AIDR-MicroMappers Image Clicker Injector

1. Overview





2. Image Parsing Constrains

* UI is out of scope.
* MicroFilters should generate max.1500 records per file. If more than 1500 records exist, the system should generate separate files.
* Accepted image extensions are .jpg, .png, .jpeg
* Image search only apply for twitpic & instragram : see <https://github.com/MicroMappers/MicroFilters/blob/master/core/tasks.py>
* Duplicate image urls should be removed
* Should expect up to 40-60mb files to process

3. File input & Output

* AIDR will generate file and, send to url. The format will be .json and, it has consistent structure. See the sample;

{"filter\_level":"low","retweeted":false,"in\_reply\_to\_screen\_name":null,"possibly\_sensitive":false,"truncated":false,"lang":"fr","in\_reply\_to\_status\_id\_str":null,"id":588734861592154112,"extended\_entities":{"media":[{"sizes":{"thumb":{"w":150,"resize":"crop","h":150},"small":{"w":340,"resize":"fit","h":204},"large":{"w":1000,"resize":"fit","h":600},"medium":{"w":600,"resize":"fit","h":360}},"id":588646652363153408,"media\_url\_https":"https://pbs.twimg.com/media/CCtLAvRUIAAp5TB.jpg","media\_url":"http://pbs.twimg.com/media/CCtLAvRUIAAp5TB.jpg","expanded\_url":"http://twitter.com/Telerama/status/588646653294354432/photo/1","source\_status\_id\_str":"588646653294354432","indices":[139,140],"source\_status\_id":588646653294354432,"id\_str":"588646652363153408","type":"photo","display\_url":"pic.twitter.com/tneFkPjX9M","url":"http://t.co/tneFkPjX9M"}]},"in\_reply\_to\_user\_id\_str":null,"timestamp\_ms":"1429200303825","in\_reply\_to\_status\_id":null,"created\_at":"Thu Apr 16 16:05:03 +0000 2015","favorite\_count":0,"place":null,"coordinates":null,"retweeted\_status":{"filter\_level":"low","retweeted":false,"in\_reply\_to\_screen\_name":null,"possibly\_sensitive":false,"truncated":false,"lang":"fr","in\_reply\_to\_status\_id\_str":null,"id":588646653294354432,"extended\_entities":{"media":[{"sizes":{"thumb":{"w":150,"resize":"crop","h":150},"small":{"w":340,"resize":"fit","h":204},"large":{"w":1000,"resize":"fit","h":600},"medium":{"w":600,"resize":"fit","h":360}},"id":588646652363153408,"media\_url\_https":"https://pbs.twimg.com/media/CCtLAvRUIAAp5TB.jpg","media\_url":"http://pbs.twimg.com/media/CCtLAvRUIAAp5TB.jpg","expanded\_url":"http://twitter.com/Telerama/status/588646653294354432/photo/1","indices":[118,140],"id\_str":"588646652363153408","type":"photo","display\_url":"pic.twitter.com/tneFkPjX9M","url":"http://t.co/tneFkPjX9M"}]},"in\_reply\_to\_user\_id\_str":null,"in\_reply\_to\_status\_id":null,"created\_at":"Thu Apr 16 10:14:33 +0000 2015","favorite\_count":13,"place":null,"coordinates":null,"text":"#Cannes2015 MERCI ! Et d\u00E9sol\u00E9 pour cette avalanche de tweets. Retrouvez toute la s\u00E9lection ici http://t.co/5QfCELXFHr