

In [21]:

```
import zipfile

from PIL import Image
from IPython.display import display
import pytesseract
import cv2 as cv
import numpy as np
import os

face_cascade = cv.CascadeClassifier('readonly/haarcascade_frontalface_default.xml')
```

In [22]:

```
with zipfile.ZipFile("readonly/images.zip", "r") as zip_ref:
    zip_ref.extractall("extracted/")
```

In [23]:

```
def scan_text(page):
    page_img = Image.open(page)
    text_str = pytesseract.image_to_string(page_img).lower()
    return text_str.split()
```

In [24]:

```
def find_faces(page):
    cv_page = cv.imread(page)
    gray_page = cv.cvtColor(cv_page, cv.COLOR_BGR2GRAY)
    faces = face_cascade.detectMultiScale(gray_page, scaleFactor=1.3, minNeighbors=5, minSize=(25,25))
    contact_sheet = None
    if len(faces) > 0:
        PIL_page = Image.open(page)
        contact_sheet = Image.new(PIL_page.mode, (100*5, 100*(len(faces)//5+1)))
        c_x = 0
        c_y = 0
        for x, y, w, h in faces:
            face = PIL_page.crop((x, y, x+w, y+h))
            contact_sheet.paste(face.resize((100,100), Image.ANTIALIAS), (c_x,c_y))
            if c_x == 400:
                c_x = 0
                c_y += 100
            else:
                c_x += 100
        return contact_sheet
```

In [25]:

```
def search(tag):
    print("Search tag: " + tag)
    for page in os.listdir("extracted/"):
        page_path = "extracted/" + page
        if tag.lower() in scan_text(page_path):
            print("Results found in file " + page)
            if find_faces(page_path) is None:
                print("But there were no faces in that file!")
            else:
                display(find_faces(page_path))
```

In [26]:

```
search("Christopher")  
search("Mark")
```

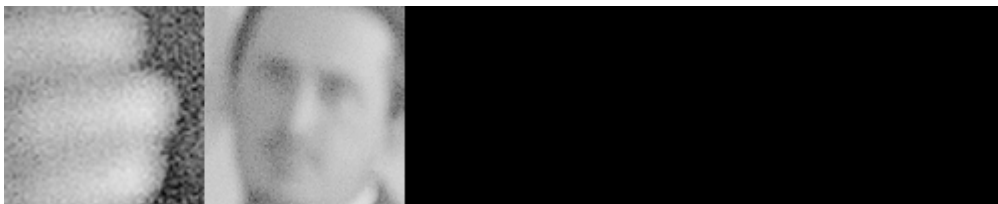
Search tag: Christopher  
Results found in file a-3.png



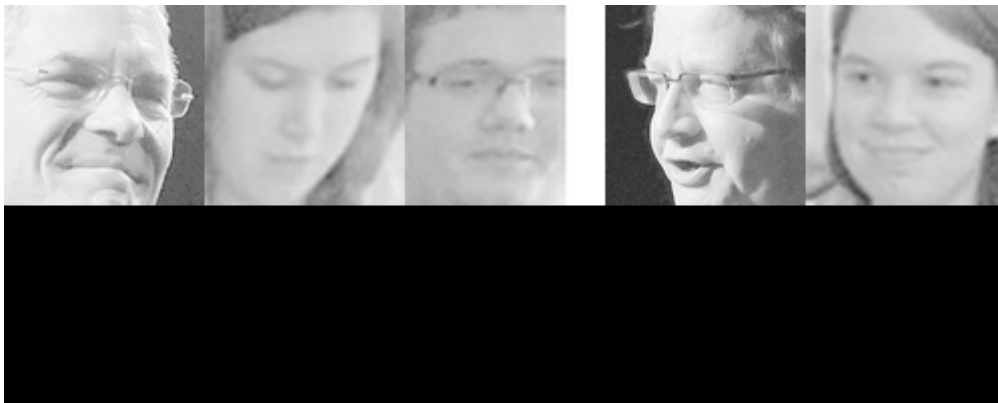
Results found in file a-0.png



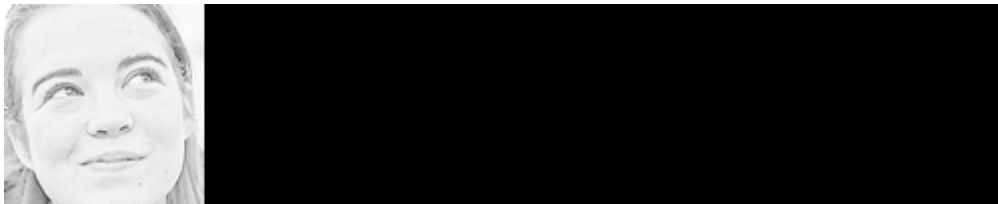
Search tag: Mark  
Results found in file a-3.png



Results found in file a-10.png  
But there were no faces in that file!  
Results found in file a-1.png



Results found in file a-9.png



Results found in file a-5.png



Results found in file a-0.png



Results found in file a-8.png  
But there were no faces in that file!  
Results found in file a-2.png



In [ ]: