

XML

TP 1: Documentos XML

Documento bien formado – Instalación del parser

Instalación del parser para validar

Existen dos tipos de parsers para realizar validaciones de documentos:

- **Document Object Model (DOM)**: construye todo el árbol del documento (crea un DOM document) completamente en memoria.
- **Simple API for Xml (SAX)**: sólo crea el árbol de a pedazos por demanda (crea un SAX stream).

Si el documento XML es muy grande y quizás no se pueda mantenerlo íntegramente en memoria, conviene usar SAX.

Más allá de cuál sea el parser a utilizar, precisamos de uno para poder validar nuestros documentos. Hay muchas implementaciones de estos parsers. Nosotros vamos a usar la proporcionada por Apache: **Xerces**.

Tener en cuenta que los navegadores sólo se fijan si el documento está bien formado y no realizan validaciones contra esquemas (las ignora).

La implementación Xerces está diseñada para ejecutar por línea de comandos bajo cualquier entorno (Windows, Linux) porque está programada en Java.

[Instrucciones para Instalar en Linux \(Ubuntu\)](#)

[Instrucciones para Instalar en Mac](#)

[Instrucciones para Instalar en Windows](#)

[Instrucciones para Instalar en Windows WSL](#)

[Instalación de plugins en Visual Studio Code](#)

Pasos para su instalación en Linux (Ubuntu)

1) Crear un directorio “XML” en “Carpeta Personal”.

Nota: NO crear la carpeta XML dentro de directorios que se encuentren dentro de iCloud/OneDrive/Google Drive (o cualquier otro sistema de archivos en la nube). En su reemplazo, utilizar el directorio home local del usuario.

2) Bajar a dicha carpeta, desde el Campus ITBA (campus.itba.edu.ar), los archivos: **Xerces-J-bin.2.12.1.zip**, **ejemplo_bashrc** y **pruebaTP1.xml**. El archivo **Xerces-J-bin.2.12.1.zip** también puede bajarse de <https://dlcdn.apache.org/xerces/j/binaries/>

3) Descompactar el archivo **Xerces-J-bin.2.12.1.zip** en la carpeta “XML”. Las bibliotecas java que precisamos son: **xercesImpl.jar**, **xml-apis.jar** y **xercesSamples.jar**

Debemos tener instalado java y seteadas las variables de entorno PATH y CLASSPATH. En el paso siguiente se explica cómo hacerlo en Linux.

4) Configuración de las variables de entorno:

Mostrar los archivos ocultos de “Carpeta Personal” mediante el explorador de Ubuntu: Ver->Mostrar Archivos ocultos.

Asumimos que hemos creado el directorio XML en nuestra carpeta. Si no es así reemplazar lo indicado en el ítem 3-c) por el path dónde haya descomprimido los 3 archivos que mencionados en el punto 2): xercesSamples.jar, xml-apis.jar y xercesImpl.jar.

Editar el archivo **.bashrc** de nuestro home. Si no existe, crearlo.

EL USUARIO “SALERNO” DEBE SER REEMPLAZADO POR EL USUARIO DE CADA UNO

Incluir las siguientes 2 líneas (ver archivo de Campus **ejemplo_bashrc**)

```
export
CLASSPATH=$CLASSPATH:/home/salerno/XML/xerces-2_12_1/xercesSamples.jar:/home
/salerno/XML/xerces-2_12_1/xml-apis.jar:/home/salerno/XML/xerces-2_12_1/xer
cesImpl.jar
export PATH=$PATH:/usr/lib/jvm/java-7-openjdk-amd64/
```

Si las variables ya existen y se está editando el archivo (no creándolo), agregar lo indicado usando : (dos puntos).

Nota: NO es conveniente llevar al **.bashrc** las líneas del recuadro del punto c) de este documento pdf, mediante **copy&paste**. Se pueden copiar estas líneas desde el archivo **ejemplo_bashrc.txt** bajado de Campus en el punto 2)

Grabar el archivo **.bashrc** y reiniciar la sesión para abrir y cerrar la terminal para que los cambios tengan efecto.

Una vez ejecutado el script, la validación (evaluar si el XML está bien formado), procede de alguna de estas 2 formas según se quiera usar parser DOM o parser SAX:

```
$ java dom.Writer nombreDocumento.xml

O bien

$ java sax.Writer nombreDocumento.xml
```

Pasos para su instalación en Mac

1) Crear un directorio “XML” en el directorio home del usuario.

Nota: NO crear la carpeta XML dentro de directorios que se encuentren dentro de iCloud/OneDrive/Google Drive (o cualquier otro sistema de archivos en la nube). En su reemplazo, utilizar el directorio home local del usuario (el que suele estar dentro del directorio /Users/).

