

Generación automática de resúmenes extractivos genéricos de múltiples documentos basado en Word2Vec



Álvaro José Echeverría Restrepo

Manual de instalación

Director: PhD. Carlos Alberto Cobos Lozada

Co-Director: PhD. Martha Eliana Mendoza Becerra

**Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Sistemas
Grupo de I+D en Tecnologías de la Información (GTI)
Área de Interés: Gestión de la Información y Sistemas Inteligentes
Popayán, Abril de 2018**

Tabla de contenido

INTRODUCCIÓN.....	3
REQUERIMIENTOS DEL SISTEMA.....	3
ESTRUCTURA DE LOS DIRECTORIOS DE TRABAJO	3
INSTALACIÓN DEL ENTORNO DE DESARROLLO	4
Instalación de Python 2.7	4
Instalación de Perl	4
Instalación de NLTK WordNet para lematización y búsqueda de sinónimos.....	8

INTRODUCCIÓN

El objetivo de este manual es proporcionar una guía para la instalación del entorno de desarrollo y complementos necesarios para la ejecución de los algoritmos involucrados en el trabajo de investigación: Generación automática de resúmenes extractivos genéricos de múltiples documentos basado en Word2Vec. De igual forma se describe la estructura de los directorios de trabajo necesarios para la ejecución de los mismos. Previamente a las instalaciones descritas en este manual, se debe haber instalado Visual Studio 2010.

REQUERIMIENTOS DEL SISTEMA

Los algoritmos desarrollados en el presente trabajo de investigación se podrán ejecutar en máquinas con las siguientes características:

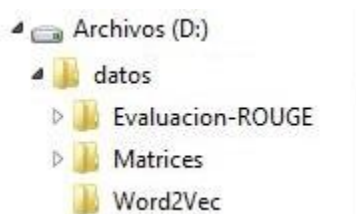
Sistema Operativo: Windows.

Procesador: mínimo recomendado 1.8 GHz.

RAM: mínimo recomendado 4 GB.

ESTRUCTURA DE LOS DIRECTORIOS DE TRABAJO

La carpeta **datos** debe estar ubicada en el disco **D:** para que los algoritmos puedan hacer uso de los datos allí presentes.

