

Contenido

[INSTALACIÓN](#)

[WINDOWS](#)

[MAC OS](#)

[CONVENCIONES DE LOS EJERCICIOS](#)

[USO BÁSICO](#)

[1. CARGA DE DATOS](#)

[FACETING](#)

[FILTERING](#)

[CLUSTERING](#)

[EXPORTACIÓN](#)

[USO AVANZADO](#)

[USO DEL API DE EOL](#)

[ENLACES ADICIONALES](#)

INSTALACIÓN

ANTES DE COMENZAR

1. Instale [Java JRE](#) si no lo tiene.
2. Diríjase a la [sección de descargas](#) del sitio web de *OpenRefine* y realice la descarga de acuerdo a su sistema operativo:

WINDOWS

1. Un archivo ZIP se descargará en su computador
2. Extraiga el archivo a una carpeta nueva y nómbrala **OpenRefine**
3. Ingrese en la carpeta y ejecute la aplicación:


Si *openrefine.exe* no funciona intente con *openrefine.bat*, se abrirá automáticamente una ventana de línea de comandos y luego la interfaz del programa en el navegador de Internet. **No utilice Internet Explorer**, copie y pegue la dirección en otro navegador de ser necesario.

MAC OS

1. Un archivo DMG se descargará en su computador
2. Abra el archivo y arrastre el icono del programa a la carpeta "Aplicaciones":
3. Ejecute la aplicación y se abrirá el programa en el navegador de Internet.

Se abrirá automáticamente la interfaz del programa en el navegador de Internet. **No utilice Safari**, copie y pegue la dirección en otro navegador de ser necesario.

CONVENCIONES DE LOS EJERCICIOS

| | |
|-------------------------------|---|
| Fórmulas para copiar y pegar | cell.recon.match.id |
| Comandos en OpenRefine | Edit column |
| Nombres de las columnas | nombreRecon |
| Enlaces a sitios informativos | www.sibcolombia.net |
| Menú Columna |  |

USO BÁSICO

Al terminar cada ejercicio recuerde hacer clic en el botón **Remove todos** que se encuentra debajo del logo de OpenRefine

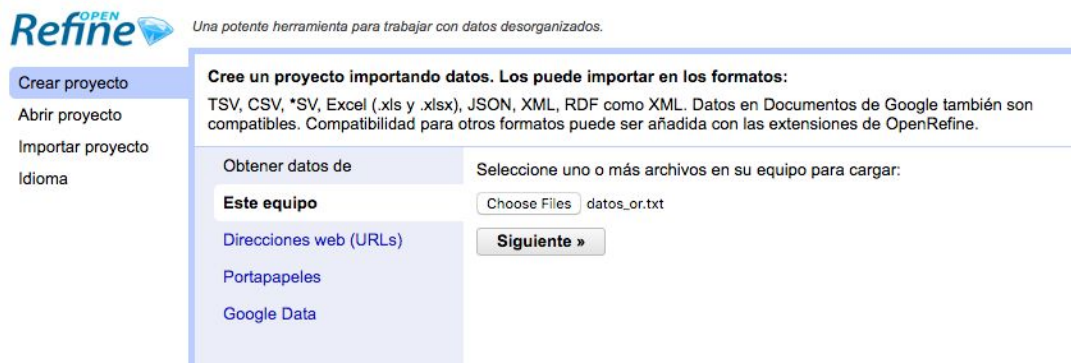
1. CARGA DE DATOS

ANTES DE COMENZAR

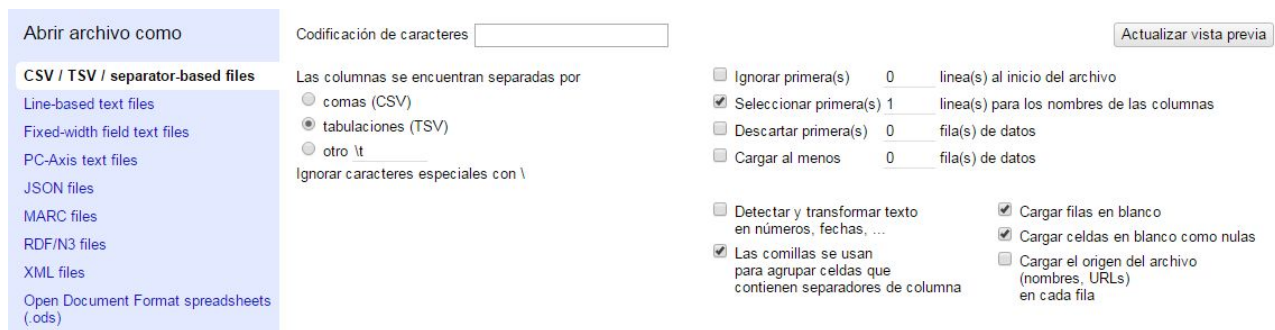
Puede seleccionar el idioma de su preferencia en [Language Settings](#) o [Idioma](#), esta guía utiliza español como lenguaje de la herramienta.

CREAR UN PROYECTO

1. Descargue el archivo de ejemplo **datos_or**
2. En la ventana de *OpenRefine* vaya a [Crear Proyecto](#). Seleccione su equipo y cargue el archivo con sus datos y haga clic en [Siguiete](#).



