Concurrent programming lab01 - simon.barras



```
1: import pokemon
 2: import matplotlib.pyplot as plt
 3: import matplotlib.gridspec as gridspec
 4: import time
 5:
 6: k_nb_td_max: int = 15
 7: k_nb_t_min: int = 1
 8: k_step = 1
9: args_ = {
10: "min_pokemon": 1,
10:
          "max_pokemon": 898,
11:
          "tp": "T",
12:
         "nbr_tp": 4,
13:
14:
15: }
16: result_ = {
17: "T": [],
18: "T_key": [],
19:
          "P": [],
20:
          "P_key": [],
          "A_load": [],
"Mi_load": [],
21:
22:
          "Ma_load": []
23:
24: }
25: color_ = {
26: "T": "b",
          "P": "r"
27:
28: }
29:
30: def speed_thread( nb: int = k_nb_td_max):
          global result_
args_['tp'] = "T"
31:
33:
          for i in range(k_nb_td_min, nb, k_step):
34:
               args_['nbr_tp'] = i
35:
               start = int(round(time.time()))
36:
               results_load = pokemon.main(args_, True)
              end = int(round(time.time()))
print(str(i) + " threads -> "+str(end- start) +" [s]")
result_['T'].append(end - start)
37:
38:
39:
40:
               result_['T_key'].append(i)
41:
               sum = 0
42:
              min = results_load[0]
               max = results_load[0]
43:
44:
               for val in results_load:
45:
                    sum += val
                    if min > val: min = val
46:
47:
                    if max < val : max = val
48:
               sum /= len(results_load)
              result_['A_load'].insert(i, sum)
result_['Mi_load'].insert(i, min)
result_['Ma_load'].insert(i, max)
49:
50:
51:
52:
53:
          fig, (ax1, ax2) = plt.subplots(2, 1)
54:
          fig.suptitle('Speed test threads')
55:
56:
          ax1.plot(result_['T_key'], result_['T'], c=color_['T'], label='Threads')
57:
          ax1.set_ylabel('Time (s)')
58:
          ax1.set_xlabel('Number of threads')
59:
60:
          ax2.plot(result_['T_key'], result_['A_load'], c='b', label='Load average')
ax2.plot(result_['T_key'], result_['Ma_load'], 'g--', label='Max load')
ax2.plot(result_['T_key'], result_['Mi_load'], 'r--', label='Min load')
61:
62:
63:
64:
65:
          ax2.set_ylabel('Image by thread average')
          ax2.set_xlabel('Number of threads')
66:
67:
          return ax1, ax2
68:
69: def speed_process( nb: int = k_nb_td_max):
70:
          global result_
args_['tp'] = "P"
71:
72:
          for i in range(k_nb_td_min, nb, k_step):
73:
               args_['nbr_tp'] = i
74:
               start = int(round(time.time()))
75:
               results_load = pokemon.main(args_, True)
              end = int(round(time.time()))
print(str(i) + " process -> "+str(end- start) +" [s]")
result_['P'].append(end - start)
76:
77:
78:
79:
               result_['P_key'].append(i)
80:
81:
          fig, (ax1) = plt.subplots(1, 1)
82:
          fig.suptitle('Speed test process')
83:
          ax1.plot(result_['P_key'], result_['P'], c=color_['P'], label='Process')
84:
85:
86:
          ax1.set_ylabel('Time (s)')
87:
          ax1.set_xlabel('Number of process')
88:
          return ax1
89:
90: def load_experiment(type: str, number_of_tp: int = 4):
          args_["tp"] = type
91:
92:
          args_["nbr_tp"] = number_of_tp
93:
          results = pokemon.main(args_, True)
94:
          somme = 0
95:
          for i in range(len(results)):
96:
               somme += results[i]
```

```
97:
            plt.title("Load by " + type)
plt.ylabel("Number of images")
 98:
 99:
            plt.xlabel("Id")
100:
101:
           id = 0
102:
           # for nb in results:
           # plot.bar(id, somme / nb, str(results))
# id += 1
103:
104:
105:
           plt.bar(range(len(results)), results)
106:
107:
108: if __name__ == '__main__':
109:    print("Speed test with threads...")
110:    speed thread()
110:
           speed_thread()
           #plt.show()
111:
           print("Speed test with process...")
speed_process()
plt.show()
112:
113:
114:
           print("Load test with threads...")
#load_experiment("T", )
115:
116:
117:
           #plt.show()
          print("Load test with process...")
#load_experiment("P", )
118:
119:
          #plt.show()
120:
121:
```