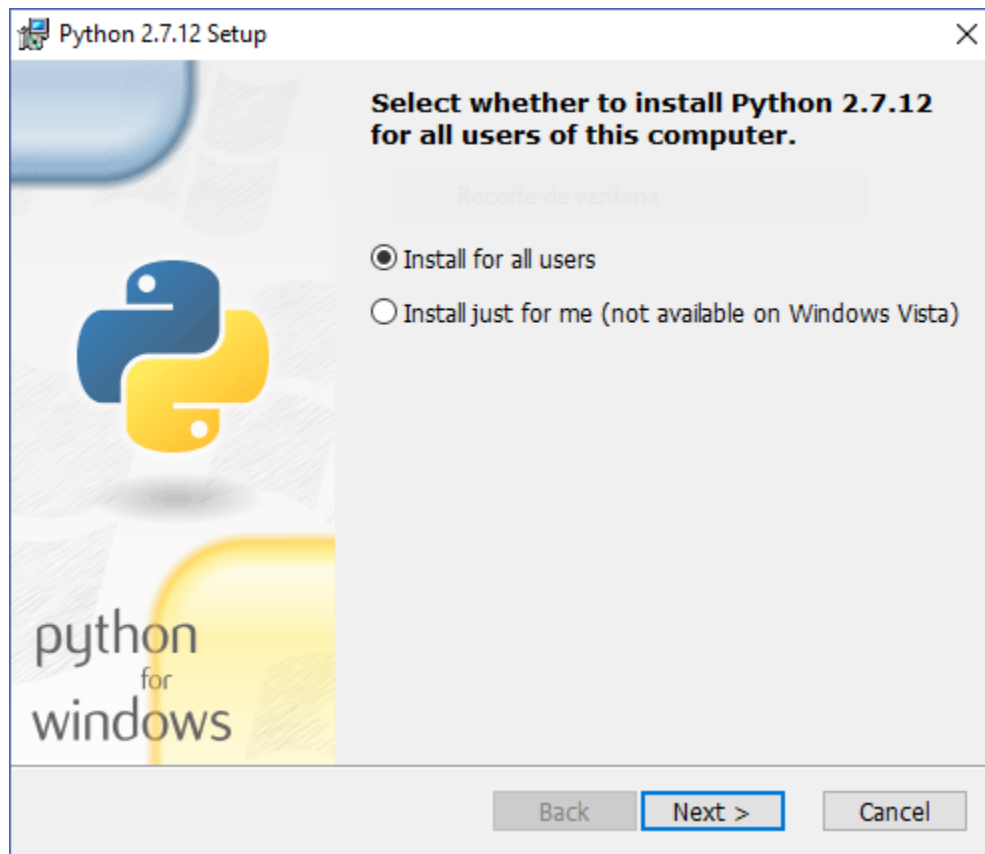


En la carpeta **datos** se encuentra el código en Python para realizar la lematización y búsqueda de sinónimos, además los datasets (Documentos de DUC) y matrices empleadas en la ejecución de los algoritmos.

INSTALACIÓN DEL ENTORNO DE DESARROLLO

Instalación de Python 2.7

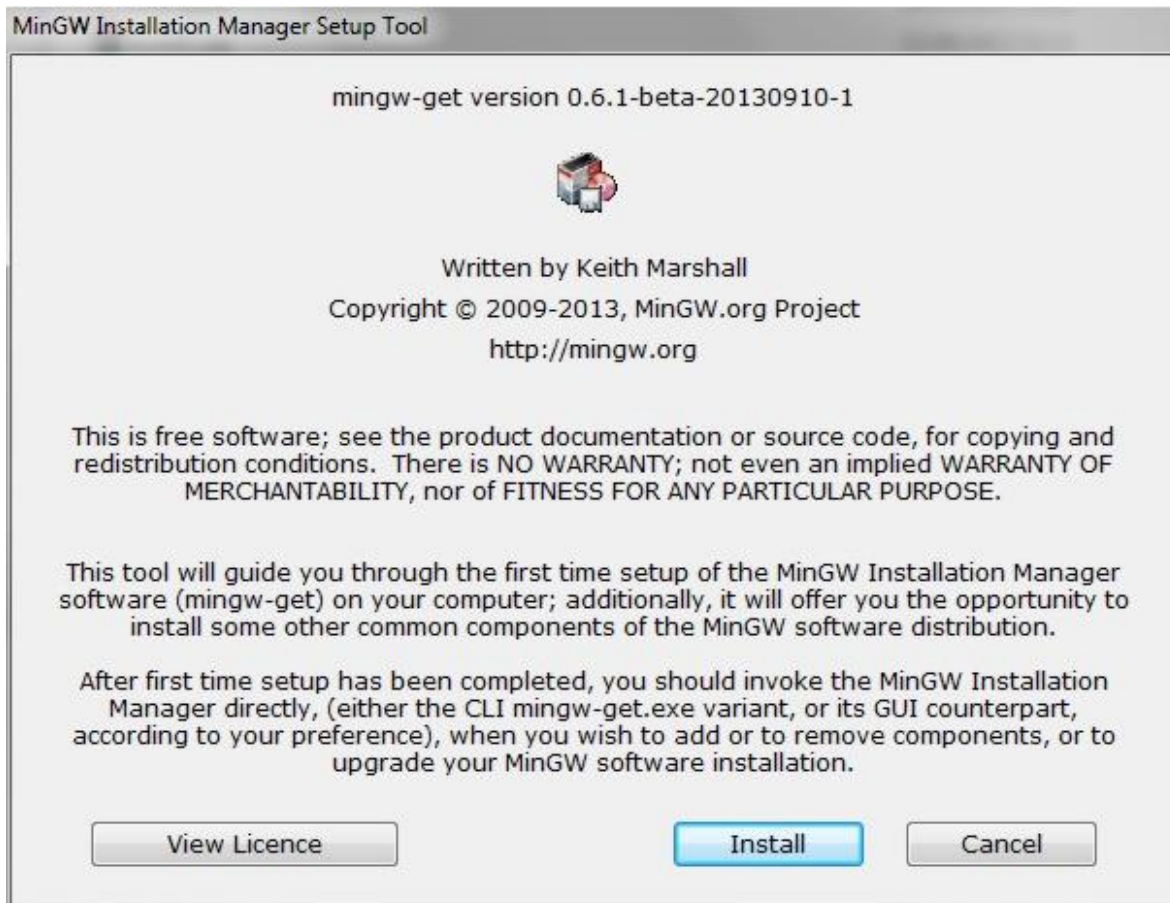
Primero se debe descargar Python 2.7.12.msi desde la página oficial de Python (<https://www.python.org/>) que por defecto están los enlaces para Windows. Luego se ejecuta el instalador mostrando la siguiente ventana.