3. Un panel de selección aparecerá, este le permite especificar el tipo de datos que se cargan y configurar la manera en la que los datos son leídos. Deje las opciones tal y como se muestran en la imagen:



Nota: si en la vista previa encuentra caracteres extraños (p.ej: símbolos como ♦ en lugar de las tildes) pruebe otra codificación de caracteres en el cuadro **Codificación de caracteres**.

4. En la esquina superior derecha verá un cuadro de texto en el que puede cambiar el nombre del proyecto, haga clic en el botón **Crear Proyecto** y espere a que cargue el archivo:




Nombre del proyecto

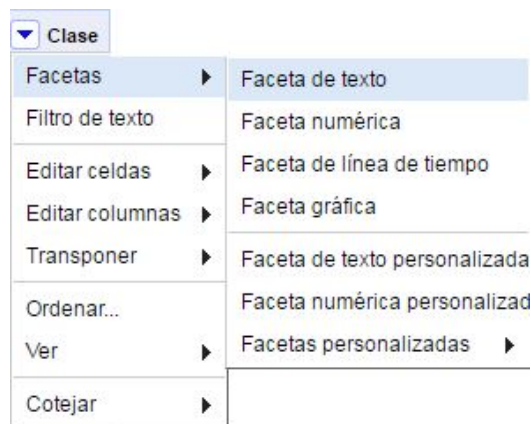
2. FACETAS

ANTES DE EMPEZAR

Las facetas permiten filtrar los datos en conjuntos más pequeños para facilitar su uso y análisis, pueden hacerse para texto, números y fechas.

EJERCICIO 1. Facetas y correcciones masivas

1. Diríjase a la columna **Clase**, haga clic en  y siga la ruta que se muestra en la imagen para hacer una **Faceta de texto**:



2. A su izquierda aparecerá una ventana con el nombre de la columna y las facetas encontradas:




Haga clic en **conteo** para organizar las clases de la más a la menos abundante y en **A-Z** para organizarlas en orden alfabético.

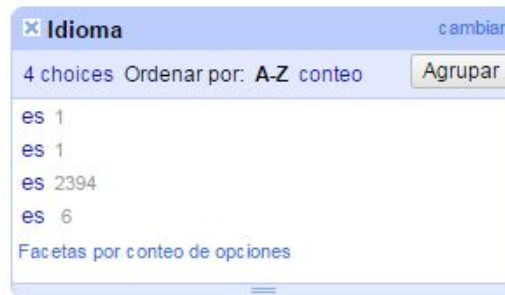
3. Corrija las inconsistencias en los nombres de las Clases Ave (**Aves**) y Mamalia (**Mammalia**). Para esto acerque el cursor al valor que desea corregir y haga clic en **editar**, luego en el cuadro de texto

que aparece corrija el error y haga clic en **Aplicar**, todos los valores serán corregidos de manera automática:



EJERCICIO 2. Facetas y espacios en blanco

1. Diríjase a la columna **Idioma**, haga clic en  y realice una **Faceta de texto**.
2. A su izquierda aparecerá la ventana con el nombre de la columna y las facetas encontradas:




Al realizar este procedimiento vemos que el programa ha detectado espacios extra y por eso nos muestra cuatro opciones diferentes.

3. Corrija las inconsistencias desde el Menú Columna de **Idioma**, siguiendo la ruta **Editar Celdas > Transformaciones comunes > Quitar espacios al inicio y al final**, verá un mensaje de notificación:

Text transform on 8 cells in column Idioma: value.trim() Deshacer

2. Si observa la ventana de las facetas de **Idioma**, notará que ahora solo existe una opción con 2402 registros y que los espacios fueron eliminados.

EJERCICIO 3. Facetas y duplicados



1. Diríjase a la columna **Número de Catálogo**, haga clic en  y siga la ruta **Facetas > Facetas personalizadas > Faceta por duplicados**.
3. A su izquierda verá la ventana de las facetas encontradas:

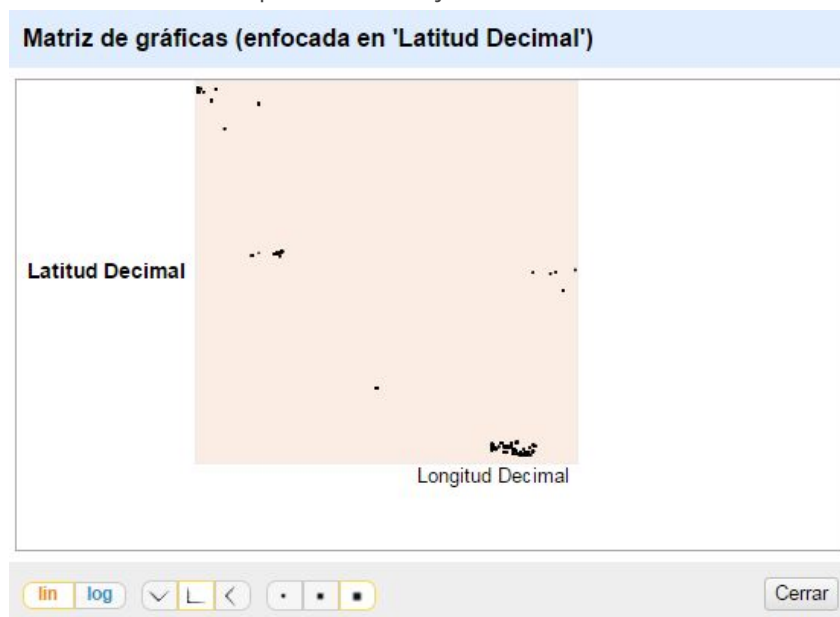


Podemos ver que el programa ha detectado valores únicos (false) y valores duplicados (true).