63:

```
1: # -*- coding: utf-8 -*-
 3: # Copyright 2021, School of Engineering and Architecture of Fribourg
 4: #
 5: # Licensed under the Apache License, Version 2.0 (the "License");
     # you may not use this file except in compliance with the License.
     # You may obtain a copy of the License at
 8: #
 9: #
           http://www.apache.org/licenses/LICENSE-2.0
10: #
11: # Unless required by applicable law or agreed to in writing, software 12: # distributed under the License is distributed on an "AS IS" BASIS,
13: # WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14: # See the License for the specific language governing permissions and
15: # limitations under the License.
16:
17: """
18: This file contains a function, which downloads a specific Pokemon image from assets.pokemon.com website.
19: It is a helper funcition that wil be used in the Concurrent Systems lab01.
20: """
21:
22: __author__ = 'Michael Maeder'
23: __date__ = "04.10.2021"
24: _version_ = "1.1"
25: _email_ = "michael.maeder@hefr.ch"
26: _userid_ = "michael.maeder"
27:
28:
29: import cv2
30: import numpy as np
31: import urllib.request
33: URL = 'https://assets.pokemon.com/assets/cms2/img/pokedex/detail/'
34:
35:
36: def download_image(img_nbr: int, path: str, verbose_print: bool = False):
37:
38:
         Downloads a specific pokemon image in PNG format. The function gets a index number and downloads the corresponding
         image. The PNG-converted image will be written to the given filesystem path
39:
40:
41:
         :argument
42:
             img_nbr
                                -- the number of the image (1 .. 898)
-- the path where the image will be written to
43:
              path
              verbose_print -- selector if some debugging information should be printed or not
44:
45:
46:
        urllib.error.HTTPError -- will be raised by the request library. The exception will be forwarded to the caller
47:
48:
49:
50:
              # creation of the request object with the given URL
51:
              req = urllib.request.Request(URL + '{:03d}'.format(img_nbr) + '.png')
53:
              response = urllib.request.urlopen(req)
                                                                                          # response contains the raw response data
54:
              raw_image_data = response.read()
                                                                                          # get the bytes of the image
              image = np.asarray(bytearray(raw_image_data), dtype="uint8")  # convert it to an array of uint8s
image = cv2.imdecode(image, cv2.IMREAD_UNCHANGED)  # the code the array as an image
cv2.imwrite(path + "/" + '{:03d}'.format(img_nbr) + ".png", image)  # finally write the PNG to the given path
55:
56:
57:
             if verbose_print:
    print("Saved " + '{:03d}'.format(img_nbr) + ".png")
58:
59:
        except Exception as e:
60:
61:
             print("Error occured for Pokemon " + '{:03d}'.format(img_nbr))
62:
              raise e
                                                                                          # forward the exception to the caller
```

```
1: #!/usr/bin/env python3
 2: # -*- coding: utf-8 -*-
 3:
 4: # Copyright 2021, School of Engineering and Architecture of Fribourg
 5: #
    # Licensed under the Apache License, Version 2.0 (the "License");
    # you may not use this file except in compliance with the License.
 8:
    # You may obtain a copy of the License at
 9: #
10: #
          http://www.apache.org/licenses/LICENSE-2.0
11: #
12: # Unless required by applicable law or agreed to in writing, software
13: # distributed under the License is distributed on an "AS IS" BASIS,
14: # WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
15: \# See the License for the specific language governing permissions and
16: # limitations under the License.
17:
18:
19: """
20: Download a range of pokemon's picture.
