

# **Pubmed Separador Guía de Usuario**

**Instituto de Investigaciones en  
Matemáticas Aplicadas y en Sistemas**

**Programa: Minería de Datos Aplicada  
a Sistemas Biológicos**

**Reyes Bolaños Martín  
Marzo de 2021**

## Índice

Pubmed-Batch-Download .....	1
Descripción .....	1
Instalación .....	1
Pubmed Separador .....	2
Parámetros del archivo.....	2
Funcionamiento del programa .....	3

## **Pubmed-Batch-Download**

### **Descripción**

Este software elaborado por Bill Greenwald es indispensable para el funcionamiento de este proyecto, ya que nos ayuda a descargar documentos PDF provenientes de la plataforma Pubmed con tan solo introducir el ID del artículo.

### **Instalación**

El enlace para consultar dicho proyecto a través de Github es el siguiente:  
<https://github.com/billgreenwald/Pubmed-Batch-Download>.

Después que se haya descargado es importante tener alguna versión instalada de Anaconda para su correcta ejecución.

Una vez tengamos descargados e instalados dichos componentes es necesario abrir una terminal y ejecutar el siguiente comando:

```
Linux: conda env create -f pubmed-batch-downloader-py3.yml  
Windows: conda env create -f pubmed-batch-downloader-py3-windows.yml
```

Para activar el entorno usamos la siguiente sentencia:

```
conda activate pubmed-batch-downloader-py3
```

En caso de utilizar el Sistema Operativo Windows es necesario seguir los siguientes comandos para instalar otros paquetes no incluidos en la instalación predeterminada:

```
conda install requests beautifulsoup4 lxml  
conda install requests3
```

## Pubmed Separador

Después de instalar correctamente el programa anterior en nuestro ordenador, es necesario que en la carpeta Pubmed-Batch-Download-master se incluya el archivo PubmedSeparador.py para que el programa se ejecute de manera correcta.

### Parámetros del archivo

Este programa utiliza tres parámetros, de los cuáles dos de ellos requieren leer archivos excel (.xlsx) con un nombre específico en sus columnas para extraer la información importante de manera correcta:









- En el apartado 'Factores' utiliza la información general sobre los genes, para esto se necesitan tres columnas obligatorias (Locus Tag, Gene Name y Product Synonyms ), mientras que existen dos opcionales, las cuales deben llevar el nombre de 'Type1' y 'Type2' en caso de que también exista información sobre la interacción los genes.

Locus Tag	Gene Name	Product Synonyms	Type1	Type2
PA4147	acoR		TF	TG
PA0601	agtR	two-component response regulator, AgtR	TF	TG
PA0294	aguR	transcriptional regulator AguR	TF	TG
PA5483	algB		TF	TG
PA5255	algQ	transcriptional regulatory protein AlgR2	TF	TG
PA5261	algR	AlgR1	TF	TG
PA4446	algW		TF	TG
PA5200	amgR	AmgR	TF	TG

- Keyword: Este segundo apartado se encarga de extraer las palabras que indican alguna interacción entre una o varios genes. Para esto el archivo tipo excel solo debe tener una columna denominada 'KEYWORD'

KEYWORD
activation
repression
regulation
activate
repress
regulated
activator
repressor
regulate
inhibition
Promoter
enhancer
binding sites
operon
receptor
cis-regulatory

- ID - PDF: En este último apartado solo es necesario introducir el ID del archivo localizado en la plataforma de PUBMED con el fin de descargarlo a la carpeta `fetched_pdfs` y leer su contenido.