4. Haga clic en **true** y verá los registros. De esta manera se pueden detectar los duplicados para un análisis posterior.

EJERCICIO 4. **Faceta gráfica**

1. Diríjase a la columna **Latitud Decimal**, haga clic  y siga la ruta **Editar celdas > Transformaciones comunes > A número**. De esta manera convertirá los valores de esta columna a números que el programa puede reconocer. Realice el mismo procedimiento para la columna **Latitud Decimal**.
2. Diríjase nuevamente a la columna **Latitud Decimal**, haga clic en  y siga la ruta **Facetas > Faceta gráfica**.
3. Aparecerá una ventana con las posibles relaciones numéricas de la columna seleccionada y las demás columnas que contienen valores numéricos. Seleccione las opciones de los recuadros amarillos de la imagen y luego haga clic en la gráfica que se muestra. De esta manera aparecerá la ventana del Facet en nuestro espacio de trabajo.




5. Diríjase a la columna **Municipio**, haga clic en  y realice un una **Faceta de texto**, de esta manera OpenRefine mostrará:

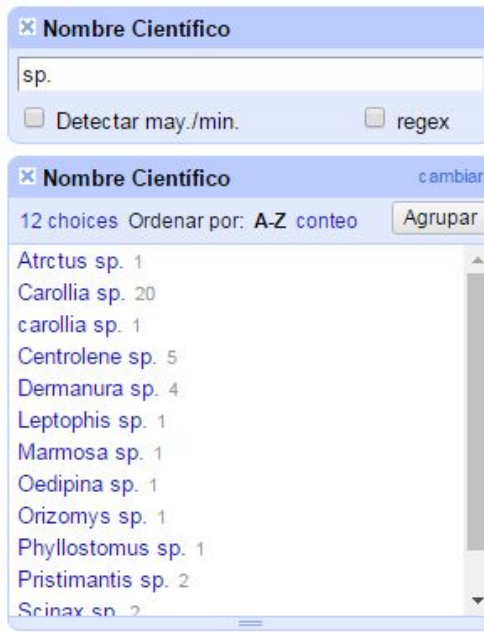



6. Gracias a esto puede realizar un análisis preliminar de las coordenadas fuera de lugar en una gran cantidad de registros. Haga clic en cada uno de los municipios para ver como estos se agrupan, de la misma manera puede seleccionar los grupos formados directamente en la gráfica. Al realizar esto se irán visualizando los registros que seleccione en la tabla de datos. Si desea ver la grafica mas grande aumente el zoom del navegador o haga clic en **Exportar gráfica**.
7. Seleccione el municipio de **Salaminas** y observe como forma dos grupos en la gráfica.
8. Seleccione directamente en la gráfica el grupo de la parte superior izquierda y observe como cinco registros presentan coordenadas extrañas.
9. Para volver a ver todos los registros haga clic en **restaurar** en las dos ventanas,

3. FILTROS

EJERCICIO 5. Filtro básico

1. Diríjase a la columna **Nombre Científico**, haga clic en  y luego en **Filtro de texto**, aparecerá la ventana del Filtro. Escriba en el campo de texto **sp.** y realice una **Faceta de texto** en **Nombre Científico** para visualizar los registros con este valor:



2. Para corregir este error vaya a  en **Nombre Científico** y siga la ruta **Editar celdas > Transformar...**, ingrese la fórmula `value.replace(" sp.", "")` tal y como se muestra a continuación:

Transformación personalizada en Nombre Científico

Expresión Lenguaje

`value.replace(" sp.", "")` No hay error de sintaxis.

Vista previa Historial Con estrella Ayuda

| row | value | value.replace(" sp.", "") |
|------|--------------|---------------------------|
| 118. | Atrctus sp. | Atrctus |
| 494. | Carollia sp. | Carollia |
| 495. | Carollia sp. | Carollia |
| 496. | Carollia sp. | Carollia |
| 497. | Carollia sp. | Carollia |
| 498. | Carollia sp. | Carollia |
| 499. | Carollia sp. | Carollia |

En error mantener original Re-transformar hasta veces hasta que no haya cambios
 cambiar a en blanco
 guardar error

Haga clic en **Aceptar** y verá el mensaje de confirmación de los cambios realizados.

