**Clase 10**

**Expresiones regulares:**

Para hacer scraping primero hace que usar expresiones regulares. Es un texto especial para describir un patrón de búsqueda. Como comodines en esteroides. Regex es el módulo o librería de las expresiones regulares. Se escribe import re. Si uso find me encuentra la posición del carácter. El resultado de utilizar el .search es un objeto del que se puede extraer cualquier cosa. .group se pone para que haya salida. \w+ sirve para buscar todas las letras a la izquierda o a la derecha del carácter que se indicó. Findall encuentra todas las ocurrencias dentro de un texto. En método findall las reglas son las mismas que en el search, la diferencia es que el search solo busca la primera ocurrencia. Si el código del método findall se pone entre paréntesis solo salen las ocurrecias, si se pone sin paréntesis sale todo el grupo (ejemplo todo el correo). El + es one or more, el \* es cero or more. Dentro de los corchetes busca los caracteres en específico.

**Web Scraping:**

Detrás de toda página hay un código fuente. El elemento central en html son las etiquetas. Scrape en inglés quiere decir raspar. Es el proceso de extraer información de páginas de internet. La ventaja es que no se depende de que las páginas tengan API (link de descarga de datos).