

# Guía de Usuario para la Aplicación de Extracción de Identificadores de FASTA

## Introducción

Esta herramienta está diseñada para extraer identificadores específicos de archivos FASTA obtenidos de UniProt y guardarlos en un archivo de texto. La aplicación es especialmente útil para bioinformáticos y otros investigadores que trabajan con grandes volúmenes de datos de secuencias proteicas.

## Propósito

La aplicación extrae identificadores en el formato `>sp|IDENTIFIER|` de archivos FASTA y los guarda en un archivo de texto. Esto facilita la organización y análisis de datos de secuencias.

## Formato de Entrada

La aplicación espera archivos FASTA en el formato estándar de UniProt, donde las líneas de descripción de las secuencias tienen la siguiente estructura:

```
>sp|P0C829|CA4B_CONST Conotoxin SIVB OS=Conus striatus OX=6493 PE=1 SV=1
MGMRMMFTVFLSVVLATTVVSTPSDRASDGRNAAVHERQKELVPSVITTCGGYDPGTMCP
PCRCTNSCPTKPKKPGRND
```

## Identificador Extraído

En el ejemplo anterior, el identificador extraído es `P0C829`.

## Uso de la Aplicación

## Ejecución en Windows

Para ejecutar la aplicación en Windows, simplemente haga doble clic en el archivo `extraerIDs.exe`.

## Interfaz Gráfica

### 1. \*\*Selección del Archivo FASTA\*\*

- Al ejecutar la aplicación, aparecerá un cuadro de diálogo para seleccionar el archivo FASTA de entrada. Navegue hasta el archivo deseado y selecciónelo.

### 2. \*\*Selección del Directorio de Salida\*\*

- Después de seleccionar el archivo FASTA, aparecerá un cuadro de diálogo para seleccionar el directorio donde se guardará el archivo de salida. Navegue hasta el directorio deseado y selecciónelo.

## Archivo de Salida

El archivo de salida será un archivo de texto llamado `identifiers.txt`, que contendrá todos los identificadores extraídos, uno por línea.

1. \*\*Archivo FASTA de Entrada\*\*:

```
>sp|P0C829|CA4B_CONST Conotoxin SIVB OS=Conus striatus OX=6493 PE=1 SV=1
MGMRMMFTVFLSVVLATTVVSTPSDRASDGRNAAVHERQKELVPSVITTCCGYDPGTMCP
PCRCTNSCPTKPKKPGRND
```

## Ejecución

Doble clic en `extraerIDs.exe`.

## Archivo de Salida

identifiers.txt

P0C829

## Contexto

Esta herramienta es particularmente útil en bioinformática para el análisis de datos de secuencias proteicas. Los identificadores extraídos pueden ser utilizados para acceder a información adicional en bases de datos, realizar anotaciones funcionales, o integrar datos en flujos de trabajo de análisis bioinformáticos.

## Detalles Técnicos

### 1. **Expresión Regular**:

- La aplicación utiliza la expresión regular `>sp\|(.+?)\|` para encontrar y extraer los identificadores en las líneas de descripción de las secuencias.

### 2. **Interfaz Gráfica**:

- Utiliza `tkinter` para mostrar cuadros de diálogo para la selección de archivos y directorios.

## Conclusión

Esta herramienta proporciona una forma sencilla y eficiente de extraer identificadores específicos de archivos FASTA de UniProt, ayudando a los investigadores a manejar y analizar grandes conjuntos de datos de secuencias proteicas de manera organizada.