EJERCICIO 6. Filtro avanzado I

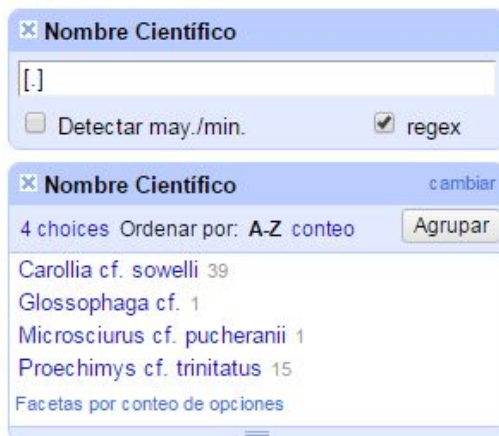
1. Diríjase a la columna **familia**, haga clic en el Menú Columna y luego en **Filtro de texto**, aparecerá la ventana del Filtro.
2. Realice una **Faceta de texto** en **familia** para visualizar los registros con este elemento:
3. Marque la casilla **regex**. Pegue en el campo de texto la expresión regular `((?!ae).)$` esta expresión nos permite excluir todas las palabras de la columna que no terminan en "ae".




Podrá observar como los registros que no corresponden a familias han sido filtrados, usted puede editarlos o marcarlos para hacer una revisión posterior de los datos. Para conocer más de las expresiones regulares haga clic [aquí](#).

EJERCICIO 7. Filtro avanzado II

1. Diríjase a la columna **Nombre Científico**, haga clic en el Menú Columna realice un **Filtro de texto**.
2. Realice una **Faceta de texto** en **Nombre Científico** para visualizar los registros con este elemento.
3. Marque la casilla **regex**. Pegue en el campo de texto la expresión regular `[.]`



4. Para editarlo masivamente vaya a  en **Nombre Científico** y siga la ruta **Editar celdas > Transformar...**, ingrese la fórmula `value.replace(" cf.", "")` tal y como se muestra a continuación:

Transformación personalizada en Nombre Científico

Expresión Lenguaje General Refine Expression Language (GREL) ▾

```
value.replace(" cf.", "")
```

No hay error de sintaxis.

[Vista previa](#) [Historial](#) [Con estrella](#) [Ayuda](#)

| row | value | value.replace(" cf.", "") |
|------|----------------------|---------------------------|
| 299. | Carollia cf. sowelli | Carollia sowelli |
| 300. | Carollia cf. sowelli | Carollia sowelli |
| 301. | Carollia cf. sowelli | Carollia sowelli |
| 302. | Carollia cf. sowelli | Carollia sowelli |
| 303. | Carollia cf. sowelli | Carollia sowelli |
| 304. | Carollia cf. sowelli | Carollia sowelli |
| 305. | Carollia cf. sowelli | Carollia sowelli |

En error
 mantener original
 Re-transformar hasta veces hasta que no haya cambios
 cambiar a en blanco
 guardar error

Haga clic en **Aceptar** y verá el mensaje de confirmación de los cambios realizados.

4. AGRUPACIONES

EJERCICIO 8. Agrupación básica

- Diríjase a la columna **Colector**, y realice una **Faceta de Texto**, aparecerá la ventana del Facet:

✕ **Colector**
cambiar

7 choices Ordenar por: **A-Z** conteo

Ana gonzales 1

Ana Gonzales 1

G. Ricardo 1

GonzalesAna 1

ricardo g 40

Ricardo G. 1

Ricardo G 15

(blank) 2342

Facetas por conteo de opciones

- En la parte superior derecha verá el botón **Agrupar**. haga clic y aparecerá la ventana de **Agrupar y editar valores** para la columna **Colector**.

Agrupar y editar valores en la columna "Colector"

Esta función le permite encontrar agrupaciones de diferentes valores que pueden ser representaciones alternativas de la misma cosa. Por ejemplo, "New York" y "new york" probablemente se refieren al mismo concepto, solo se presenta diferencia en la capitalización. De la misma manera "Gödel" y "Godel" probablemente se refieren a la misma persona. [Más información ...](#)

Método Colisión de llaves

Función Huella

2 clusters Encontrado

| Número de valores | Número de filas | Valores en la agrupación | ¿Unir? | Nuevo valor de las celdas | # Valores en la agrupación |
|-------------------|-----------------|---|--------------------------|---|---|
| 3 | 42 | <ul style="list-style-type: none">ricardo g (40 rows)G. Ricardo (1 rows)Ricardo G. (1 rows) | <input type="checkbox"/> | <input type="text" value="ricardo g"/> | <input type="text" value="2 — 3"/> |
| 2 | 2 | <ul style="list-style-type: none">Ana gonzales (1 rows)Ana Gonzales (1 rows) | <input type="checkbox"/> | <input type="text" value="Ana gonzales"/> | <input type="text" value="2 — 42"/> |
| | | | | | <input type="text" value="Longitud promedio de los valores"/> |
| | | | | | <input type="text" value="9.66 — 13.5"/> |
| | | | | | <input type="text" value="Varianza de los valores"/> |
| | | | | | <input type="text" value="0.47000000000000003 — 1.5"/> |

Seleccionar todos

Seleccionar ninguno

Exportar agrupaciones

Unir seleccionados y reagrupar

Unir seleccionados y cerrar

Cerrar

Nota: en esta ventana podrá ver la siguiente información

Número de valores: la cantidad de diferentes representaciones que el algoritmo muestra como similares.

Número de filas: El número de registros por agrupación.

Valores en la agrupación: Los valores textuales seleccionados por el algoritmo para esa agrupación y el número de registros por valor.

Unir?: Es el cuadro de selección para convertir todos los valores en el valor seleccionado.

New cell value: Es el valor seleccionado para convertir los valores. Puede escribir un valor completamente diferente o hacer clic en cualquier valor para asignarlo como valor seleccionado.

