

[6:28:11 PM] eusebiu blindu: that is an interesting philosophical point: if people are unaware about a subject but act as its true, is it the same as to lie?

[6:28:12 PM] vamshi: freeocr is easy to use

[6:28:20 PM] Balu: [6:28 PM] Ajay Balamurugadas:

<<< which was easy to learn/use?free OCR

[6:28:26 PM] Michael Larsen: I think the FreeOCR was the most straightforward.

[6:29:14 PM] Balu: you cannot blame google docs as it is not an OCR tool.

[6:29:17 PM] Narasimha Reddy: free OCR better , but number 6 recognized as 5

[6:29:38 PM] Michael Larsen: the google docs looked cooler (and had different interface if used in Chrome than in Firefox 4 or IE8, may be leveraging HTML5 to display items differently).

[6:29:43 PM] eusebiu blindu: I think there are some opensource code source for OCRs and anyone can try to improve it

[6:29:49 PM] Balu: [6:29 PM] Narasimha Reddy:

<<< free OCR better , but number 6 recognized as 5Where did you get the image from? is that a screenshot?

[6:30:03 PM] eusebiu blindu: Anyway based on this mission next time we can try the porn filter! :)

[6:30:14 PM] Michael Larsen: But there were more details to have to be aware of to make behave properly (Google Docs).

[6:30:16 PM] eusebiu blindu: that would attract people!

[6:30:17 PM] Narasimha Reddy: free OCR for pdf files first page only covering

[6:30:41 PM] Michael Larsen: Narasimha, I noticed the same thing (thanks for the reminder).

[6:30:46 PM] Balu: [6:30 PM] Narasimha Reddy:

<<< free OCR for pdf files first page only coveringthat's a limitation of free OCR mentioned in FAQs section :)

[6:30:54 PM | Edited 6:31:29 PM] Michael Larsen: #issue, disregard ;)

[6:31:10 PM] Ajay Balamurugadas: why is it an issue?

[6:31:16 PM] Narasimha Reddy: @Balu , i did not read thanks for pointing

[6:31:31 PM] Ajay Balamurugadas: it says 1st page only, scanned 1st page only

[6:31:39 PM] Ajay Balamurugadas: what's the issue here?

[6:32:04 PM] Michael Larsen: It's not.

[6:32:08 PM] Narasimha Reddy: @Balu i have created image in mspaint

[6:32:12 PM] Markus Gärtner: it's mentioned in the FAQs? I looked there, but didn't see that

[6:32:30 PM | Edited 6:32:35 PM] Michael Larsen: the site explains that one page only is the expected behavior, and it looks to follow through. I didn't read that.

[6:33:06 PM] Balu: [6:32 PM] Markus Gärtner:

<<< it's mentioned in the FAQs? I looked there, but didn't see thatlast FAQ

[6:33:28 PM] Michael Larsen: However, compared to other sites, if they allowed more than one page, I think we could make a case that it's not consistent with other products.

[6:33:32 PM] European Weekend Testers: there were 3 products, each with its own FAQ

[6:34:11 PM] Balu: [6:32 PM] Narasimha Reddy:

<<< @Balu i have created image in mspaintFree OCR requires image resolution to be 150 dpi or higher. refer question 7 in FAQs section

[6:34:57 PM] European Weekend Testers: @Balu How do we know that an image resolution has 150dpi or higher?

[6:35:21 PM] Balu:

[6:33 PM] European Weekend Testers:

<<< there were 3 products, each with its own FAQ@European - I was referring to Free OCR FAQ's section

[6:35:43 PM] Michael Larsen: We can import an image into a graphical application (IrfanView, for instance) and then check via the resample option (if it's an image file).

[6:37:31 PM] Michael Larsen: Granted, taht will tell you the resolution, adjusting it to a higher resolution may not net you much if the original file is much lower.