http://t.co/tneFkPjX9M","contributors":null,"geo":null,"entities":{"trends":[],"hashtags":[{"text":"Cannes2015","indices":[0,11]}],"symbols":[],"media":[{"sizes":{"thumb":{"w":150,"resize":"crop","h":150},"small":{"w":340,"resize":"fit","h":204},"large":{"w":1000,"resize":"fit","h":600},"medium":{"w":600,"resize":"fit","h":360}},"id":588646652363153408,"media\_url\_https":"https://pbs.twimg.com/media/CCtLAvRUIAAp5TB.jpg","media\_url":"http://pbs.twimg.com/media/CCtLAvRUIAAp5TB.jpg","expanded\_url":"http://twitter.com/Telerama/status/588646653294354432/photo/1","indices":[118,140],"id\_str":"588646652363153408","type":"photo","display\_url":"pic.twitter.com/tneFkPjX9M","url":"http://t.co/tneFkPjX9M"}],"urls":[{"expanded\_url":"http://po.st/l0L8MW","indices":[95,117],"display\_url":"po.st/l0L8MW","url":"http://t.co/5QfCELXFHr"}],"user\_mentions":[]},"source":"<a href=\"https://about.twitter.com/products/tweetdeck\" rel=\"nofollow\">TweetDeck<\/a>","favorited":false,"in\_reply\_to\_user\_id":null,"retweet\_count":33,"id\_str":"588646653294354432","user":{"location":"Paris, France","default\_profile":false,"profile\_background\_tile":true,"statuses\_count":25929,"lang":"fr","profile\_link\_color":"009999","id":67278158,"following":null,"protected":false,"favourites\_count":80,"profile\_text\_color":"333333","contributors\_enabled":false,"description":"Twitter officiel de T\u00E9l\u00E9rama. Retrouvez-nous \u00E9galement sur Facebook (https://www.facebook.com/Telerama) Et aussi @TeleramaSortir","verified":true,"name":"T\u00E9l\u00E9rama","profile\_sidebar\_border\_color":"FFFFFF","profile\_background\_color":"FFFFFF","created\_at":"Thu Aug 20 10:06:08 +0000 2009","default\_profile\_image":false,"followers\_count":162440,"geo\_enabled":false,"profile\_image\_url\_https":"https://pbs.twimg.com/profile\_images/1966519362/Logo\_Telerama\_normal.jpg","profile\_background\_image\_url":"http://pbs.twimg.com/profile\_background\_images/736846160/9d08fa7f713839d77ff53f1d3172108d.jpeg","profile\_background\_image\_url\_https":"https://pbs.twimg.com/profile\_background\_images/736846160/9d08fa7f713839d77ff53f1d3172108d.jpeg","follow\_request\_sent":null,"url":"http://www.telerama.fr","utc\_offset":7200,"time\_zone":"Paris","notifications":null,"profile\_use\_background\_image":true,"friends\_count":1978,"profile\_sidebar\_fill\_color":"EFEFEF","screen\_name":"Telerama","id\_str":"67278158","profile\_image\_url":"http://pbs.twimg.com/profile\_images/1966519362/Logo\_Telerama\_normal.jpg","listed\_count":3002,"is\_translator":false}},"contributors":null,"text":"RT @Telerama: #Cannes2015 MERCI ! Et d\u00E9sol\u00E9 pour cette avalanche de tweets. Retrouvez toute la s\u00E9lection ici http://t.co/5QfCELXFHr http://\u2026","geo":null,"entities":{"trends":[],"hashtags":[{"text":"Cannes2015","indices":[14,25]}],"symbols":[],"media":[{"sizes":{"thumb":{"w":150,"resize":"crop","h":150},"small":{"w":340,"resize":"fit","h":204},"large":{"w":1000,"resize":"fit","h":600},"medium":{"w":600,"resize":"fit","h":360}},"id":588646652363153408,"media\_url\_https":"https://pbs.twimg.com/media/CCtLAvRUIAAp5TB.jpg","media\_url":"http://pbs.twimg.com/media/CCtLAvRUIAAp5TB.jpg","expanded\_url":"http://twitter.com/Telerama/status/588646653294354432/photo/1","source\_status\_id\_str":"588646653294354432","indices":[139,140],"source\_status\_id":588646653294354432,"id\_str":"588646652363153408","type":"photo","display\_url":"pic.twitter.com/tneFkPjX9M","url":"http://t.co/tneFkPjX9M"}],"urls":[{"expanded\_url":"http://po.st/l0L8MW","indices":[109,131],"display\_url":"po.st/l0L8MW","url":"http://t.co/5QfCELXFHr"}],"user\_mentions":[{"id":67278158,"indices":[3,12],"screen\_name":"Telerama","id\_str":"67278158","name":"T\u00E9l\u00E9rama"}]},"aidr":{"features":[{"words":["","et\_d\u00E9sol\u00E9","la\_s\u00E9lection","cette","s\u00E9lection","tweets\_retrouvez","merci","de","d\u00E9sol\u00E9","ici","d\u00E9sol\u00E9\_pour","toute","pour","cette\_avalanche","avalanche","merci\_et","#cannes2015","\_#cannes2015","de\_tweets","#cannes2015\_merci","pour\_cette","s\u00E9lection\_ici","la","avalanche\_de","toute\_la","retrouvez\_toute","tweets","et","retrouvez"],"type":"wordvector"}],"crisis\_code":"150312104037\_emsc\_landslides\_by\_kw\_fra","nominal\_labels":[{"label\_name":"Not physical landslide","source\_id":746,"from\_human":false,"attribute\_description":"The message is related to a physical landslide","label\_code":"03\_not\_landslide","confidence":1,"label\_description":"The item does not refer to a physical landslide","attribute\_code":"landslide\_fr","attribute\_name":"Landslide"}],"doctype":"twitter","crisis\_name":"EMSC Landslides by KW fra"},"source":"<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client<\/a>","favorited":false,"in\_reply\_to\_user\_id":null,"retweet\_count":0,"id\_str":"588734861592154112","user":{"location":"","default\_profile":true,"profile\_background\_tile":false,"statuses\_count":7955,"lang":"en","profile\_link\_color":"0084B4","profile\_banner\_url":"https://pbs.twimg.com/profile\_banners/515262132/1419158044","id":515262132,"following":null,"protected":false,"favourites\_count":8458,"profile\_text\_color":"333333","contributors\_enabled":false,"description":"L'\u00E9criture enchant\u00E9e - Enchanted writing.","verified":false,"name":"Passionissimo","profile\_sidebar\_border\_color":"C0DEED","profile\_background\_color":"C0DEED","created\_at":"Mon Mar 05 08:57:36 +0000 2012","default\_profile\_image":false,"followers\_count":969,"geo\_enabled":false,"profile\_image\_url\_https":"https://pbs.twimg.com/profile\_images/572824763580444672/xAo1UDc2\_normal.jpeg","profile\_background\_image\_url":"http://abs.twimg.com/images/themes/theme1/bg.png","profile\_background\_image\_url\_https":"https://abs.twimg.com/images/themes/theme1/bg.png","follow\_request\_sent":null,"url":null,"utc\_offset":7200,"time\_zone":"Ljubljana","notifications":null,"profile\_use\_background\_image":true,"friends\_count":1030,"profile\_sidebar\_fill\_color":"DDEEF6","screen\_name":"Passionissimo","id\_str":"515262132","profile\_image\_url":"http://pbs.twimg.com/profile\_images/572824763580444672/xAo1UDc2\_normal.jpeg","listed\_count":56,"is\_translator":false}}

* Image Clicker file should have a header info with .csv extension. See the sample ;

User-Name,Tweet,Time-stamp,Location,Latitude,Longitude,Image-Link,TweetID

Oyezscotus,"4 hrs ago #earthquake 5.1 has hit Calatagan , Philippines, 88.9mi, 3:26 AM CST (GEOFON) http://t.co/E2641THHTk",2014-12-31 13:57:00,https://twitter.com/Oyezscotus/status/550289576321744896/photo/1,,,https://pbs.twimg.com/media/B6MFdDnCQAAGq8g.jpg,550289576321744896^M