3. Vaya a **Función**, seleccione **Huella del n-grama** y en **Tamaño ngrama** escriba **1**

Agrupar y editar valores en la columna "Colector"

Esta función le permite encontrar agrupaciones de diferentes valores que pueden ser representaciones alternativas de la misma cosa. Por ejemplo, "New York" y "new york" probablemente se refieren al mismo concepto, solo se presenta diferencia en la capitalización. De la misma manera "Gödel" y "Godel" probablemente se refieren a la misma persona. [Más información ...](#)

Método Función Tamaño ngrama 2 clusters Encontrado

| Número de valores | Número de filas | Valores en la agrupación | ¿Unir? | Nuevo valor de las celdas |
|-------------------|-----------------|---|--------------------------|---|
| 4 | 57 | <ul style="list-style-type: none">ricardo g (40 rows)Ricardo G (15 rows)G. Ricardo (1 rows)Ricardo G. (1 rows) | <input type="checkbox"/> | <input type="text" value="ricardo g"/> |
| 3 | 3 | <ul style="list-style-type: none">Ana gonzales (1 rows)Ana Gonzales (1 rows)GonzalesAna (1 rows) | <input type="checkbox"/> | <input type="text" value="Ana gonzales"/> |

Valores en la agrupación:

Filas en la agrupación:

Longitud promedio de los valores:

Varianza de los valores:

Para conocer más acerca de los algoritmos (altamente recomendado) haga clic [aquí](#).

- Para la primera agrupación asigne un valor nuevo, para esto vaya al cuadro de texto de **Nuevo valor de las celdas** y escriba **Ricardo Gonzalez**. Luego marque el cuadro de **Unir?**: para esta agrupación.
- Para la segunda agrupación haga clic en **Ana Gonzales**, automáticamente el valor en **Nuevo valor de las celdas** cambiará y la casilla **Unir?** se marcará.
- Haga clic en **Unir seleccionados y cerrar** para fusionar los valores y volver a la ventana principal. El resultado del proceso debería verse así:



5. EXPORTACIÓN

EJERCICIO 9. Exportar el archivo

Existen múltiples maneras de exportar los archivos en OpenRefine, la siguiente es la que ha mostrado más utilidad en la mayoría de los casos

- En la esquina superior derecha haga clic en el botón **Exportar**
- Seleccione **Configurar exportación...** aparecerá la ventana de exportación:

Configurar exportación

Contenido Descarga Cargar Código

Seleccione y ordene las columnas a exportar

- ID
- Idioma
- Base del registro
- Número de Catálogo
- Colector
- country
- Código del País
- Departamento
- Municipio

Optpciones para **ID**

Para celdas cotejadas, descargar

- Nombre cotejado
- ID cotejado
- Enlace a la página cotejada
- El contenido de las celdas
- Descargar nada para celdas sin correspondencia
- ISO 8601, p. ej., 2011-08-24T18:36:10+08:00
- Formato corto local
- Formato largo local
- Personalizado
- Formato mediano local
- Formato local completo
- Usar zona horaria local
- Omitir hora/fecha

Incluir encabezados de columnas
 Incluir filas en blanco (p. ej. todas las celdas nulas)
 Ignorar facetas y filtros filters y exportar todas las filas

3. En la pestaña **Contenido** puede seleccionar las columnas que quiere exportar, si selecciona **Ignorar facetas, filtros y exportar todas las filas**, todos los facets y filtros serán ignorados. Esto es útil cuando solo queremos hacer visualizaciones de los datos.
4. Vaya a la pestaña **Descarga** y seleccione el separador de caracteres que desee, en este caso se recomienda dejar las opciones que vienen por defecto. Haga clic en **Descargar** y guarde su archivo.
5. También puede exportar el proyecto (archivo para OpenRefine) siguiendo la ruta **Exportar > Exportar proyecto**. De esta manera puede descargar el proyecto para continuar trabajando en otro equipo.

USO AVANZADO

1. USO DEL API DE EOL

ANTES DE EMPEZAR

En este ejercicio se utilizará el API de la *Encyclopedia of Life* (EOL) a través de OpenRefine para obtener las categorías taxonómicas de una lista de nombres determinada.

Nota: EOL agrupa las clasificaciones de los grupos de organismos de diversos proveedores de contenido, cada uno de las cuales es soportado por una comunidad de científicos. Para una lista completa de los proveedores y descripción de los mismos ingrese a:

http://eol.org/api/docs/provider_hierarchies

Entre los proveedores de clasificaciones se encuentran:

AntWeb, Avibase, CoL, FishBase, Index Fungorum, ITIS, IUCN, Metalmark Moths of the World, NCBI, The Paleobiology Database, The Reptile Database, WoRMS.