21: The number of thread or process is configurable
22: """
23:
24: __author__ = "Simon Barras"
25: __date__ = "2021-10-15"
26: __version__ = "0.0.1"
27: __email__ = "simon.barras@edu.hefr.ch"
28: __userid__ = "simon.barras"
29:
30: import time
31:
32: from getpokemon import download_image
33: import sys
                                                        Il plat les los
Jocurented 2 cons
34: import threading
35: import multiprocessing
36:
37:
38: k_max_pokemon = 898
39: k_default_verbose = False
40: k_nb_args = 5
41: args_ = {
42:
         "min_pokemon": 1,
        "max_pokemon": k_max_pokemon,
43:
44:
        "tp": "T",
45:
        "nbr_tp": 6,
        "path": "./tmp"
46:
47: }
48: tp_list_ = []
49: tp_result_ = []
50: count_ = 0
51: do_stats_ = False
52:
53:
54: # This method can have some concurrent problems ~ . Why
55: def worker_canHaveConcurrentProblem(index):
56:
        global count_
        while count_ < args_['max_pokemon']:
    # Protection de la section critique</pre>
57:
58:
             count_ += 1
             #print("TD: " + str(index) + " is downloading " +str(count))
60:
61:
             try:
62:
                 download_image(count_, args_["path"], k_default_verbose)
63:
             except Exception as e:
                 continue
64:
65:
             if do_stats_:
66:
                 global tp_result_
67:
                 tp_result_[index] += 1
68:
        describe worker (index):
69:
70: def
        # Download all image from the beginning to the end
71:
72:
         # Increment value with the number of TP
        for i in range(1 + index, args_['max_pokemon'], args_['nbr_tp']
    #print("TD: " + str(index) + " is downloading " +str(i))
73:
74:
75:
             try:
76:
                 download_image(i, args_["path"], k_default_verbose)
77:
             except Exception as e:
    continue
                                       C3
78:
79:
             if do_stats_:
80:
                 global tp_result_
                 tp_result_[index] += 1
81:
82:
83: (l) (l) 84: def initialize():
85:
        global count_
86:
        count_ = args_['min_pokemon'] - 1
87:
        for i in range(0, args_["nbr_tp"]):
88:
             if args_["tp"] == "T":
89:
                 tp = threading.Thread(target=worker, args=(i, ))
             else:
90:
91:
                 tp = multiprocessing.Process(target=worker, args=(i,
92:
93:
             tp_list_.append(tp)
94:
95:
96: def clear():
```

```
global tp_result_
 98 .
         tp_list_.clear()
 99:
         tp_result_.clear()
         tp_result_ = [0] * args_["nbr_tp"]
100:
101:
102:
103: def main(args=args_, give_stat=False):
104:
          "" main function ""
         global args_
105:
106:
         args_ = args
global do_stats_
107:
108:
         do_stats_ = give_stat
109:
110:
         # Clear all residual values
111:
        clear()
112:
         # Initialize return value if is not necessary
113:
114:
        if not do_stats_:
115:
            tp_result_.append(0)
116:
117:
         # Initialize all values
118:
            Create thread/process
        initialize()
119:
120:
121:
         # Start threads/process
         for tp in tp_list_:
123:
            tp.start()
124:
         # Wait until threads/process finish
125:
         for tp in tp_list_:
126:
127:
            tp.join()
129:
         return tp_result_
130:
131:
                                                                      exceptions.
132: # main program entry point
if i == 1:
                 args_["min_pokemon"] = int(sys.argv[1])
136:
137:
             elif i == 2:
                args_["max_pokemon"] = int(sys.argv[1])
138:
             elif i == 3:

    args_["tp"] = sys.argv[i]

elif i == 4:
139:
140:
141:
             args_["nbr_tp"] = int (sys.argv[i])
elif i == 5:
143:
144:
                 args_["path"] = sys.argv[i]
         start = time.time()
145:
146:
         main()
147:
         end = time.time()
        print("Time %f" % (end - start))
148:
```

