Extractor de datos de Scopus

Autor: Pablo Mazzucchi

Contenidos

[Archivos de configuración 3](#_Toc430868991)

[Procesos para extracción de datos 4](#_Toc430868992)

[Proceso Semestral 5](#_Toc430868993)

[Query 1 y Query 2 6](#_Toc430868994)

[Datos extraídos 6](#_Toc430868995)

# 

# Archivos de configuración

Básicamente son dos, el que se muestra en la Figura 1, que contiene las variables y parámetros que serán utilizados durante la ejecución de la aplicación, y el mostrado en la Figura 2, que permite visualizar la cantidad de llamadas a las APIs que se hacen desde la aplicación, con una API Key determinada.

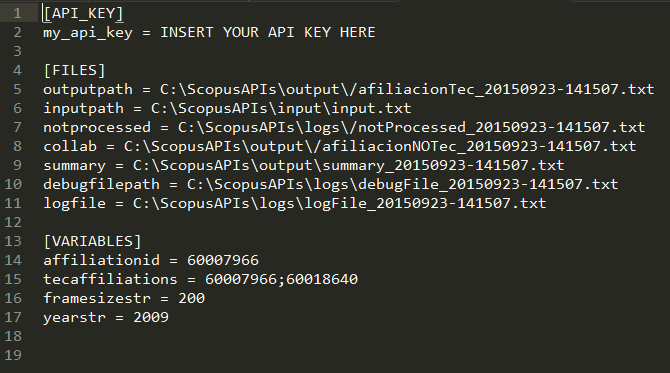


Figura 1- Archivo configuración: C:\ScopusAPIs-master\config\config.cfg

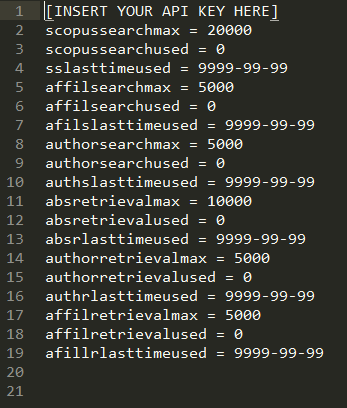


Figura 2 - Archivo configuración: C:\ScopusAPIs-master\config\limits.cfg

En la Tabla 1 se describe con detalle las variables definidas en el archivo de configuración mostrado en la Figura 1.

|  |  |
| --- | --- |
| Variable | Descripción |
| outputpath | Ruta del archivo que contendrá a todos los autores (y la respectiva información de sus publicaciones) cuya afiliación sea alguna de las especificadas en **tecaffiliations**. Para más información: Datos extraídos |
| inputpath | Ruta del archivo que contendrá el ScopusID de todos los autores de los que se requiera información sobre sus publicaciones |
| Notprocessed | Ruta del archivo que contendrá los IDs de todos aquellos autores a los cuales no se les haya podido extraer la información solicitada |
| Collab | Ruta del archivo que contendrá a todos los autores (y la respectiva información de sus publicaciones) cuya afiliación sea **DIFERENTE** a las especificadas en **tecaffiliations**. Para más información: Datos extraídos |
| Summary | Ruta del archivo que contendrá información relevante a las afiliaciones procesadas. Para más información: Datos extraídos |
| Debugfilepath | Ruta del archivo que contendrá información útil para encontrar “bugs” dentro de la aplicación. |
| Logfile | Ruta del archivo que contendrá mensajes de error ocurridos al procesar tanto autores como afiliaciones. |
| Affiliationid | ID de la afiliación de la cual se requieran datos. |
| Tecaffiliations | Los IDs que estén en esta lista (separados por “;”) definirán qué autores se guardarán en el archivo **outputpath** |
| Framesizestr | Cantidad de registros que se obtienen en cada llamada a las APIs de Scopus. El máximo definido por Elsevier es 200. |
| yearstr | Año a partir del cual se obtiene información sobre las afiliaciones, autores o publicaciones. El filtro que se hace es “>yearstr”. Osea, si se requieren publicaciones del 2010 hasta la fecha en que se ejecute la aplicación, entonces se deberá poner en esta variable el valor “2009”, ya que “> 2009” es equivalente a “>=2010”. |

Tabla 1 - Descripción de las variables definidas en el archivo de configuración

En la Tabla 2 se describe con detalle las variables definidas en el archivo de configuración mostrado en la Figura 2.