Se realiza la instalación ejecutando la opción “Next >” hasta finalizar el proceso.

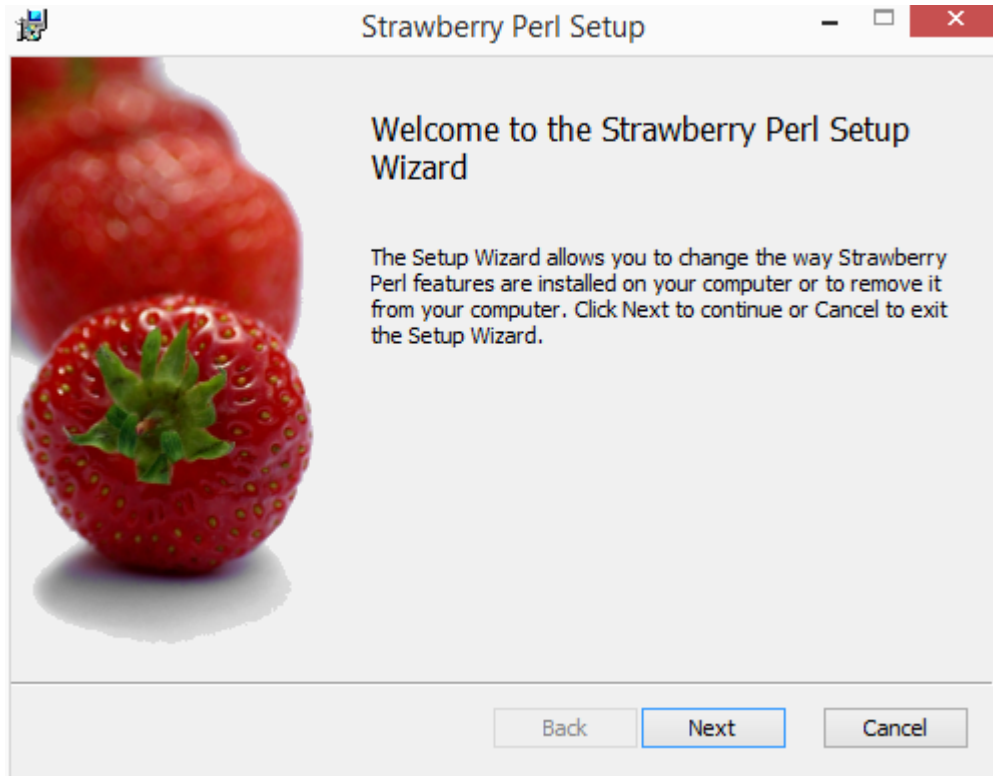
Instalación de Perl

Si no se tiene instalado un compilador de c++, primero se debe instalar MinGW descargándolo desde la página oficial (<http://www.mingw.org/>), Una vez descargado, ejecuta el archivo **mingw-get-setup.exe** e inicia la instalación del software:



Una vez instalado MinGW, se confirma que en las variables de entorno del sistema se encuentre el path c:MinGw y se procede a la instalación de Perl.

Se debe descargar el archivo **strawberry-perl-5.22.2.1-32bit.msi** desde la página oficial (<http://strawberryperl.com/>) y ejecutarlo para iniciar la ejecución del software.