Git Logs

commit 7fbb461edec11afe0c0ffcb44b076832c23d2d17 (HEAD -> refs/heads/main,
refs/remotes/origin/main, refs/remotes/origin/HEAD)

Author: Simon Barras <simon.barras@edu.hefr.ch>

Date: Fri, 15 Oct 2021 17:14:14 +0000

Update README.md

commit 1cfeba08c8edacae0bba4d858de0126bcf43de00

 ${\tt Author: Simon \ Barras}\ < 66463606 + {\tt simbarras@users.noreply.github.com} >$

Date: Fri, 15 Oct 2021 18:50:43 +0200

add a little report

commit e03701968f3af11f796ffa42417a561839fbaf3c

 ${\tt Author: Simon \ Barras}\ <\!66463606\!+\!simbarras@users.noreply.github.com\!>$

Date: Fri, 15 Oct 2021 18:50:26 +0200

finilasie experiment

commit 1655aabb3a13c69a15f684f8933953af4d8ea330

Author: Simon Barras <66463606+simbarras@users.noreply.github.com>

Date: Fri, 15 Oct 2021 13:36:17 +0200

rename variable

commit 202114bad6a0dfe6e807a05771b6646da7913947

Author: Simon Barras <66463606+simbarras@users.noreply.github.com>

Date: Fri, 8 Oct 2021 13:11:20 +0200

update result's graph

commit 9393ee14b247a64a7ad67eb972b3a3036e9250c5

 ${\tt Author: Simon \ Barras}\ < 66463606 + {\tt simbarras@users.noreply.github.com} >$

Date: Fri, 8 Oct 2021 10:52:49 +0200

Fix load graph bar

commit 096e2d5de4c13cc96890f549565396fe73f35fc5

Author: Simon Barras <66463606+simbarras@users.noreply.github.com>

Date: Fri, 8 Oct 2021 10:17:31 +0200

Fix load average

commit 834e87d839a8d1085edabcc638df87560d799781

 ${\tt Author: Simon \ Barras}\ < 66463606 + {\tt simbarras@users.noreply.github.com} >$

Date: Fri, 8 Oct 2021 09:39:34 +0200

update test

commit d5e019bda734c88102cfafb924e0b5c4dad33907

 ${\tt Author: Simon \ Barras}\ < 66463606 + {\tt simbarras@users.noreply.github.com} >$

Date: Tue, 5 Oct 2021 23:42:16 +0200

first trytp work in same time

commit 54f45a04654bcd2ddc903542335eaf87d8515995

Author: Simon Barras <66463606+simbarras@users.noreply.github.com>

Date: Tue, 5 Oct 2021 22:46:43 +0200

first try

commit b6e016b606ab4ee6b937b5b19b123932a354cd5c

Author: Simon Barras <66463606+simbarras@users.noreply.github.com>

Date: Tue, 5 Oct 2021 15:51:29 +0200

downloading files

commit 60fcb53c196d05eee076a6028b71b2a027f04bf5

Merge: 6892738 a39309c

Author: Simon Barras <66463606+simbarras@users.noreply.github.com>

Date: Tue, 5 Oct 2021 13:43:24 +0200

Merge branch 'main' of https://gitlab.forge.hefr.ch/concurp/2021-

2022/concurp-student-labs

commit a39309cfa63e7882a5a9eb053a41ea5bf6cfb817

Author: Michael Mäder <michael.maeder@hefr.ch>

Date: Tue, 5 Oct 2021 11:42:16 +0000

img_nbr argument description fixed

.

commit 68927386afb9226e35024dd494b6af27544b3ae5

Merge: 63b4c71 4420bdc

Author: Simon Barras <66463606+simbarras@users.noreply.github.com>

Date: Tue, 5 Oct 2021 13:35:04 +0200

Merge branch 'main' of https://gitlab.forge.hefr.ch/concurp/2021-2022/concurp-student-labs

commit 63b4c714df47a2a07588f16d32abec13faaf5b21

 ${\tt Author: Simon \ Barras}\ < 66463606 + {\tt simbarras@users.noreply.github.com} >$

Date: Tue, 5 Oct 2021 13:32:14 +0200

add tests

.

commit 4420bdc331f3756219fdbbe0a973c44e2de2d5d4

Merge: cb7835c e33f5b2

Author: Michael Mäder <michael.maeder@hefr.ch>

Date: Tue, 5 Oct 2021 09:28:53 +0200

Merge branch 'main' of https://gitlab.forge.hefr.ch/concurp/2021-2022/concurp-student-labs

commit cb7835c50510512efcd2773a3ca6132ca646ab76

Author: Michael Mäder <michael.maeder@hefr.ch>

Date: Tue, 5 Oct 2021 09:28:45 +0200

.gitignore for Python stuff.