EJERCICIO 7. Obtención de las categorías

1. Vaya a la columna **Nombre Científico**. Es importante que estos nombres no contengan calificadores de cómo "cf.", "sp." o "spp.", de ser este el caso elimínelos.
3. Haga clic en el menú columna y luego siga la ruta **Editar columnas > Agregar columna basada en esta columna...**, en el cuadro **Nuevo nombre de la columna** escriba **nombreRecon**. No realice ningún otro cambio.
4. El siguiente paso es reconciliar los nombres con EOL, utilizando el servicio implementado por Roderic D.M. (<http://iphylo.blogspot.com>). Para esto haga clic en el Menú Columna de **nombreRecon**, y siga la ruta **Cotejar > Iniciar...**
5. Haga clic en **Agregar servicio estándar...** y pegue la siguiente dirección:
http://iphylo.org/~rpage/phyloinformatics/services/reconciliation_eol.php
6. Haga clic en **Add Service** y espere un momento. En la lista de servicios ahora podrá encontrar Encyclopedia of Life, selecciónelo.

Cotejar columna "nombreRecon" » Ir a API del servicio

Cotejar cada celda con los valores de una de estas clases: Usar también detalles relevantes de otras columnas:

Scientific name
/biology/organism_classification/scientific_name

| columna | ¿Incluir? | Como propiedad |
|--------------------|--------------------------|----------------------|
| ID | <input type="checkbox"/> | <input type="text"/> |
| Idioma | <input type="checkbox"/> | <input type="text"/> |
| Base del registro | <input type="checkbox"/> | <input type="text"/> |
| Número de Catálogo | <input type="checkbox"/> | <input type="text"/> |
| Colector | <input type="checkbox"/> | <input type="text"/> |
| country | <input type="checkbox"/> | <input type="text"/> |
| Código del País | <input type="checkbox"/> | <input type="text"/> |
| Departamento | <input type="checkbox"/> | <input type="text"/> |
| Municipio | <input type="checkbox"/> | <input type="text"/> |
| Localidad | <input type="checkbox"/> | <input type="text"/> |
| Elevación Mínima | <input type="checkbox"/> | <input type="text"/> |

Cotejar contra la clase:

Reconciliar contra ninguna clase en particular

Cotejar automáticamente candidatos con alta confianza

Máximo número de candidatos a devolver

7. No haga ninguna modificación, haga clic en **Cotejar** y verá la notificación de inicio:

**Reconcile cells in column nombreRecon to type
/biology/organism_classification/scientific_name**
3% hecho

8. Luego de un rato, dependiendo del número de nombres y la conexión a internet, verá un resultado como el siguiente:

Facetas / Filtros Deshacer / Rehacer 10 **2402 filas**

Actualizar Restablecer todos Remover todos

Mostrar como: **filas** registros Mostrar: 5 10 25 50 filas

| Todo | Nombre Científico | nombreRecon |
|------|--------------------|---|
| 1. | Actitis macularius | Actitis macularius <input checked="" type="checkbox"/> Actitis macularius (Linnaeus, 1766) (1) <input checked="" type="checkbox"/> Actitis macularius (Linnaeus, 1766) (1) <input checked="" type="checkbox"/> Crear nuevo elemento |
| 2. | Alouatta seniculus | Alouatta seniculus <input checked="" type="checkbox"/> Alouatta seniculus (Linnaeus, 1766) (1) <input checked="" type="checkbox"/> Alouatta macconnelli Elliot 1910 (1) <input checked="" type="checkbox"/> Alouatta macconnelli Elliot 1910 (1) <input checked="" type="checkbox"/> Crear nuevo elemento |
| 3. | Alouatta seniculus | Alouatta seniculus <input checked="" type="checkbox"/> Alouatta seniculus (Linnaeus, 1766) (1) <input checked="" type="checkbox"/> Alouatta macconnelli Elliot 1910 (1) <input checked="" type="checkbox"/> Alouatta macconnelli Elliot 1910 (1) <input checked="" type="checkbox"/> Crear nuevo elemento |
| 4. | Alouatta seniculus | Alouatta seniculus <input checked="" type="checkbox"/> Alouatta seniculus (Linnaeus, 1766) (1) <input checked="" type="checkbox"/> Alouatta macconnelli Elliot 1910 (1) <input checked="" type="checkbox"/> Alouatta macconnelli Elliot 1910 (1) <input checked="" type="checkbox"/> Crear nuevo elemento |

nombreRecon: judgment cambiar
 2 choices Ordenar por: **A-Z** conteo
 (blank) 13
 none 71
 Facetas por conteo de opciones

nombreRecon: best candidate's score cambiar restaurar
 1.00 — 1.00
 Numérico 2318 Non-numeric 0 Blank 0 Error 84

9. Como se puede observar algunos nombres aparecen en color azul y otros en negro. Para aceptar el mejor candidato del cotejo automáticamente, vaya al menú Columna de **nombreRecon** y siga la ruta **Cotejar > Acciones > Asociar cada celda con su mejor candidato**

Match each of 2318 cells to its best candidate in column nombreRecon Deshacer

10. Notará que existen filas sin valor puesto que desde un principio no tenían nombre científico documentado. Remueva todas las facetas haciendo clic en **Remover todos** y en el menú columna de **nombreRecon** siga la ruta **Cotejar > Facetas > Por parámetro**

nombreRecon: judgment cambiar
 3 choices Ordenar por: **A-Z** conteo
 (blank) 13
 matched 2318
 none 71
 Facetas por conteo de opciones

Seleccione **matched** para trabajar únicamente con las filas cuyo nombre científico fue cotejado.