Luego de instalar Strawberry Perl, se procede a entrar al cmd de Windows y ejecutar el comando **CPAN** para entrar a la instalación de módulos cpan.

```
Microsoft Windows [Versión 6.3.9600]
(c) 2013 Microsoft Corporation. Todos los derechos reservados.

C:\Users\lenovo-pc>CPAN
Loading internal null logger. Install Log::Log4perl for logging messages

cpan shell -- CPAN exploration and modules installation (v2.11)
Enter 'h' for help.

cpan>
```

Luego, en el cpan ejecutar el siguiente comando: **install XML::DOM** y esperar a que termine toda la instalación.

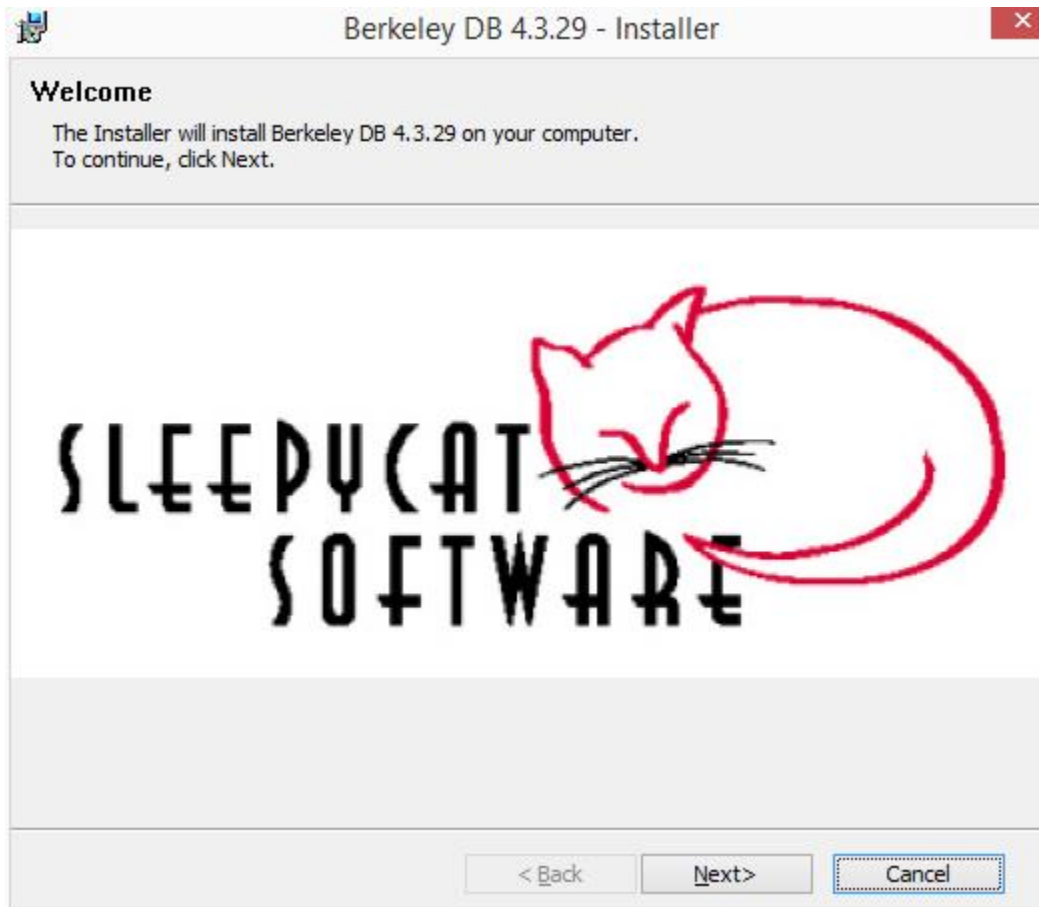
```
Símbolo del sistema - CPAN
Microsoft Windows [Versión 6.3.9600]
(c) 2013 Microsoft Corporation. Todos los derechos reservados.

C:\Users\lenovo-pc>CPAN
Loading internal null logger. Install Log::Log4perl for logging messages

cpan shell -- CPAN exploration and modules installation (v2.11)
Enter 'h' for help.

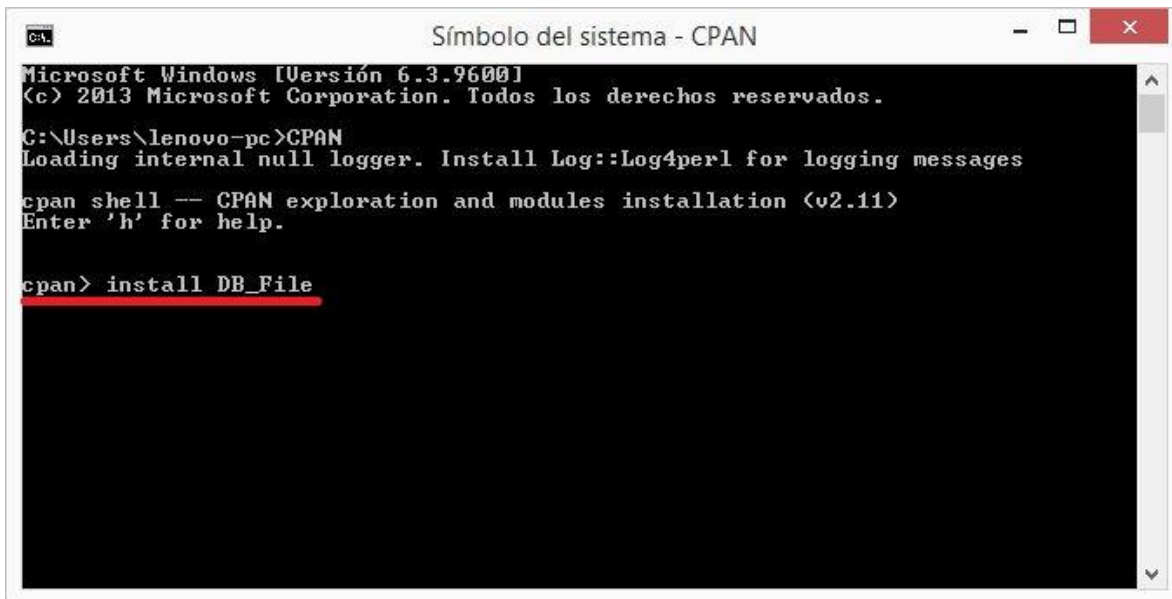
cpan> install XML::DOM
```