[6:37:32 PM] Balu: [6:34 PM] European Weekend Testers:

<<< @Balu How do we know that an image resolution has 150dpi or higher?you can see it in the image properties

[6:37:39 PM] Ajay Balamurugadas: mspaint - open file and do a CTRL + E

[6:37:58 PM] Markus Gärtner: there is no mspaint on my mac :)

[6:38:13 PM] Michael Larsen: Always more than one way to do things :).

[6:38:13 PM] Markus Gärtner: but there is an image information dialogue in gimp as well

[6:38:20 PM] Balu: @Ajay - Useful shortcut to find dpi :)

[6:38:27 PM | Edited 6:38:40 PM] Michael Larsen: Markus, check to see if your imaging tool allows you to resize/resample

[6:38:32 PM] Narasimha Reddy: @Ajay thanks

[6:38:37 PM] European Weekend Testers: Would anyone use more any on these OCRs after the session as a personal need or consider it?

[6:38:49 PM] Ajay Balamurugadas: hey future sessions, online resolution (viewer/modifier tool :)

[6:39:13 PM] Michael Larsen: Looking at the google docs tool, it looks interesting/promising.

[6:39:37 PM] Michael Larsen: But I have found one area that's problematic with all three , and that's the table with borders issue.

[6:39:51 PM] Ajay Balamurugadas: i don't see the immediate use of these tools but when I need OCR tools, i know which one to try first... :)

[6:39:54 PM] Markus Gärtner: I would give it some more time

[6:40:02 PM] Markus Gärtner: maybe try one in one or two years again

[6:40:08 PM] Markus Gärtner: (don't nail me on that)

[6:40:15 PM] European Weekend Testers: I would give it some more time maybe try one in one or two years again

(don't nail me on that)

[6:40:16 PM] European Weekend Testers: lol

[6:40:25 PM] Michael Larsen: this is a personal bias, but I actualy have a lot of files that are in this format. It would be great to copy and use the tet for other purposes, but the random placement of the text is a turnm off in all of them.

[6:40:27 PM] European Weekend Testers: That is a distinct possibility

[6:40:36 PM] Ajay Balamurugadas: I feel, 3-4 hrs on each would give me confidence...

[6:40:52 PM] Balu: @European - Where can we actually use OCR tools to aid testing?

[6:41:43 PM] European Weekend Testers: @Balu I have no idea. But we can test OCR tools to aid testing :)

[6:41:51 PM] Balu: :)

[6:41:58 PM] Michael Larsen: I think it could be helpful when it comes to document sitings that haven't been fully formatted to the web. I could se it being valuable to help import faxed items, or forms that have data we want to retain ina cleaner method.

[6:42:07 PM] Michael Larsen: OCR may be one way to deal with some of that.

[6:42:21 PM] European Weekend Testers: OCR are not accurate and never will be in certain contexts.

[6:42:28 PM] Michael Larsen: Or to deal with older correspondence, letters, notes, etc.

[6:42:53 PM] Michael Larsen: For me, though, the biggest challenges with OCR are those related to handwriting, and have been for the last two decades.

[6:43:26 PM] Ajay Balamurugadas: [6:41 PM] Balu:

<<< Where can we actually use OCR tools to aid testingGive a tester one of these tools to see how he tests. :)

[6:43:34 PM] Michael Larsen: Now \*that\* would be a valuable tool if anyone ever figures out how to relably and with the least amount of pain handle diaparate handwriting styles.

[6:44:15 PM] Michael Larsen: I consider it more of a value for ubiquitous capture than for actual testing, unless I'm actually testing OCR software, of course ;).

[6:44:40 PM] eusebiu blindu: I am wondering if Google uses in the search

[6:45:19 PM] eusebiu blindu: like you google some text and the results are also from parsing images with ocr

[6:45:35 PM | Edited 6:45:44 PM] Michael Larsen: @sebui, interesting question... do you mean can their scanning technology allow them to view image documents and pull out actual words to aid in their search queries?