11. El siguiente paso es obtener el ID que asigna EOL para cada taxón, vaya al Menú Columna de **nombreRecon** y siga la ruta **Editar columna > Agregar columna basada en esta columna...**, en la ventana que aparece pegue en el campo de texto **Expresión** la expresión **cell.recon.match.id** y nombre la columna como **EOL_ID**

Agregar columna basada en la columna nombreRecon

Nuevo nombre de la columna

cambiar a en blanco guardar error copiar valor de la columna original

Expresión Lenguaje No hay error de sintaxis.

Vista previa

| row | value | cell.recon.match.id |
|-----|--------------------|---------------------|
| 1. | Actitis macularius | 1064984 |
| 2. | Alouatta seniculus | 323922 |
| 3. | Alouatta seniculus | 323922 |
| 4. | Alouatta seniculus | 323922 |
| 5. | Alouatta seniculus | 323922 |
| 6. | Amazilia amabilis | 1048672 |
| 7. | Amazilia amabilis | 1048672 |

12. Ahora se utilizará el API de EOL para obtener una lista de proveedores en los que se puede encontrar la taxonomía de cada taxón, vaya al Menú Columna de **EOL_ID** y siga la ruta **Editar columna > Agregar columna accediendo a URLs...** e introduzca la siguiente fórmula tal y como aparece:
- ```
"http://eol.org/api/pages/1.0/"+value+".json?images=2&videos=0&sounds=0&maps=0&text=2&iucn=false&subjects=overview&licenses=all&details=false&common_names=false&synonyms=false&references=false&vetted=0&cache_ttl="
```

Cambie el nombre de la columna a **json\_proveedores** y en el campo **Tiempo de retraso** escriba 250, haga clic en **Aceptar** y tenga paciencia ya que la consulta puede tomar un tiempo.



## Agregar columna accediendo a URIs basada en la columna EOL\_ID

Nuevo nombre de la columna  Tiempo de retraso   
milisegundos

En error  cambiar a en blanco  guardar error

Ingrese las URLs a acceder:

Expresión  Lenguaje

```
"http://eol.org/api/pages/1.0/" + value + ".json?
images=2&videos=0&sounds=0&maps=0&text=2&iucn=false&subjects=overview&licens
es=all&details=false&common_names=false&synonyms=false&references=false&vett
ed=0&cache_ttl="
```

Vista previa Historial Con estrella Ayuda

| row | value                                                                                                                                                                                                                          |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|     | "http://eol.org/api/pages/1.0/" + value + ".json?<br>images=2&videos=0&sounds=0&maps=0&text=2&iucn=false&subjects=overview&licenses=all&details=false&common_names=false&synonyms=false&references=false&vetted=0&cache_ttl="  |
| 1.  | 1064984 http://eol.org/api/pages/1.0/1064984.json?<br>images=2&videos=0&sounds=0&maps=0&text=2&iucn=false&subjects=overview&licenses=all&details=false&common_names=false&synonyms=false&references=false&vetted=0&cache_ttl=" |
| 2.  | 323922 http://eol.org/api/pages/1.0/323922.json?<br>images=2&videos=0&sounds=0&maps=0&text=2&iucn=false&subjects=overview&licenses=all&details=false&common_names=false&synonyms=false&references=false&vetted=0&cache_ttl="   |
| 3.  | 323922 http://eol.org/api/pages/1.0/323922.json?<br>images=2&videos=0&sounds=0&maps=0&text=2&iucn=false&subjects=overview&licenses=all&details=false&common_names=false&synonyms=false&references=false&vetted=0&cache_ttl="   |
| 4.  | 222022 http://eol.org/api/pages/1.0/222022.json?<br>images=2&videos=0&sounds=0&maps=0&text=2&iucn=false&subjects=overview&licenses=all&details=false&common_names=false&synonyms=false&references=false&vetted=0&cache_ttl="   |

Aceptar Cancelar

**Nota:** EOL ofrece una serie de proveedores que se pueden consultar a través del API ([http://eol.org/api/provider\\_hierarchies/1.0.json](http://eol.org/api/provider_hierarchies/1.0.json)) al momento de realización de esta guía se encuentran:

"AntWeb (Ant Species)"  
"WORMS Species Information (Marine Species)"  
"FishBase (Fish Species)"  
"IUCN Red List (Species Assessed for Global Conservation)"  
"Index Fungorum"  
"Cephaloleia LifeDesk resource"  
"The Reptile Database"  
"GBIF Nub Taxonomy"  
"Avibase - IOC World Bird Names (2011)"  
"Integrated Taxonomic Information System (ITIS)"  
"Paleobiology Database"  
"Clements Checklist resource"  
"Taxonomic Hierarchy of COL-China 2012"  
"South African Vertebrates"  
"Continenticola Classification"  
"NCBI Taxonomy"  
"Species 2000 & ITIS Catalogue of Life: April 2013"  
"Metalmark Moths of the World"  
"Algeabase resource"  
"Algae"

12. Para obtener una lista legible agregue una columna basada en `json_proveedores` (Editar columna > Agregar columna ) usando la expresión:

```
forEach(value.parseJson().get('taxonConcepts'),entry,entry.get('nameAccordingTo')).
uniques().join(", ")
```

Nómbrela `lista_proveedores`.