modified example script

.

commit e33f5b243a0934416bafe819fb3898d1105ec66f

Author: Michael Mäder <michael.maeder@hefr.ch>

Date: Mon, 4 Oct 2021 15:46:55 +0000

pip requirements for the helper function

commit 78181db37b4f31bbd7820b665bbd68fdc1d18a2e

Author: Michael Mäder <michael.maeder@hefr.ch>

Date: Mon, 4 Oct 2021 15:29:09 +0000

main program skeleton

commit d3c5201d25cc7163d96f367dcd5310b03e4cfab7

Author: Michael Mäder <michael.maeder@hefr.ch>

Date: Mon, 4 Oct 2021 15:28:02 +0000

Add new file

commit 628ee2696269da6283e79b242ab566ae2e605fb4

Author: Michael Mäder <michael.maeder@hefr.ch>

Date: Mon, 4 Oct 2021 15:27:12 +0000

Add new directory

commit 59a5d5681ba0eced659e4ca453a5336eedebe96d

Author: Michael Mäder <michael.maeder@hefr.ch>

Date: Mon, 4 Oct 2021 14:51:37 +0000

helpfer function for the Pokémon lab uploaded

commit 0a548c28737e4747255b6058c45c2d1bb543b552

Author: Michael Mäder <michael.maeder@hefr.ch>

Date: Mon, 4 Oct 2021 14:50:04 +0000

Add new directory

commit 46a33f06042a72406d72d7f8b69e84d04005f061

Author: Michael Mäder <michael.maeder@hefr.ch>

Date: Mon, 4 Oct 2021 14:49:51 +0000

Add new directory

Run params: 10-13-P-8—tmp-pokemon

Directory statistics for 10-13-P-8--tmp-pokemon



should read only 4 incres!

Run params: 10-13-P-8-tmp-pokemon

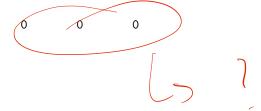
Time 18.815633 libpng warning: iCCP: known incorrect sRGB profile libpng warning: iCCP: known incorrect sRGB profile

```
libpng warning: iCCP: known incorrect sRGB profile
```

libpng warning: iCCP: known incorrect sRGB profile libpng warning: iCCP: known incorrect sRGB profile libpng warning: iCCP: known incorrect sRGB profile

Run params: 1-800-P-32-tmp-pokemon

Directory statistics for 1-800-P-32--tmp-pokemon



Run params: 1-800-P-32-tmp-pokemon

Time 28.269413 libpng warning: iCCP: known incorrect sRGB profile libpng warning: iCCP: known incorrect sRGB profile

```
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profilelibpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
```

```
libpng warning: iCCP: known incorrect sRGB profile
```

```
libpng warning: iCCP: known incorrect sRGB profile
#######
#######
libpng warning: iCCP: known incorrect sRGB profile
```

libpng warning: iCCP: known incorrect sRGB profile libpng warning: iCCP: known incorrect sRGB profile

```
libpng warning: iCCP: known incorrect sRGB profile
```

```
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profilelibpng warning: iCCP: known incorrect sRGB profilelibpng
warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profilelibpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
```

```
libpng warning: iCCP: known incorrect sRGB profile
```

Run params: 1-8-M-32-tmp-pokemon

Directory statistics for 1-8-M-32--tmp-pokemon

0 0 0

what is

Run params: 1-8/M-32-tmp-pokemon

Time 33.365499 libpng warning: iCCP: known incorrect sRGB profile libpng warning: iCCP: known incorrect sRGB profile

```
libpng warning: iCCP: known incorrect sRGB profile
```

```
libpng warning: iCCP: known incorrect sRGB profile
```

```
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profilelibpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
#######
#######
libpng warning: iCCP: known incorrect sRGB profile
```

```
libpng warning: iCCP: known incorrect sRGB profile
```

```
libpng warning: iCCP: known incorrect sRGB profile
```

```
libpng warning: iCCP: known incorrect sRGB profile
```