[6:45:43 PM] eusebiu blindu: yes

[6:46:28 PM] Michael Larsen: I don't know, but considering some of ther esults I've seen come back in searches, where the documents have not been able to select text when opened, I think it may well be the case.

[6:46:38 PM] Michael Larsen: Who knows a Googlist we can ask ;)?

[6:46:43 PM] eusebiu blindu: using an application its like a book, it gives ideas and wonder what others are doing :)

[6:47:01 PM] European Weekend Testers: ok guys, last 15 min

[6:47:52 PM] European Weekend Testers: Can anyone give its summarised ideas about the session today?

[6:48:12 PM] Ajay Balamurugadas: [6:45 PM] eusebiu blindu:

<<< like you google some text and the results are also from parsing images with ocrWhat a great point!

[6:49:07 PM] Michael Larsen: We tested three different systems, each with the intended goal of allowing the user to upload documents and have the testx read and presented in a way that could be modified and edited. Each of the three tols had its own features and limitations.

[6:49:30 PM] Michael Larsen: straight text files with minimal formatting had the best chance of success in all three applications.

[6:50:26 PM] Michael Larsen: More formatted documents had greather challenges in having accurate information cross over. tables with hard defined borders overall performed the most

poorly on all three systems.

[6:50:56 PM | Edited 6:51:02 PM] Michael Larsen: The variety of document types was broad, though the 2nd option only allowed for image files.

[6:51:09 PM] Michael Larsen: how's that :)?

[6:51:23 PM] European Weekend Testers: @Balu Do you have a resume for this session?

[6:51:32 PM] European Weekend Testers: @Michael great!

[6:53:00 PM] Ajay Balamurugadas: [http://dl.dropbox.com/u/8987774/EWT39\\_Ajay\\_TestReport.docx](http://dl.dropbox.com/u/8987774/EWT39_Ajay_TestReport.docx)

With the limited tests, I recommend (order of preference - high to low) Google docs, Free OCR and then Free Online OCR

[6:53:51 PM] Balu: @European - I've found a message when using free OCR that i've completed 10 uploads and it wouldn;t allow any more for an Hour. any one else observed this?

[6:54:14 PM] Markus Gärtner: I didn't find google docs that nice, and would rank it even behind free online ocr

[6:54:30 PM] Ajay Balamurugadas: Clearing cache, new instance, incognito window - does these change the limit?

[6:54:33 PM | Edited 6:54:53 PM] Michael Larsen: @balu, yes, the first two both state that there is a limit to the number of diocuments that can be uploaded in an hour. You just verified one of the claims ;).

[6:55:10 PM] European Weekend Testers: @Ajay nice report!

[6:55:23 PM] Balu: @Ajay- yes, when cleared cache it worked again.

[6:55:24 PM] European Weekend Testers: @Balu nice catch!

[6:55:53 PM] Michael Larsen: @Balu, that's cool. feature or bug ;)?

[6:55:56 PM] European Weekend Testers: ok we have 5 more minutes

[6:56:09 PM] Ajay Balamurugadas: feature for testers ;)

[6:56:14 PM] Ajay Balamurugadas: bug for users

[6:56:19 PM] Balu: :)

[6:57:06 PM] European Weekend Testers: @madhusudhan still there?

[6:57:20 PM] European Weekend Testers: @Narasimha Reddy?

[6:57:27 PM] European Weekend Testers: David?

[6:57:29 PM] Narasimha Reddy: yes,

[6:57:42 PM] European Weekend Testers: what are your thoughts ?

[6:58:14 PM] Balu: @ European - I dont have a consolidated report for this session as i couldn't test much of the three OCR tools. I took more time on learning. but within what i've done, its Online OCR, Free OCR and then google docs. I don;t consider Google docs as OCR tool though.

[6:59:36 PM] Ajay Balamurugadas: And for those who don't know who is operating under Europe Weekend Testers account, its Eusebiu Blindu (Sebi) .