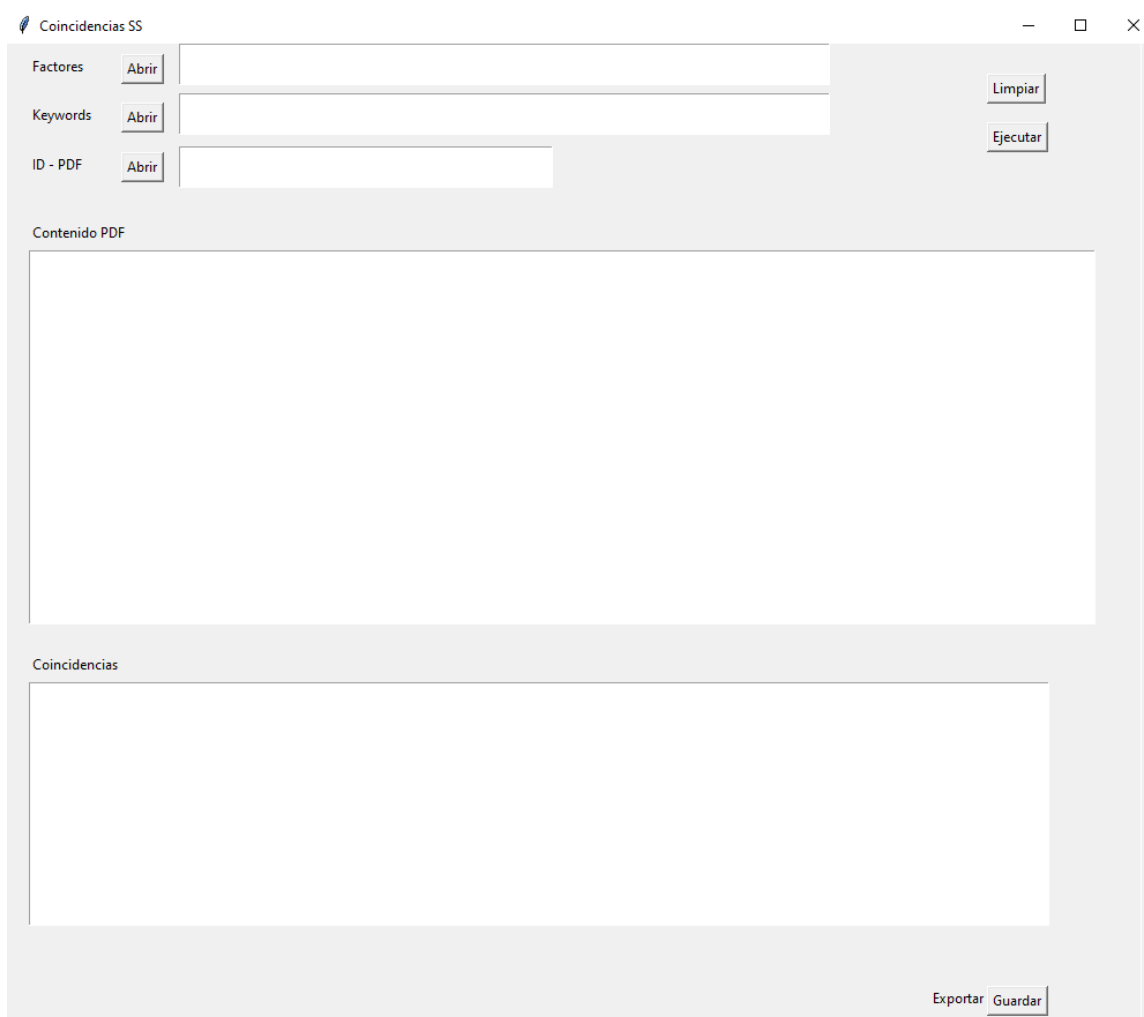
Este equipo > Escritorio > Pubmed-Batch-Download-master > <b>fetched_pdfs</b>						
Nombre		Fecha de modificación	Tipo	Tamaño		
	8930909.pdf	02/03/2021 10:26 p. m.	Foxit Reader PDF ...	117 KB		
	9150205.pdf	02/03/2021 10:29 p. m.	Foxit Reader PDF ...	186 KB		
	9645431.pdf	02/03/2021 10:32 p. m.	Foxit Reader PDF ...	63 KB		
	10564474.pdf	02/03/2021 10:33 p. m.	Foxit Reader PDF ...	51 KB		
	10894749.pdf	02/03/2021 10:34 p. m.	Foxit Reader PDF ...	96 KB		
	11092854.pdf	02/03/2021 10:36 p. m.	Foxit Reader PDF ...	3,065 KB		
	11381121.pdf	02/03/2021 11:00 p. m.	Foxit Reader PDF ...	207 KB		
	19243444.pdf	04/03/2021 12:26 p. m.	Foxit Reader PDF ...	234 KB		

### Funcionamiento del programa

Para ejecutar el programa es necesario abrir una terminal ubicada en la carpeta *Pubmed-Batch-Download-master* y ejecutar el siguiente comando:

python pubmedSeparador.py

A los pocos segundos se abrirá la siguiente ventana, el cual se procederá a explicar a detalle



En la primera parte podemos ver tres botones con los nombres factores, Keywords y ID-PDF, en los cuales se buscará la ruta de los archivos tipo excel que contienen la información ya explicada anteriormente.

