# El proceso de Text Mining.

Diagrama

Descripción generada automáticamente

## 1. Data Acquisition

- Esta parte se basa en de dónde voy a obtener la información. Algunas técnicas son: Datasets, Scraping, Product Intervention, Data Augmentation.

- **Datasets**: Conjuntos de datos disponibles.

- **Scaping**: Utilizar un código para obtener datos de varios sitios web.

- Usar las capacidades que tienen algunos productos de recaudar datos y usarlos.

- **Data Augmentation**: Hacer que un Dataset se vuelva mucho más grande usando técnicas.

\* *Remplazo de palabras por Sinónimos*

\* *Back translation:* traducir de un idioma a otro y luego volverlo al original.

\* *Replacing entities*: Se identifican entes y se remplazan por cosas similares.

\* *Utilizar algoritmos estadísticos* para generar más datos a partir de una base.

## 2. Text Excraction and Clean-up

- Se hace un proceso de limpieza para que se pueda procesar. Algunos ejemplos son:

- HTML parsing and clean-up: Identificar partes del documento y eliminar etiquetas HTML.

- Unicode Normalization: Identificar y remplazar aquellos caracteres que no son letras.