A continuación se procede a instalar Berkeley db desde el archivo **db-4.3.29.msi** disponible en la página oficial de Oracle (<http://www.oracle.com/>).



Se completa la instalación ejecutando la opción "Next >" hasta finalizar el proceso.

Por último se abre el cmd de Windows y se entra al instalador de módulos cpan ejecutando el comando **CPAN** y a continuación, dentro de cpan se ejecuta el comando: **install DB_File**.



```
Símbolo del sistema - CPAN
Microsoft Windows [Versión 6.3.9600]
(c) 2013 Microsoft Corporation. Todos los derechos reservados.

G:\Users\lenovo-pc>CPAN
Loading internal null logger. Install Log::Log4perl for logging messages

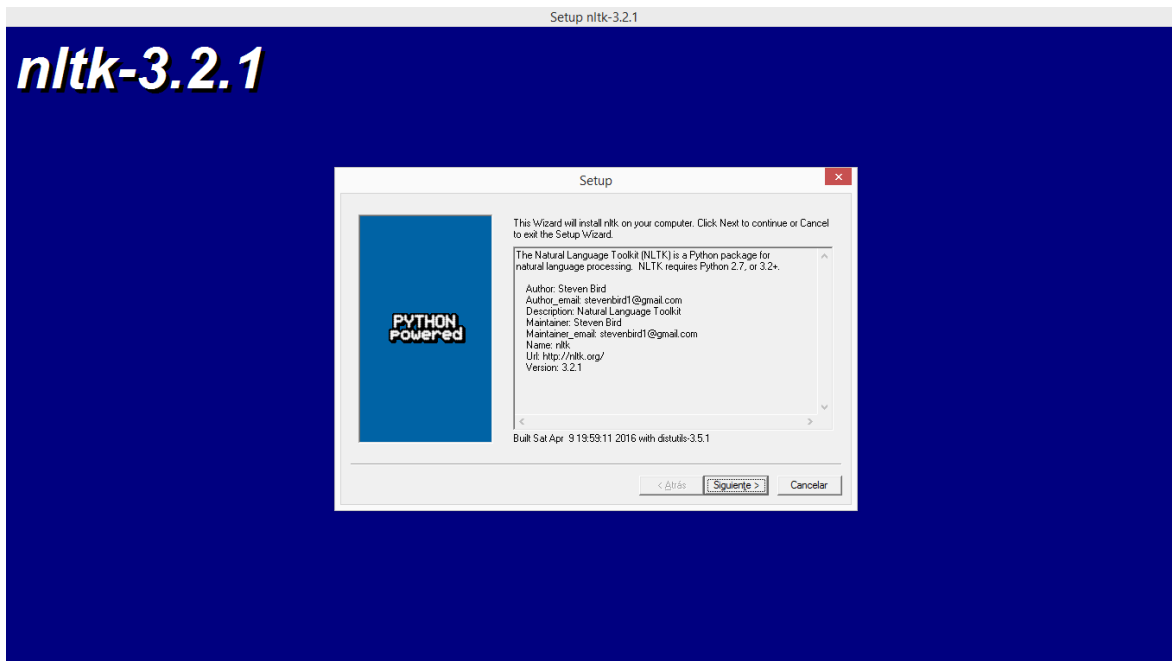
cpan shell -- CPAN exploration and modules installation (v2.11)
Enter 'h' for help.

cpan> install DB_File
```

