Kerala Congress Ad Texts

Suhaas Ramani

10/27/2020

library(tesseract)

## Warning: package 'tesseract' was built under R version 4.0.3

if(is.na(match("hin", tesseract\_info()$available))) tesseract\_download("hin")  
hin <- tesseract("hin")  
eng <- tesseract("eng")  
  
#Sample set of ads with Hindi text  
tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/168565304095241.jpg", engine = hin)

## [1] "<U+0939><U+0947> <U+0915><U+093E>\n<U+092A><U+0902> <U+091B><U+093E> <U+096B>\n<U+0939> \"3 2”!\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/243650359807244.jpg", engine = hin)

## [1] "| <U+0932><U+0915> <U+0939><U+0941> <U+0917><U+094D><U+0930><U+093E><U+092E> <U+092A><U+0902><U+091A><U+093E><U+092F><U+0924><U+094B><U+0902> <U+0915><U+094B> <U+091C><U+094B><U+0921><U+093C><U+0947><U+0902><U+0917><U+0947>\n'<U+0939><U+0938><U+094D><U+0906> <U+0939><U+093E><U+0930><U+094D><U+0921> <U+0938><U+094D><U+092A><U+0940><U+0921> <U+0913><U+092A><U+094D><U+091F><U+0940><U+0915><U+0932> <U+092B><U+093E><U+0921><U+092C><U+0930>\n<U+0927><U+0915><U+094D><U+0915><U+093E> <U+0928><U+0947><U+091F><U+0935><U+0930><U+094D><U+0915> <U+0938><U+0947>\n<U+0915><U+093E><U+092E> <U+0930><U+0941><U+0915><U+0947> <U+0928><U+093E>, <U+0926><U+0947><U+0936> <U+091D><U+0941><U+0915><U+0947> <U+0928><U+093E>\n<U+092A><U+094D><U+091F><U+093E><U+0903>\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/250661195779041.jpg", engine = hin)

## [1] "<U+0915><U+0940>\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/274408200106679.jpg", engine = hin)

## [1] "<U+0915><U+0940>\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/275754296646100.jpg", engine = hin)

## [1] "<U+0967> 8 |\n:<U+0939>\* <U+092E><U+093E><U+0928><U+0935><U+092F><U+0941><U+0915><U+094D><U+0924> <U+0917><U+0917><U+0928> <U+092F><U+093E><U+0928>\n<U+0938><U+093E> <U+092E><U+093E> <U+096D>9\n<U+092E><U+094B><U+0926><U+0940> <U+0938><U+0930><U+0915><U+093E><U+0930>\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/277223726498810.jpg", engine = hin)

## [1] "<U+094D><U+091F><U+0947>\n\n<U+0965>) > <U+0942>\n\n| <U+0915><U+093F>\n<U+0939><U+0940>]\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/282135935834941.jpg", engine = hin)

## [1] "<U+0915><U+0915> <U+0939><U+0948>\n<U+0915><U+094D><U+0937><U+0940>\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/283644765908600.jpg", engine = hin)

## [1] "#” 0 <U+0917><U+093E>\n|| <U+0964> <U+0939><U+0948> | <U+0922><U+093E><U+0902><U+091A><U+093E> <U+0915><U+094D><U+0937><U+0947><U+0924><U+094D><U+0930><U+094B><U+0902> <U+092E><U+0947><U+0902>\n<U+0938><U+094D><U+0915><U+093E> 00 <U+0932><U+093E><U+0916> <U+0915><U+0930><U+094B><U+0921><U+093C>\n<U+0966> <U+0915><U+093E> <U+0928><U+093F><U+0935><U+0947><U+0936>\n<U+0915><U+093E><U+092E> <U+0930><U+0941><U+0915><U+0947> <U+0928><U+093E>, <U+0926><U+0947><U+0936> <U+091D><U+0941><U+0915><U+0947> <U+0928><U+093E>\n<U+0928>\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/284471749166874.jpg", engine = hin)

## [1] "<U+092C><U+0939> <U+0938><U+094D> <U+0939><U+093E>\n<U+091F><U+094D>\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/308203143189449.jpg", engine = hin)

## [1] "<U+0915><U+0947>)\n<U+0964> <U+0936><U+094D><U+0903> <U+0939><U+0948> <U+092C><U+0941><U+0928><U+093F><U+092F><U+093E><U+0926><U+0940> <U+0922><U+093E><U+0902><U+091A><U+093E> <U+0915><U+094D><U+0937><U+0947><U+0924><U+094D><U+0930><U+094B><U+0902> <U+092E><U+0947><U+0902>\n<U+0915><U+0932> 00 <U+0932><U+093E><U+0916> <U+0915><U+0930><U+094B><U+0921><U+093C>\n. <U+0915><U+093E> <U+0928><U+093F><U+0935><U+0947><U+0936>\n7 <U+0926><U+0947><U+0936> <U+091D><U+0941><U+0915><U+0947> <U+0928><U+093E>\n<U+0965> <U+092E><U+094B><U+0926><U+0940> <U+0938><U+0930><U+0915><U+093E><U+0930>,\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/315507612455689.jpg", engine = hin)