Run params: 1-8-P-0-tmp-pokemon

Directory statistics for 1-8-P-0--tmp-pokemon

0 0 0

Run params: 1-8-P-0-tmp-pokemon

Time 0.000004

Run params: 30-47-T-8 tmp-pokemon

Directory statistics for 30 47-T-8--tmp-pokemon

46 46 1012

Run params: 30-47-T-8—tmp-pokemon

Time 2.296827

Run params: default

Directory statistics for default

0 0 0

Run params: default

Time 23.662107 libpng warning: iCCP: known incorrect sRGB profile libpng warning: iCCP: known incorrect sRGB profile

```
libpng warning: iCCP: known incorrect sRGB profile
```

Lab01 - Pokémon figure scraper

This is the second TP of the lesson "Concurrent system". In this exercise we need to test threads and process to download quickly 900 pictures of pokemon. The guideline is available here.

This project is made by @Simon.barras

Launch app
python3 pokemon.py [start_id [stop_id [tp [nbr_tp [out_dir]]]]]

argument	description	default value
pokemon.py	name of the python program file that contains the main entry point	
$start_id$	the ID of the first image to download	1
stop_id	the ID of the last image to download (included \rightarrow the image with this ID must also be downloaded)	898
tp	selector of multi-threading or multi-processing, T: threading, P: processing	T
nbr_tp	the number of threads or processes	4
out_dir	the directory, where the files are written to	$/\mathrm{tmp}/$

pokemon.py

The code is documented but there is 2 workers. The used worker is poke-mon.worker(index) and this how it work:

For example, we want to download 10 pictures with 4 threads

This is the load's repartition > Jump = number of thread/process

Threads / process	Image
TP 1	0, 4, 8
TP 2	1, 5, 9

Threads / process	Image
TP 3	2, 6
TP 4	3, 7

And here this is an example of the activity

Image	0	4	2	1	3	5	7	8	9	6
TP 1	-	-						-		
TP 2				-		-			-	
TP 3			-							-
TP 4					-		-			



This version of the worker work well but if a process finish earlier as his teammate he can't help them. To fix this, I have think about an algorithm where thread ask about the next image to download. The effect is that the images are downloaded in order but the thread/process aren't downloading the same number of images. If we retake the previous example we can have something like this:

Image	0	1	2	3	4	5	6	7	8	9
TP 1	-	_						-		
TP 2				-		-			-	-
TP 3			-							
TP 4					-		-			

```
Time ------
```

The benefit is the time, if a thread is slower, all other may help him. The problem with this solution is the concurrent problem.

```
def worker_canHaveConcurrentProblem(index):
    global count_
    while count_ < args_['max_pokemon']:
        # Protection de la section critique
        count_ += 1
        #print("TD: " + str(index) + " is downloading " +str(count))
        try:
            download_image(count_, args_["path"], k_default_verbose)
        except Exception as e:
            continue
        if do_stats_:</pre>
```



```
global tp_result_
tp_result_[index] += 1
```

If there is a shift between count_ += 1 and download_image(count_, args_["path"], k_default_verbose) the thread may not download the right

nood tost throads

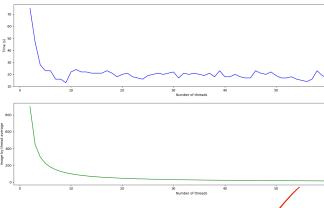
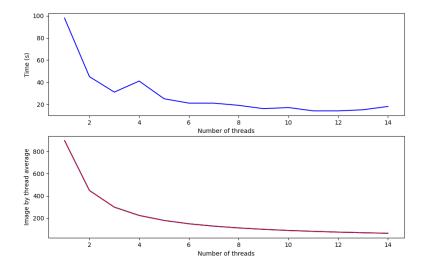
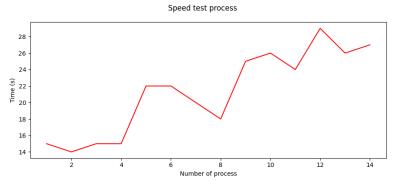


image. We can see this effect with this picture: When there are to many threads, the time isn't stable.