Y esperar a que termine el proceso de instalación.

Instalación de NLTK WordNet para lematización y búsqueda de sinónimos

Primero se debe descargar el archivo de instalación **nlk-3.2.1.win32.exe** (<http://www.nltk.org/install.html> o <https://pypi.python.org/pypi/nltk>) y ejecutarlo para iniciar la instalación.



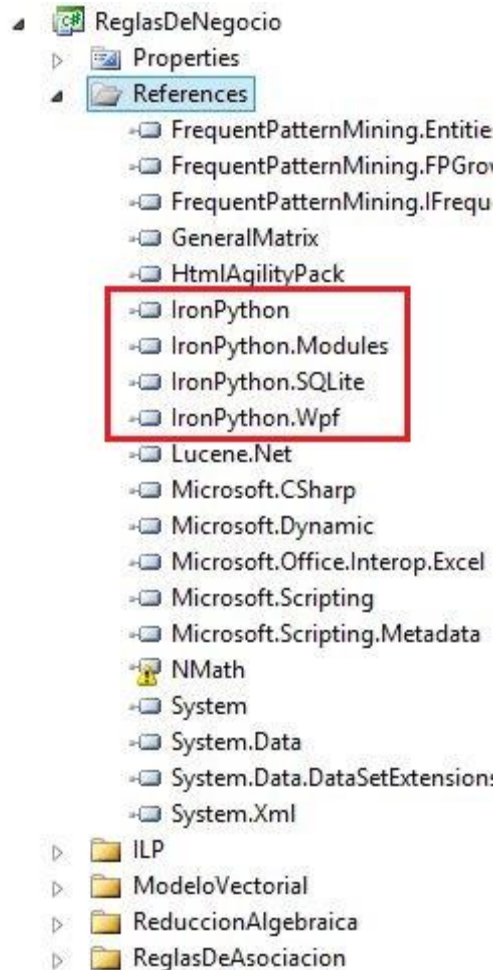
Se completa la instalación ejecutando la opción “Siguiente >” hasta finalizar el proceso.

Ahora, teniendo instalado el Python 2.7, se procede a instalar IronPython mediante el archivo **IronPython-2.7.6.3.msi** disponible en la página oficial (<http://ironpython.net/download/>).



Se completa la instalación ejecutando la opción “Next >” hasta finalizar el proceso.

Luego se deben agregar los archivos DLL de la carpeta donde se instaló IronPython a las referencias del proyecto de Visual Studio 2010.



Luego, si es necesario, agregar el archivo **fcntl.py** en la carpeta Lib de Python27 disponible en <https://pypi.python.org/pypi/micropython-fcntl/0.0.4>

Por último, descargar la carpeta **nltk_data**. Para descargar esta carpeta se hace necesario abrir el archivo **python.exe** y ejecutar el comando **nltk.download()** y copiarla en la carpeta raíz del usuario (por ejemplo c:/users/alvaro).