|  |  |
| --- | --- |
| Variable | Descripción |
| Scopussearchmax,  Scopussearchused,  sslasttimeused | Máxima cantidad de llamadas por semana, llamadas utilizadas hasta el momento y fecha de la última utilización de la API [**Scopus Search**](http://api.elsevier.com/documentation/SCOPUSSearchAPI.wadl)**.** |
| Affilsearchmax,  Affilsearchused,  Afilslasttimeused | Máxima cantidad de llamadas por semana, llamadas utilizadas hasta el momento y fecha de la última utilización de la API [**Affiliation Search**](http://api.elsevier.com/documentation/AFFILIATIONSearchAPI.wadl#N10027)**.** |
| Authorsearchmax,  Authorsearchused,  Authslasttimeused | Máxima cantidad de llamadas por semana, llamadas utilizadas hasta el momento y fecha de la última utilización de la API [**Author Search**](http://api.elsevier.com/content/search/author). |
| Absretrievalmax,  Absretrievalused,  Absrlasttimeused | Máxima cantidad de llamadas por semana, llamadas utilizadas hasta el momento y fecha de la última utilización de la API [**Abstract Retrieval**](http://api.elsevier.com/documentation/AbstractRetrievalAPI.wadl)**.** |
| Authorretrievalmax,  Authorretrievalused,  Authrlasttimeused | Máxima cantidad de llamadas por semana, llamadas utilizadas hasta el momento y fecha de la última utilización de la API [**Author Retrieval**](http://api.elsevier.com/documentation/AuthorRetrievalAPI.wadl#N101F4). |
| Affilretrievalmax,  Affilretrievalused,  afillrlasttimeused | Máxima cantidad de llamadas por semana, llamadas utilizadas hasta el momento y fecha de la última utilización de la API  **TODO: falta implementar** |

Tabla 2 - Descripción de las variables definidas en el archivo de límites

# Procesos para extracción de datos

Actualmente existen 3 procesos definidos para extraer información, como se muestra en la Figura 3, o con más detalle en la Tabla 3.

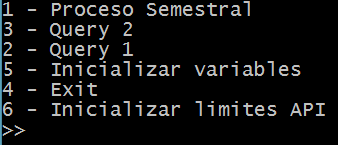


Figura 3 - Interfaz aplicación para extraer datos de las BBDD de Scopus

|  |  |
| --- | --- |
| Proceso Semestral | Se extrae información sobre las publicaciones de los investigadores del Tec, y a su vez se contabiliza las afiliaciones de todo el lote de publicaciones extraídas. |
| Query 1 | Idem proceso semestral, excepto que el input viene dado por la siguiente query:   ( ( AFFIL ( tecnológico  de  monterrey )  OR  AFFIL ( itesm ) )  OR  AFFIL ( hosp\*  san  josé  tec\*  monterrey ) ) and not ( ( AF-ID ( "Hospital San Jose Tec de Monterrey"  60018640 ) ) )  OR  ( AF-ID ( "Tecnologico de Monterrey"  60007966 ) ) |
| Query 2 | Idem proceso semestral, excepto que el input viene dado por la siguiente query:  ( AFFIL ( tec  de  monterrey ) )  AND NOT  ( ( AFFIL ( tecnológico  de  monterrey )  OR  AFFIL ( itesm ) )  OR  AFFIL ( hosp\*  san  josé  tec\*  monterrey ) ) |

Tabla 3 - Descripción de los procesos de extracción de datos de la BBDD de Scopus

## Proceso Semestral

Para ejecutarlo es necesario seguir los siguientes pasos:

1. Abrir el archivo de configuración **config.cfg** (ver Archivos de configuración).
2. En la sección **[API\_KEY]**, reemplazar el texto “INSERT YOUR API KEY HERE”, por la API Key que se haya obtenido en el [portal de Elsevier](http://dev.elsevier.com/myapikey.html).
3. En la sección **[VARIABLES]**, modificar las variables **tecaffiliations** y **yearstr** según corresponda (para más información sobre el significado de las variables ver Archivos de configuración)
4. Guardar los cambios.
5. Ejecutar la aplicación ScopusAPIs (situada en la ruta C:\ScopusAPIs-master\bin)
6. Seleccionar la opción **Inicializar variables** presionando el número 5 seguido de la tecla enter (la aplicación se cerrará)
7. Abrir el archivo **limits.cfg** (ver Archivos de configuración). Si todo funcionó correctamente, en la línea 1 del archivo podrá ver la API Key ingresada en el punto 2.
8. Asegúrese que existe el archivo “input” en la ruta “C:\ScopusAPIs\input” y que tenga una estructura similar a la mostrada en la Figura 4: **Solo 1 Scopus ID (de los autores) por cada línea**.
9. Ejecutar la aplicación ScopusAPIs nuevamente.
   1. Se recomienda inicializar los límites de las APIs (opción 6) sólo si es la primera vez que utilizará la aplicación para extraer datos o si pasó más de una semana de la última utilización (porque pasada una semana, se puede hacer uso nuevamente de la máxima cantidad de llamadas que permite Elsevier para cada API)
   2. Seleccionar las opción 1.