Result

I have run experiment.py from my home and I have get this result: $$^{\mbox{\scriptsize Speed}}$$ threads





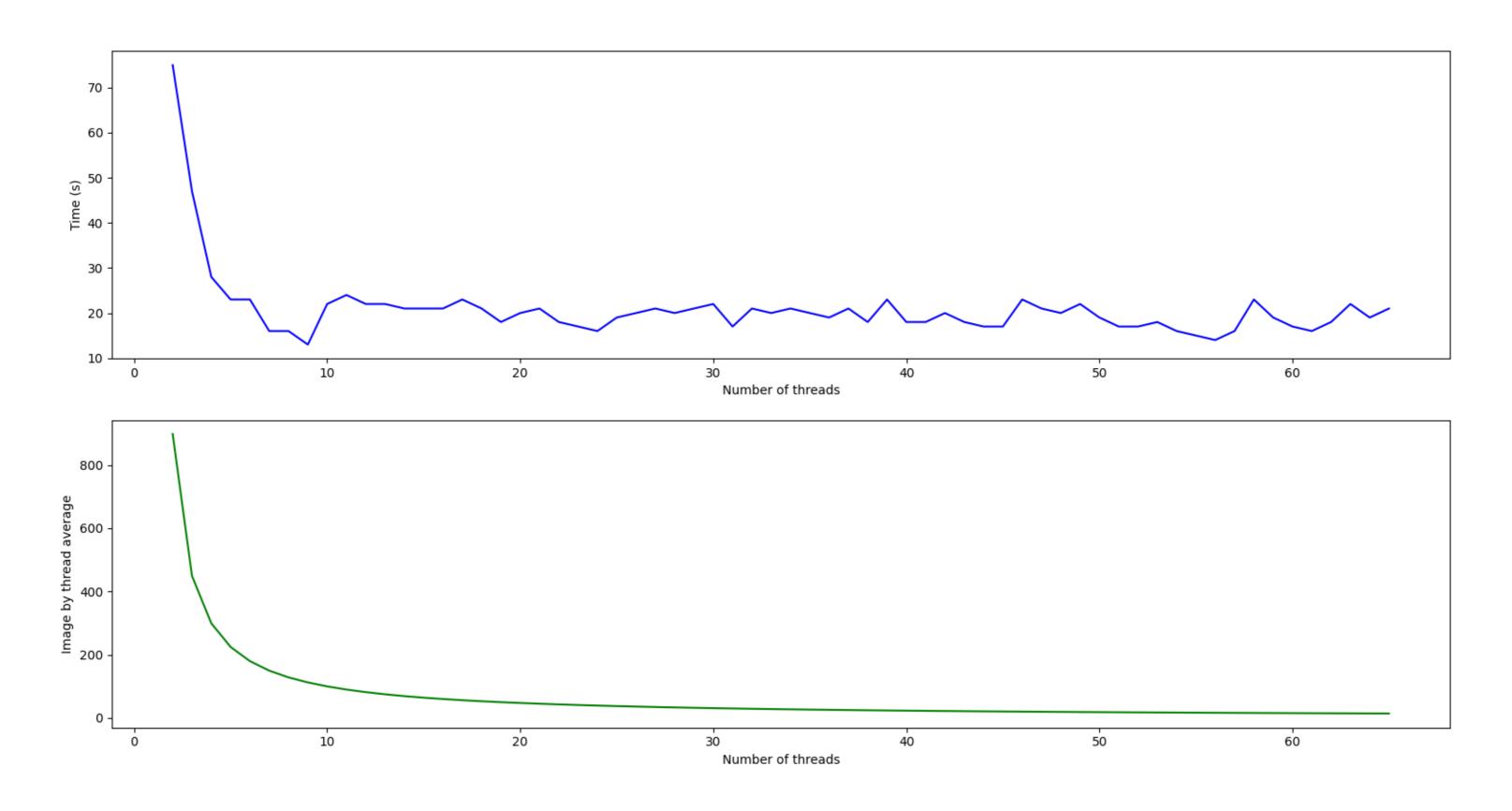
> I have use *matplotlib* to do this graph

For the threads, the time is always lower but from 8 there is no relevant improvement. This is due to the time to create the threads. Compared to the threads, the process is faster but we can quickly see that the performance decreases.