13. El siguiente paso es usar el ID del proveedor de interés. En este caso se utilizará *Catalogue of Life*. Para obtener este ID agregue una columna llamada `COL_ID` basada en `json_proveedores` utilizando la expresión:

```
forEach(value.parseJson().get('taxonConcepts'),entry,if(entry.get('nameAccordingTo'
)=="Species 2000 & ITIS Catalogue of Life: April
2013",entry.get("identifier"),"").uniques().join(",")
```

**Nota:** Puede usar cualquier proveedor de la lista del recuadro anterior (siempre y cuando aparezca en `lista_proveedores`), para esto reemplace en las fórmulas del punto 13 al XX de este ejercicio el valor "`Species 2000 & ITIS Catalogue of Life: April 2013`" por el nombre del proveedor que desee.

14. Realice una **Faceta por blanco** en la columna `COL_ID`, seleccione **false** para trabajar únicamente con las filas que tienen ID.



15. Cree una columna basada en `COL_ID` utilizando **Editar columna > Agregar columna accediendo a URLs...**

```
"http://eol.org/api/hierarchy_entries/1.0/"+value+".json?common_names=false&synonyms=false&cache_ttl="
```

Recuerde nombrar la columna como `COL_tax` y en el campo **Tiempo de retraso** escriba 250, haga clic en **Aceptar** y tenga paciencia ya que la consulta puede tomar un tiempo.

**Nota:** si su conexión es lenta:

1. Haga clic en **Cancelar**.
2. Realice una **Faceta de Texto** en **Clase**.
3. Seleccione allí los 369 registros de la Clase *Mammalia*.



4. Repita el punto 15.

16. Haga esta información legible agregando una o varias columnas basadas en `COL_tax`. Tiene dos opciones:

- a. Utilizar la expresión:

```
forEach(value.parseJson().get('ancestors'),entry,if(entry.get('taxonRank')=="family",entry.get("scientificName"),"")).uniques().join("")
```

Y reemplazar **family** por la categoría taxonómica que desee, deberá hacerlo para cada una de las categorías que desee: **kingdom**, **phylum**, **class**, **order**, **family**.

- b. Concatenar las expresiones de reino, familia, filo, clase, orden y familia con un signo + después de cada expresión.

```
forEach(value.parseJson().get('ancestors'),entry,if(entry.get('taxonRank')=="kingdom",entry.get("scientificName"),"")).uniques().join("")+forEach(value.parseJson().get('ancestors'),entry,if(entry.get('taxonRank')=="phylum",entry.get("scientificName"),"")).uniques().join("")
```

Y así sucesivamente, esto nos arrojará una columna con todos los valores.

En este ejemplo se utilizará la opción b. Agregue una columna basada en **COL\_tax** utilizando la expresión:

```
forEach(value.parseJson().get('ancestors'),entry,if(entry.get('taxonRank')=="kingdom",entry.get("scientificName"),"")).uniques().join("")+forEach(value.parseJson().get('ancestors'),entry,if(entry.get('taxonRank')=="phylum",entry.get("scientificName"),"")).uniques().join("")+forEach(value.parseJson().get('ancestors'),entry,if(entry.get('taxonRank')=="class",entry.get("scientificName"),"")).uniques().join("")+forEach(value.parseJson().get('ancestors'),entry,if(entry.get('taxonRank')=="order",entry.get("scientificName"),"")).uniques().join("")+forEach(value.parseJson().get('ancestors'),entry,if(entry.get('taxonRank')=="family",entry.get("scientificName"),"")).uniques().join("")
```

Nómbrela **COL**.

18. Para separar las categorías en varias columnas, en **COL** siga la ruta **Editar columnas > Dividir en varias columnas...**,

**Dividir columna COL en varias columnas**

**Dividir por**

Separador  
 separador   regex

Dividir en  columnas máximo (deje en blanco para no limitar)

Longitud del campo

Separe listas de números con comas, p. ej., 5, 7, 15

**Luego de la división**

Interpretar el tipo de celda

Eliminar esta columna

Aceptar Cancelar

Haga clic en **Aceptar** y verá los valores separados. Si desea cambiar los nombres de las columnas, ingrese al Menú columna y siga la ruta **Editar columnas > Renombrar esta columna** y asigne el nombre que corresponda.

## ENLACES ADICIONALES

Name validation Tutorial:

[https://docs.google.com/document/d/1tkDRXIYhmassYAk5T4v5oac5prF0jAiSMr\\_JEGTvhRo/edit](https://docs.google.com/document/d/1tkDRXIYhmassYAk5T4v5oac5prF0jAiSMr_JEGTvhRo/edit)

Higher Taxonomy Tutorial:

[https://docs.google.com/document/d/1XZ\\_pM9gldQzHzl8wfUCVea-52yub5T\\_3tc-snBgPRa0/edit](https://docs.google.com/document/d/1XZ_pM9gldQzHzl8wfUCVea-52yub5T_3tc-snBgPRa0/edit)

Más servicios de reconciliación:

<http://iphylo.blogspot.com/2012/02/using-google-refine-and-taxonomic.html>

Documentación para usuarios del programa:

<https://github.com/OpenRefine/OpenRefine/wiki/Documentation-For-Users>

Api de EOL:

<http://eol.org/api>

Lista de recursos disponibles para OpenRefine:

<https://github.com/OpenRefine/OpenRefine/wiki/External-Resources>