# Metodología

A través de la página del Ministerio de Educación se descargó la base de datos de datos SNIES que contiene todos los programas educativos existentes en Colombia (que están activos). Complementando la información con la institución, la localización, el área de conocimiento y los respectivos códigos de identificación.

Teniendo en cuenta las columnas de delimitación del área de conocimiento, se filtró todos los programas que estuvieran relacionados a alguno de los siguientes términos de búsqueda: salud, ambiente, biología o natural. Al tener la base de datos filtrada, se unificó los criterios de búsqueda tomando como referencia el nombre de la institución y el programa.

Después de tener una base de datos exclusiva con los criterios, se automatizó el proceso de búsqueda y escrapeo en Google a través de la librería selenium de Python. Al hacerlo de esta manera, fue posible capturar cada página web existente en cada programa. De esta manera, se complementó la base de datos existente con la dirección web de cada programa.

El proceso de escrapeo de la información continúo en dos líneas. La primera fue automatizar la búsqueda de plan de estudios de cada programa para generar una base de datos con el contenido de cada programa. La segunda línea implicó copiar otra información relevante del programa (perfil del aspirante o egresado, descripción del programa) y adjuntarlo

Aunque a través del proceso se logró capturar gran parte de la información, fue importante un ajuste manual en todas las etapas ya que había páginas que por su propia arquitectura limitaban el uso de una herramienta automatizada.

# Resultados descriptivos de la base de datos

En total la base de datos del SNIES cuenta con 30094 programas diferentes distribuidos en 309 instituciones de educación superior. Al hacer el correspondiente filtrado, se redujo el número total de cursos a 720 (2% de la muestra), distribuido en 135 universidades. La distribución según el filtro por término de referencia fue:

|  |  |  |
| --- | --- | --- |
| Tipo de programas | Conteo | % |
| Ciencias biológicas | 228 | 32% |
| Medio ambiente | 191 | 27% |
| Salud | 244 | 34% |
| Ciencias naturales | 57 | 8% |
| total | 720 |  |

Gráfico

El contenido generado por IA puede ser incorrecto.