2) Bajar a dicha carpeta, desde el Campus ITBA (campus.itba.edu.ar), los archivos: **Xerces-J-bin.2.12.1.zip**, **ejemplo_bashrc** y **pruebaTP1.xml**. El archivo **Xerces-J-bin.2.12.1.zip** también puede bajarse de <https://dlcdn.apache.org/xerces/j/binaries/>

3) Descompactar el archivo **Xerces-J-bin.2.12.1.zip** en la carpeta “XML”. Las bibliotecas java que precisamos son: **xercesImpl.jar**, **xml-apis.jar** y **xercesSamples.jar**

Debemos tener instalado java y seteadas las variables de entorno PATH y CLASSPATH. En el paso siguiente se explica cómo hacerlo en Mac.

4) Configuración de las variables de entorno:

- Abrir una nueva terminal.
- Asumimos que hemos creado el directorio XML en nuestra carpeta. Si no es así reemplazar lo indicado en el ítem 3-c) por el path dónde haya descomprimido los 3 archivos que mencionados en el punto 2): **xercesSamples.jar**, **xml-apis.jar** y **xercesImpl.jar**.
- Ejecutar el siguiente comando en la terminal:
`echo $SHELL`
Si la salida de ese comando es `/bin/bash`, el archivo que hay que modificar en el próximo paso (d.) es `$HOME/.bash_profile`.
Si la salida de ese comando es `/bin/zsh`, el archivo que hay que modificar en el próximo paso (d.) es `$HOME/.zshenv`.
Si la salida de ese comando no es ninguna de las anteriores, se recomienda cambiar el “Interpretador de comandos” (o “Shell interpreter”) a Bash, ejecutando:
`chsh -s /bin/bash` (luego abrir una nueva terminal para que el comando surta efecto). Como consecuencia, el archivo a modificar en el próximo paso (d.) es `$HOME/.bash_profile`.
- Listar en la terminal si existe el archivo mencionado en el punto c. con el comando:
`ls <archivo>` (por ejemplo: `ls $HOME/.bash_profile*`, o `ls $HOME/.zshenv*`; según el caso).
Luego, abrir ese archivo con un editor de texto (como vim, vi, nano, etc). Si no existe, se crea automáticamente.

EL USUARIO “SALERNO**” DEBE SER REEMPLAZADO POR EL USUARIO DE CADA UNO**

Incluir las siguientes 2 líneas

```
export  
CLASSPATH=/Users/salerno/XML/xerces-2_12_1/xercesSamples.jar:/Users/salerno  
/XML/xerces-2_12_1/xml-apis.jar:/Users/salerno/XML/xerces-2_12_1/xercesImpl.  
jar  
export PATH=$PATH:$(/usr/libexec/java_home)
```

- Si las variables ya existen y se está editando el archivo (no creándolo), agregar lo indicado usando : (dos puntos).

Nota: NO es conveniente llevar al **.bash_profile** las líneas del recuadro del punto c) de este documento pdf, mediante **copy&paste**.

- Grabar el archivo **.bash_profile** (o **.zshenv**) y reiniciar la sesión para abrir y cerrar la terminal para que los cambios tengan efecto.
- Una vez ejecutado el script, la validación (evaluar si el XML está bien formado), procede de alguna de estas 2 formas según se quiera usar parser DOM o parser SAX:

```
$ java dom.Writer nombreDocumento.xml  
  
O bien
```

```
$ java sax.Writer nombreDocumento.xml
```

Pasos para su instalación en Windows

1) Ubicarse en la carpeta “Documents” (o “Documentos”), usando el “Explorador de archivos”.

Nota: NO utilizar la carpeta “Documents” que se encuentre dentro de iCloud/OneDrive/Google Drive (o cualquier otro sistema de archivos en la nube). En su reemplazo, la carpeta “Documents” local (aquella que se encuentra en el directorio C:\).

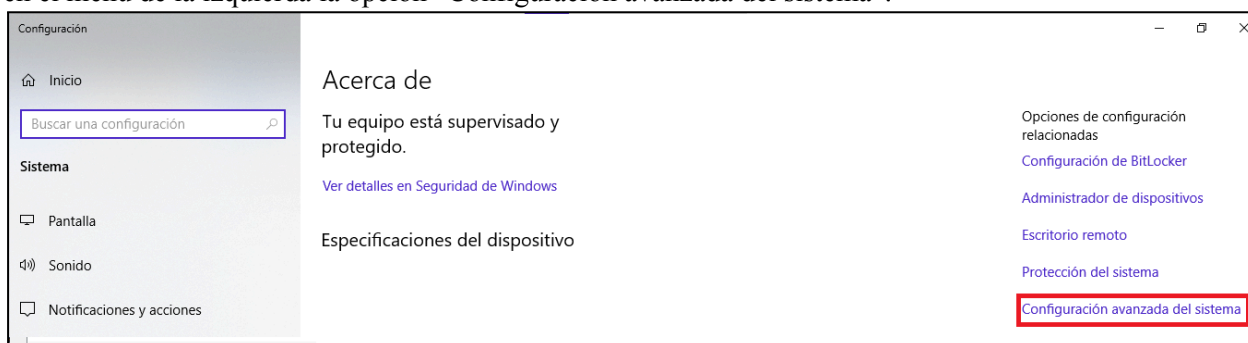
2) Bajar a dicha carpeta, desde el Campus ITBA (campus.itba.edu.ar), los archivos: **Xerces-J-bin.2.12.1.zip** y **pruebaTP1.xml**. El archivo **Xerces-J-bin.2.12.1.zip** también puede bajarse de <https://d1cdn.apache.org/xerces/j/binaries/>

