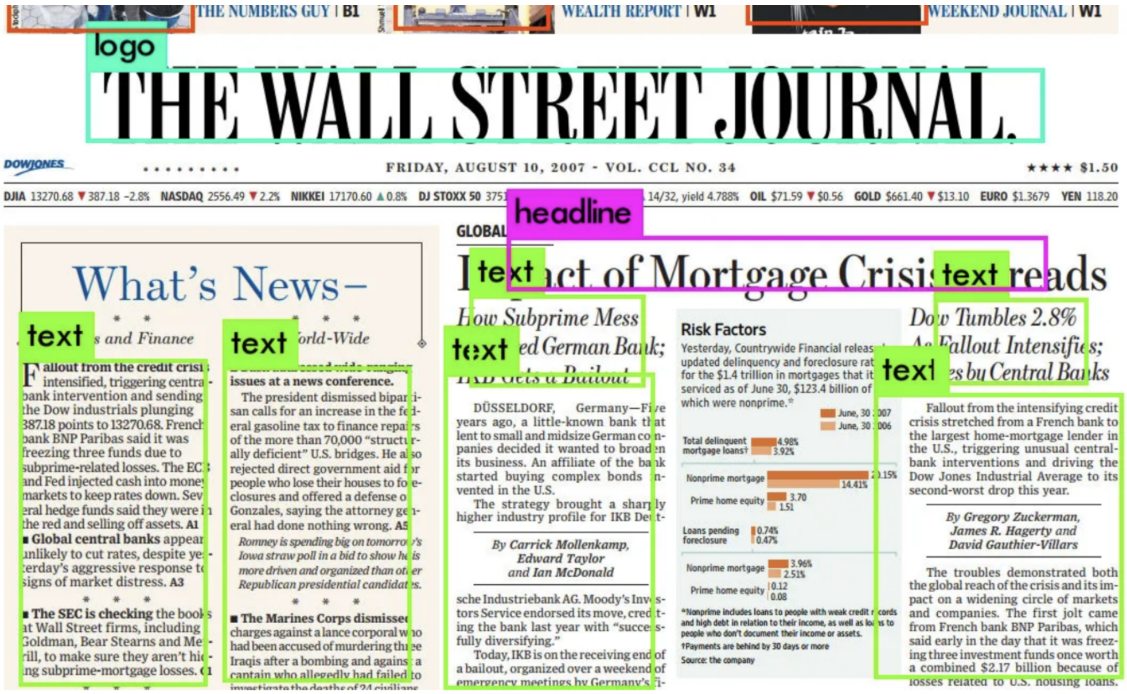
Alberto Andrés Valdés González.
Degree: Mathematical Engineer.
Work position: Data Scientist.
Mail: anvaldes@uc.cl/alberto.valdes.gonzalez.96@gmail.com
Location: Santiago, Chile.

Preprocessing PDF

In many AI problems like the related to the computer vision in which we have to use models like OCR we have to start by doing a preprocessing of the PDFs.



Steps

1. Transform PDF to images:

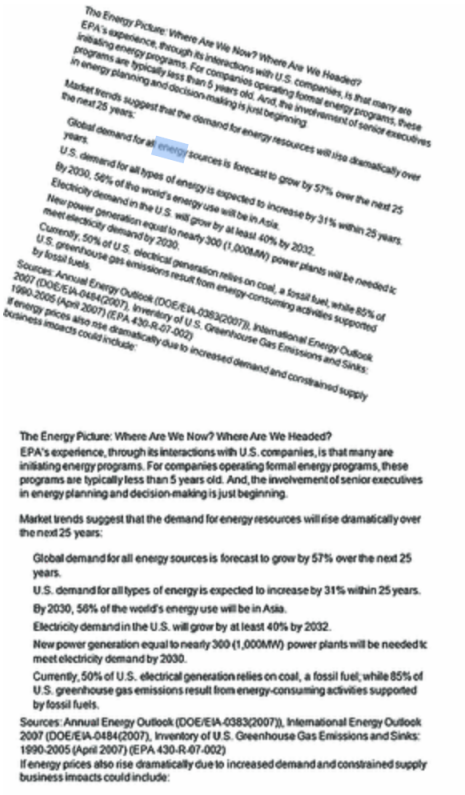
In this step we going to transform every page of the PDF to an image. This step is very important because once we have these images we can work as usual we do in the Computer Vision problems.

2. Transform from RGB to Gray:

To work with only one channel is necessary transform every image to a gray scale.

3. Skew Correction:

It is very important do a skew correction because many times the pdf to analyze corresponds to scan of differents documents and how is natural many of these documents are scanned with skew.



The Energy Picture: Where Are We Now? Where Are We Headed?

EPA's experience, through its interactions with U.S. companies, is that many are initiating energy programs. For companies operating formal energy programs, these programs are typically less than 5 years old. And, the involvement of senior executives in energy planning and decision-making is just beginning.

Market trends suggest that the demand for energy resources will rise dramatically over the next 25 years:

- Global demand for all energy sources is forecast to grow by 57% over the next 25 years.
- U.S. demand for all types of energy is expected to increase by 31% within 25 years.
- By 2030, 56% of the world's energy use will be in Asia.
- Electricity demand in the U.S. will grow by at least 40% by 2032.
- New power generation equal to nearly 300 (1,000MW) power plants will be needed to meet electricity demand by 2030.
- Currently, 50% of U.S. electrical generation relies on coal, a fossil fuel; while 85% of U.S. greenhouse gas emissions result from energy-consuming activities supported by fossil fuels.

Sources: Annual Energy Outlook (DOE/EIA-0383(2007)), International Energy Outlook 2007 (DOE/EIA-0484(2007), Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2005 (April 2007) (EPA 430-R-07-002)

If energy prices also rise dramatically due to increased demand and constrained supply business impacts could include:

4. Noise removal

The main objective of the Noise removal stage is to smoothen the image by removing small dots/-patches which have high intensity than the rest of the image. Noise removal can be performed for both Coloured and Binary images.

5. Binarization of the image

We have to determine a threshold to analyze every pixel of an image and if the value of the pixel is greather than the threshold then we going to assign the maximum value for a pixel and f the value of the pixel is lesser than the threshold then we going to assign the minimum value for a pixel.

6. Other filters

Considering our objectives or our preferences we can apply different filters to every image.