Conclusion

To summarize, we can save a lot of time with process and threads and the process are faster than threads but it's a big cost to create them. The last interrogation is, why 1 process is faster than 1 thread?





Lab01 - Pokémon figure scraper

This is the second TP of the lesson "Concurrent system". In this exercise we need to test threads and process to download quickly 900 pictures of pokemon. The guideline is available here.

This project is made by @Simon.barras

Launch app
python3 pokemon.py [start_id [stop_id [tp [nbr_tp [out_dir]]]]]

argument	description	default value
pokemon.py	name of the python program file that contains the main entry point	
$start_id$	the ID of the first image to download	1
stop_id	the ID of the last image to download (included \rightarrow the image with this ID must also be downloaded)	898
tp	selector of multi-threading or multi-processing, T: threading, P: processing	T
nbr_tp	the number of threads or processes	4
out_dir	the directory, where the files are written to	$/\mathrm{tmp}/$

pokemon.py

The code is documented but there is 2 workers. The used worker is poke-mon.worker(index) and this how it work:

For example, we want to download 10 pictures with 4 threads

This is the load's repartition > Jump = number of thread/process

Threads / process	Image
TP 1	0, 4, 8
TP 2	1, 5, 9

Threads / process	Image
TP 3	2, 6
TP 4	3, 7

And here this is an example of the activity

Image	0	4	2	1	3	5	7	8	9	6
TP 1	-	-						-		
TP 2				-		-			-	
TP 3			-							-
TP 4					-		-			

This version of the worker work well but if a process finish earlier as his teammate he can't help them. To fix this, I have think about an algorithm where thread ask about the next image to download. The effect is that the images are downloaded in order but the thread/process aren't downloading the same number of images. If we retake the previous example we can have something like this:

Image	0	1	2	3	4	5	6	7	8	9
TP 1	_	_						-		
TP 2				-		-			-	-
TP 3			-							
TP 4					-		-			

Time ------

The benefit is the time, if a thread is slower, all other may help him. The problem with this solution is the concurrent problem.

```
def worker_canHaveConcurrentProblem(index):
    global count_
    while count_ < args_['max_pokemon']:
        # Protection de la section critique
        count_ += 1
        #print("TD: " + str(index) + " is downloading " +str(count))
        try:
            download_image(count_, args_["path"], k_default_verbose)
        except Exception as e:
            continue
        if do_stats_:</pre>
```

```
global tp_result_
tp_result_[index] += 1
```

If there is a shift between count_ += 1 and download_image(count_, args_["path"], k_default_verbose) the thread may not download the right

Enough tost through

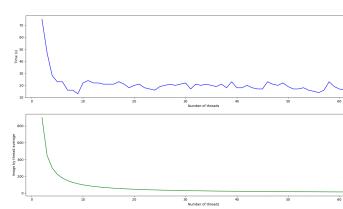
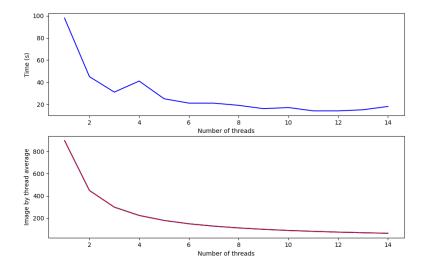
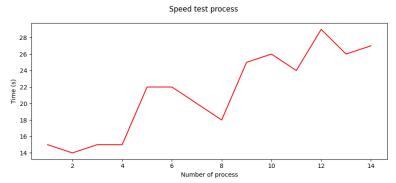


image. We can see this effect with this picture: When there are to many threads, the time isn't stable.

Result

I have run experiment.py from my home and I have get this result: $$^{\mbox{\scriptsize Speed}}$$ threads





> I have use *matplotlib* to do this graph

For the threads, the time is always lower but from 8 there is no relevant improvement. This is due to the time to create the threads. Compared to the threads, the process is faster but we can quickly see that the performance decreases.

Conclusion

To summarize, we can save a lot of time with process and threads and the process are faster than threads but it's a big cost to create them. The last interrogation is, why 1 process is faster than 1 thread?