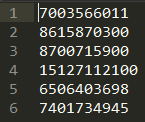


Figura 4 - Estructura del archivo de entrada para el proceso Semestral

## Query 1 y Query 2

Para ejecutarlo es necesario seguir los siguientes pasos:

1. Abrir el archivo de configuración **config.cfg** (ver Archivos de configuración).
2. En la sección **[API\_KEY]**, reemplazar el texto “INSERT YOUR API KEY HERE”, por la API Key que se haya obtenido en el [portal de Elsevier](http://dev.elsevier.com/myapikey.html).
3. En la sección **[VARIABLES]**, modificar la variable **yearstr** según corresponda (para más información sobre el significado de las variables ver Archivos de configuración)
4. Guardar los cambios.
5. Ejecutar la aplicación ScopusAPIs (situada en la ruta C:\ScopusAPIs-master\bin).
6. Seleccionar la opción **Inicializar variables** presionando el número 5 seguido de la tecla enter (la aplicación se cerrará).
7. Abrir el archivo **limits.cfg** (ver Archivos de configuración). Si todo funcionó correctamente, en la línea 1 del archivo podrá ver la API Key ingresada en el punto 2.
8. Ejecutar la aplicación ScopusAPIs nuevamente.
   1. Se recomienda inicializar los límites de las APIs (opción 6) sólo si es la primera vez que utilizará la aplicación para extraer datos o si pasó más de una semana de la última utilización (porque pasada una semana, se puede hacer uso nuevamente de la máxima cantidad de llamadas que permite Elsevier para cada API)
   2. Seleccionar las opción 2 para Query 1 o la opción 3 para Query 2.

# Datos extraídos

En la Figura 5 se puede visualizar la estructura de los archivos de salida. La misma será utilizada para los archivos que contengan a los investigadores del Tec y para los que contengan a los que no tienen afiliación del Tec. En la Tabla 4 se describen los datos recuperados.

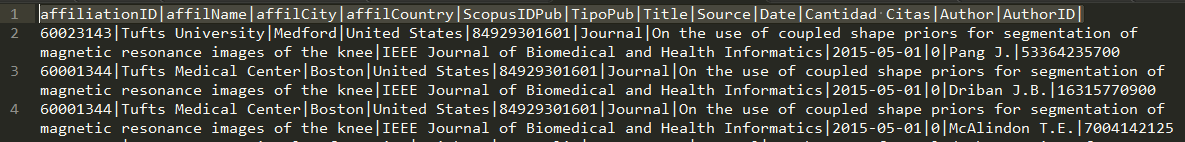


Figura 5 - Estructura de los archivos de salida del proceso semestral, query 1 y 2

|  |  |
| --- | --- |
| Atributo | Descripción |
| affiliationID | ID de la afiliación que tenía el investigador al momento de escribir la publicación |
| affilName | Nombre de la afiliación |
| affilCountry | País de la afiliación |
| ScopusIDPub | Scopus ID de la publicación |
| TipoPub | Tipo de publicación: revista, libro, congreso, etc. |
| Title | Título de la publicación |
| Source | En donde fue publicada |
| Date | Fecha en la que se publica |
| Cantidad Citas | Cantidad de citas que tiene la publicación |
| Author | Nombre del autor |
| Author ID | Scopus ID del autor |

Tabla 4 - Descripción de los atributos extraídos en el proceso Semestral, en la query 1 y en la query 2

En la Figura 6 se puede ver la estructura del archivo que resume la producción científica de todas las afiliaciones que colaboraron en publicaciones donde también participaron investigadores del Tec. En la Tabla 5 se describe en detalle cada uno de los atributos

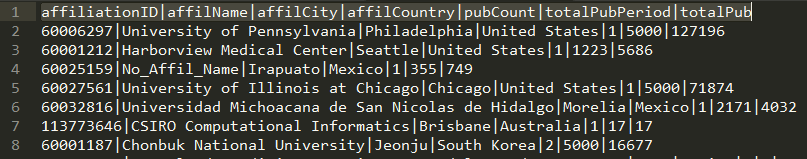


Figura 6 - Estructura del archivo que visualiza a nivel de afiliación la producción científica y la colaboración con el Tec

|  |  |
| --- | --- |
| Atributo | Descripción |
| affiliationID | ID de la afiliación que tenía el investigador al momento de escribir la publicación. |
| affilName | Nombre de la afiliación. |
| affilcity | Ciudad de la afiliación. |
| affilCountry | País de la afiliación. |
| pubcount | Cantidad de publicaciones en donde haya aparecido la afiliación en cuestión, en el lote de los datos procesados. |
| totalpubperiod | Cantidad de publicaciones que se escribieron bajo la afiliación en cuestión, en el mismo periodo de fechas definido para el proceso semestral, query 1 o 2. |
| totalpub | Cantidad de publicaciones que hay en Scopus para la afiliación en cuestión. |

Tabla 5 - Descripción de los atributos del archivo que sumariza el proceso Semestral, query 1 o 2