Factores	<input type="button" value="Abrir"/>	C:/Users/MREYESCO/Desktop/Pubmed-Batch-Download-master/Factores.xlsx
Keywords	<input type="button" value="Abrir"/>	C:/Users/MREYESCO/Desktop/Pubmed-Batch-Download-master/keywords.xlsx
ID - PDF	<input type="button" value="Abrir"/>	9150205

Una vez se escriba el ID en el tercer apartado y le demos click al botón Abrir, se descargará el documento y lo desplegará en el apartado Contenido PDF en caso de que no exista ningún problema.

ID - PDF	<input type="button" value="Abrir"/>	9150205
----------	--------------------------------------	---------

Contenido PDF

with yet another mechanism to temporally control the activation of important factors.

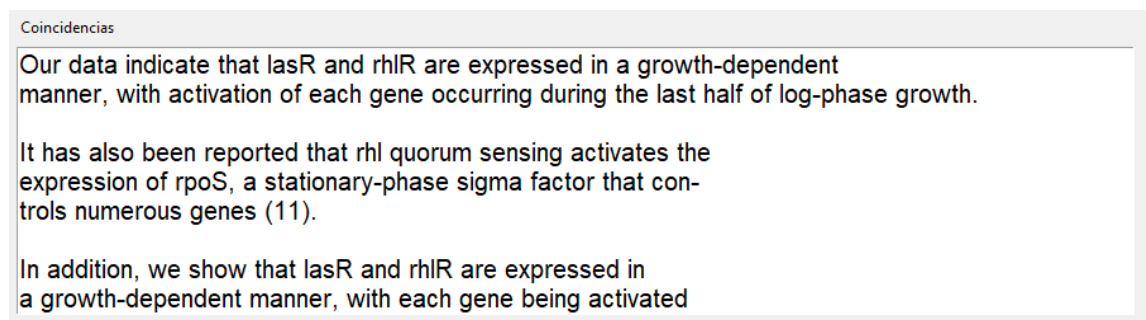
The control of RhlR at both the transcriptional and post-translational levels provides *P. aeruginosa* with an elegant mechanism through which it can use the las quorum-sensing system to control the rhl quorum-sensing system. The importance of this discovery becomes evident when one considers both the multiple virulence factors controlled by these two systems and the report (11) that RhlR-PAI-2 activates transcription of the sigma factor rpoS. The results presented here also lead us to speculate that it may be possible to develop a single therapeutic autoinducer analog of PAI-1, rather than two different analogs, which will inhibit expression of virulence genes controlled by both LasR and RhlR.

ACKNOWLEDGMENTS

E. C. Pesci was supported by Research Fellowship Grant PESCI96FO from the Cystic Fibrosis Foundation. P. C. Seed was supported by a training grant (5T32AI07362-07) in microbial pathogenesis from the National Institutes of Health. This work was also supported by National Institutes of Health research grant R01AI33713-04.




En la esquina superior derecha aparecen dos botones: limpiar nos ayuda a eliminar el contenido de todos los apartados, mientras que Ejecutar realiza el algoritmo encargado de extraer la información más importante de todo el artículo dependiendo de las palabras


encontradas en los documentos utilizados para extraer palabras clave, los resultados aparecen en la sección Coincidencias.





Finalmente, en la esquina inferior derecha aparece el botón Guardar la cual crea un archivo .txt en la carpeta Coincidencias (este directorio se crea en caso de que no exista) con la información mostrada en el apartado anterior en caso de que lo consideremos importante y se desee consultarla en otro momento.



equipo > Escritorio > Pubmed-Batch-Download-master > **Coincidencias**    Buscar en Coincid...

Nombre	Fecha de modificación	Tipo	Tamaño
 9150205.txt	16/03/2021 10:58 p. m.	Documento de te...	2 KB

 9150205.txt: Bloc de notas 

Archivo Edición Formato Ver Ayuda

Our data indicate that lasR and rhIR are expressed in a growth-dependent manner, with activation of each gene occurring during the last half of log-phase growth

It has also been reported that rhI quorum sensing activates the expression of rpoS, a stationary-phase sigma factor that controls numerous genes (11).

In addition, we show that lasR and rhIR are expressed in a growth-dependent manner, with each gene being activated during the second half of log-phase growth.

6., IPTG was omitted because the tac promoter controlling rhIR allows enough RhIR to be produced to activate rhIA in the presence of PAI-2.

It has been shown that PAI-2 and RhIR are required for the activation of rhIA in E. coli (24) and that rhIR is required for rhamnolipid production in P. aeruginosa (16).

In our experiments, we found that activation of rhIR and lasR in strain PA01 was very similar (Fig.