[6:59:38 PM] Ajay Balamurugadas: :)

[6:59:54 PM] Narasimha Reddy: Free OCR tools some what better than Google

[7:00:11 PM] Narasimha Reddy: for Image conversion

[7:00:17 PM] eusebiu blindu: thanks for help today! :)

[7:00:22 PM] Michael Larsen: In the closing minutes, I'll say that I appreciated these group approaches to ideas and methods. these are very helpful to a "Lone tester" like myself. It always

gives me better ideas to help approach testing and develop ideas I wouldn't consider on my own. For OCR or any product :).

[7:00:29 PM] Markus Gärtner: thanks for the nice session, sebi

[7:00:43 PM] Michael Larsen: Yes, thanks Sebi, great job :).

[7:00:58 PM] Markus Gärtner: was thought provoking how to come up with some more thoughts on rotating text

[7:01:01 PM] Narasimha Reddy: Thanks @Europ testers

[7:01:09 PM] eusebiu blindu: Thanks! I learned some new stuff today

[7:01:12 PM | Edited 7:01:50 PM] Michael Larsen: So this begs the question, will we be seeing more EWT Saturday sessions again?

[7:01:30 PM] Ajay Balamurugadas: Thanks Sebi, great job :)

[7:01:54 PM] Narasimha Reddy: Thanks everyone, great learning

[7:02:15 PM] Balu: Apart from different approaches what we learn from WT, i'm learning new tools every week :)

Thanks WT :))!! Thanks Ajay! Thanks Sebi for facilitating :)

[7:02:26 PM] Ajay Balamurugadas: [7:01 PM] Michael Larsen:

<<< will we be seeing more EWT Saturday sessions again? Yes, Yes, Yes :) Sebi is the new facilitator :)

[7:02:29 PM] European Weekend Testers: Thanks all for attending!

[7:02:55 PM] Michael Larsen: Awesome! We should then compare notes and times so that EWT and WTA doesn't step on each others toes :).

[7:02:57 PM] European Weekend Testers: We will try to find more and more interesting apps

[7:03:35 PM] Michael Larsen: Again, thanks for the great session, I enjoyed it. Looking forward to future sessions.

[7:03:43 PM] Michael Larsen: Blog coming shortly, too :).

[7:04:21 PM] Ajay Balamurugadas: WTA was on sunday?

[7:04:58 PM] European Weekend Testers: Thanks all!

[7:05:28 PM] Michael Larsen: Ajay, WTA is on Sunday's for you :).

[7:05:36 PM] Ajay Balamurugadas: oh ok

[7:05:38 PM] Ajay Balamurugadas: :)

[7:07:13 PM] Michael Larsen: If we had a WTA session today, it would be starting in about an hour.

[7:07:32 PM] Michael Larsen: So we would do best to not schedule them on the same days, I'd think.

[7:07:47 PM] vamshi: what's the time to start WTA?

[7:07:51 PM] Michael Larsen: But that's another discussion, not to be cluttering up this session (LOL!).

[7:07:58 PM] European Weekend Testers: we thought at some point to do it Sunday

[7:08:30 PM] Ajay Balamurugadas: so, Michael & Sebi, you know whom to contact :) [contact each other ;)]

[7:08:44 PM] Michael Larsen: Yep :).

[7:09:22 PM] Michael Larsen: @Vamshi, we tend to do them about every two to three weeks, Saturday at 11:00 AM PST, 2:00 PM EST.

[7:09:28 PM] Ajay Balamurugadas: ok, good night, over & out....

[7:09:58 PM | Edited 7:10:26 PM] Michael Larsen: Night Ajay. Have a great weekend everyone.

It's 10:00 AM at my house and my kids

are still asleep (LOL!).

[7:10:06 PM] European Weekend Testers: Bye all!

[7:10:57 PM] vamshi: ok. Thans for the session. Have a nice sunday. Over and out