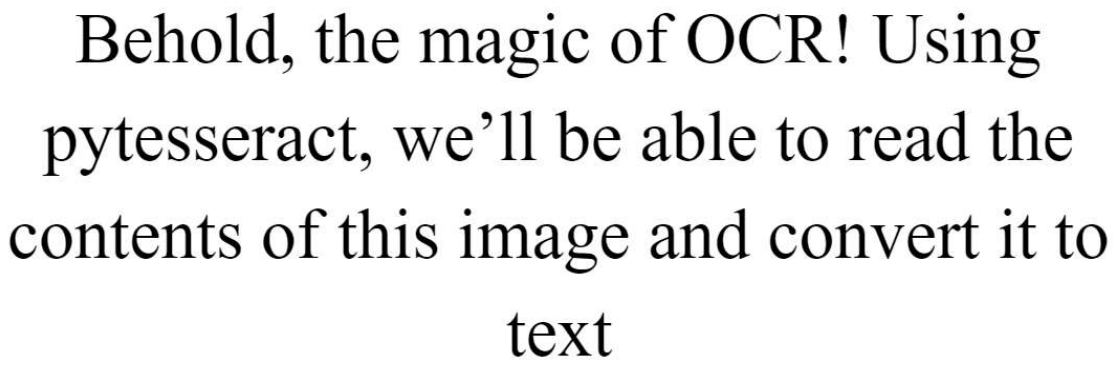
coursera Help

Lecture: The (Py)Tesseract Library

In [33]:
```
# We're going to start experimenting with tesseract using just a simple image of
# Lets first import Image from PIL and display the image text.png.
from PIL import Image

image = Image.open("readonly/text.png")
display(image)
```



In [34]:
```
# Great, we have a base image of some big clear text
# Lets import pytesseract and use the dir() fundtion to get a sense of what mighi
# functions to play with
import pytesseract
dir(pytesseract)
```

Out[34]:
```
['Output',
 'TesseractError',
 '__builtins__',
 '__cached__',
 '__doc__',
 '__file__',
 '__loader__',
 '__name__',
 '__package__',
 '__path__',
 '__spec__',
 'get_tesseract_version',
 'image_to_boxes',
 'image_to_data',
 'image_to_osd',
 'image_to_pdf_or_hocr',
 'image_to_string',
 'pytesseract']
```

In [35]:
```
# It looks like there are just a handful of interesting functions, and I think ir
# is probably our best bet. Lets use the help() function to interrogate this a bi
help(pytesseract.image_to_string)
```