## [1] "<U+0915><U+0940>\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/316595619009593.jpg", engine = hin)

## [1] "<U+092C><U+094D><U+0915><U+094D><U+093F><U+093F> <U+0967>40 <U+0967>8 |\n<U+0932><U+0915> <U+0964> <U+0917><U+094D><U+0930><U+093E><U+092E> <U+092A><U+0902><U+091A><U+093E><U+092F><U+0924><U+094B><U+0902> <U+0915><U+094B> <U+091C><U+094B><U+0921><U+093C><U+0947><U+0902><U+0917><U+0947>\n<U+0938><U+094D><U+0915><U+0964> <U+0939><U+093E><U+0908> <U+0938><U+094D><U+092A><U+0940><U+0921> <U+0913><U+0937><U+094D><U+091F><U+0940><U+0915><U+0932> <U+092B><U+093E><U+0921><U+093C><U+092C><U+0930>\n<U+0936><U+0915> 5:22: <U+0965>\n<U+0915><U+093E><U+092E> <U+0930><U+0941><U+0915><U+0947> <U+0928><U+093E>, <U+0926><U+0947><U+0936> <U+091D><U+0941><U+0915><U+0947> <U+0928><U+093E>\n<U+0924><U+091F>\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/316942372322661.jpg", engine = hin)

## [1] "<U+0964> 8 ..<U+0964> 2022 <U+0924><U+0915> <U+0905><U+0902><U+0924><U+0930><U+093F><U+0915><U+094D><U+0937> <U+092E><U+0947><U+0902> <U+092A><U+0939><U+0941><U+0902><U+091A><U+093E><U+090F><U+0902><U+0917><U+0947>\n| <U+0939><U+093F> \*«. <U+092E><U+093E><U+0928><U+0935><U+092F><U+0941><U+0915><U+094D><U+0924> <U+0917><U+0917><U+0928> <U+092F><U+093E><U+0928>\n/ <U+0932><U+093F> <U+0939><U+0948>, 2 <U+091C><U+093E>\n<U+092E><U+094B><U+0926><U+0940> <U+0938><U+0930><U+0915><U+093E><U+0930>\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/317379038942236.jpg", engine = hin)

## [1] "<U+0927><U+094D><U+091C><U+094D><U+091F><U+0932> <U+0939><U+0930>\n<U+0917><U+093E>\n(# <U+0964> '\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/326994164631380.jpg", engine = hin)

## [1] "<U+094D><U+091C><U+094D><U+091C><U+0928><U+094D><U+0924><U+094B> <U+0903>\n<U+0915><U+091F> <U+0965>\n(<U+096C> 8 <U+0939><U+0947>\n<U+0902>\n0 |\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/332488430953531.jpg", engine = hin)

## [1] "<U+093E> | <U+0948> [<U+092E><U+093E>\n"

#Sample ads with english text   
tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/248245869443408.jpg", engine = eng)

## [1] "a To \" |\n' a y eu PS :\nWea =| we\n‘Vet\na ) = ve ii ro | eg.\n‘ Ye. b t Sy\na .S\nHE WORKS STAND WITH HIM TO\nFOR THEM EMPOWER OUR YOUTH\nLolo) NTN GaN Te) oes te) eto) am ee Ati\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/252837355594021.jpg", engine = eng)

## [1] "> = oP Poe eae a\nca . ae\nean) f ; # \\\n® ie i \\\n= =| ®, ~\n~ = ke ~ =\nWey s\n( a Le i y\nZs | ae © eo\nHE WORKS ‘STAND WITH HIM TO.\nFOR THEM ERADICATE POVERTY\nDONATE NOW TO SUPPORT NAMO AGAIN IN 2019.\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/259586904965872.jpg", engine = eng)

## [1] "Se ee So ey\n\nRoa a aS aia\npote Pale algo = oe ae w\nAa ee\nOD ee eet\nee eee fe HE WORKS\nms by C xt FOR THEM\n\nCe By et\nRee) Wee es\n\nee a>\n\nei >»\n\nes >> a4 S Genus te\nos F A > NAMO AGAIN IN 2019\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/263307341255301.jpg", engine = eng)

## [1] "ey @\naa i\nPR» i.\nMOL = x 7 | HE WoRKS\n7 s Mane PY y\nPONT FOR THEM\n; ‘ee\nh . -\nNa ‘ ae pel\nVe ae\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/291975194812961.jpg", engine = eng)

## [1] "> = oP Poe eae a\nca . ae\nean) f ; # \\\n® ie i \\\n= =| ®, ~\n~ = ke ~ =\nWey s\n( a Le i y\nZs | ae © eo\nHE WORKS ‘STAND WITH HIM TO.\nFOR THEM ERADICATE POVERTY\nDONATE NOW TO SUPPORT NAMO AGAIN IN 2019.\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/300621067316452.jpg", engine = eng)

## [1] "~e f\n= ay\nnit eo %, v ze bs b fe 2\n4 a: ? Wl\nE. Joe Re axe!\n“ea \\i ON SEY\nVE: we\nC — A . eZ\n\\s oh in iit, =< 4 ZA\n= Nae\nHE WORKS\nFOR THEM\nDONATE NOW TO SUPPORT NAMO AGAIN IN 2019. 0\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/303723603829773.jpg", engine = eng)

## [1] "2. wer we SS...\nder itt) aie\nSOR PN iy ies gs Se\nby te Dee er ree ee\n§ SiS SE SD >; Or Se,\nRN NE SS Ys\n» ee ad ) Val te\nper a PON ima 9 A a a\na Ney eee\nRG PSA a. gtr Pg, re RY 3 ~\ni ie ae Ee Or,\nChemie yy MON = Re ieee\neRe i ee Se\narts See ee 2\nc ~S XN TE scone »- Es\nHE WORKS | stan witnHim to\nFOR THEM | SECURE INDIA\nDON Ua MACUL um OLN PED\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/308393213154426.jpg", engine = eng)

## [1] "etcatalalee e (\nWty yea eel \_ Ds\n\nry N\n\nro pS inh | HE WORKS\n\nvig G3 og FOR THEM\n) i\n\n“Lage \\ a\n\nJ as\n\nob. = Perennial\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/308557383146512.jpg", engine = eng)

## [1] "Sa >. =\na? EA Rod\n~ — = ml ny\nY = (5 . By \_ \\\n\* ae aS) =\nod re 4,\na\nHE WORKS\nFOR THEM\nENE MOR ee te o\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/316391559016515.jpg", engine = eng)

## [1] "2. wer we SS...\nder itt) aie\nSOR PN iy ies gs Se\nby te Dee er ree ee\n§ SiS SE SD >; Or Se,\nRN NE SS Ys\n» ee ad ) Val te\nper a PON ima 9 A a a\na Ney eee\nRG PSA a. gtr Pg, re RY 3 ~\ni ie ae Ee Or,\nChemie yy MON = Re ieee\neRe i ee Se\narts See ee 2\nc ~S XN TE scone »- Es\nHE WORKS | stan witnHim to\nFOR THEM | SECURE INDIA\nDON Ua MACUL um OLN PED\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/317264939138170.jpg", engine = eng)

## [1] "ee oS a ~\npat eee ben Were HE WorKS\neee © eee, FOR THEM\n‘aoe yee\nHae SW Ae\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/319174032278450.jpg", engine = eng)

## [1] "Se ee So ey\n\nRoa a aS aia\npote Pale algo = oe ae w\nAa ee\nOD ee eet\nee eee fe HE WORKS\nms by C xt FOR THEM\n\nCe By et\nRee) Wee es\n\nee a>\n\nei >»\n\nes >> a4 S Genus te\nos F A > NAMO AGAIN IN 2019\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/320897145439050.jpg", engine = eng)

## [1] "Se ee So ey\n\nRoa a aS aia\npote Pale algo = oe ae w\nAa ee\nOD ee eet\nee eee fe HE WORKS\nms by C xt FOR THEM\n\nCe By et\nRee) Wee es\n\nee a>\n\nei >»\n\nes >> a4 S Genus te\nos F A > NAMO AGAIN IN 2019\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/327450167915330.jpg", engine = eng)

## [1] "a 7 ) )\na “aie b Aw\nEC 4 AKS @ Na me >\n—\_— ; Ax en ‘Way te\ni ee y cs\nie ‘oe ow Sf\n~ wh in fo aN\ni \_—\_-\nHE WORKS\nFOR THEM qa\n"

tesseract::ocr("C:/Users/Suhaas Ramani/Desktop/Kerala Text Ads/333318100860253.jpg", engine = eng)

## [1] "em aie be os\nbere. &.\n) aif SS / [ y\n“eu 4\na i a, . ~ a)\nPOR THEM, | SPSWen Sut SSdonrens\nFOR THEM\nre Na etek Ue fe) ae CCL ai\n"