3) Descomprimir el archivo **Xerces-J-bin.2.12.1.zip** en la carpeta “Documents”. Las bibliotecas java que precisamos son: **xercesImpl.jar**, **xml-apis.jar** y **xercesSamples.jar**

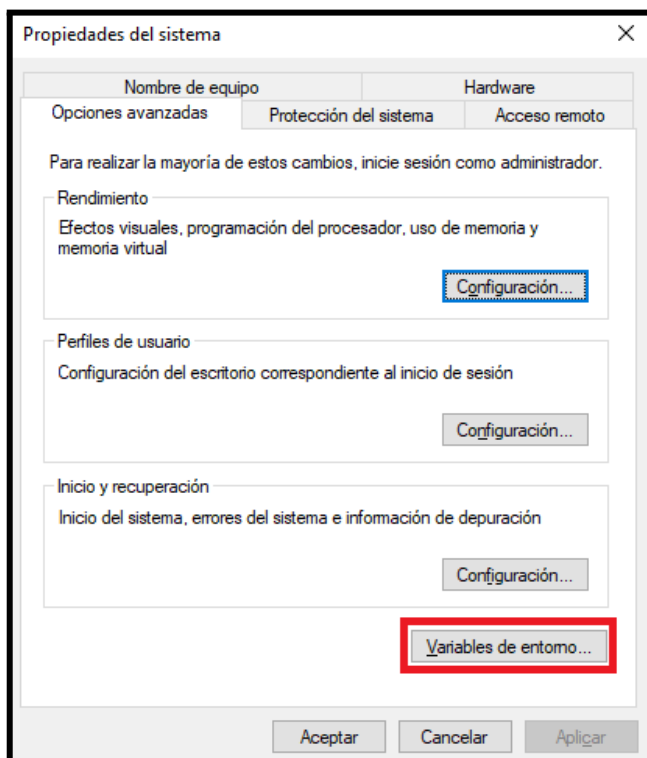
Debemos tener instalado java y seteadas las variables de entorno PATH y CLASSPATH. En el paso siguiente se explica cómo hacerlo en Windows.

4) Configuración de las variables de entorno:

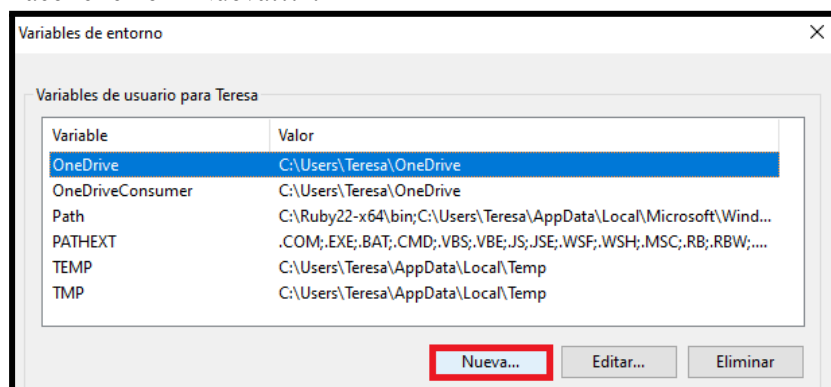
- a. Ir a “Panel de control” y luego ir a Sistema y Seguridad -> Sistema. Allí, seleccionar en el menú de la izquierda la opción “Configuración avanzada del sistema”.



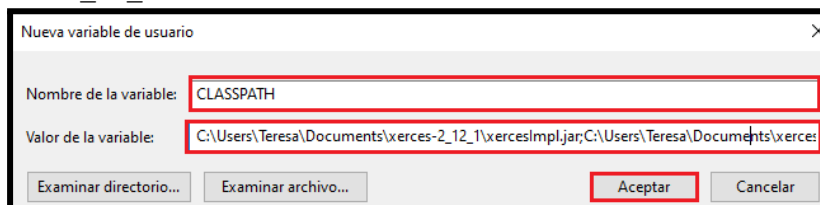
- b. En la ventana “Propiedades del sistema”, hacer click en “Variables de entorno...”.



Hacer click en “Nueva...”.

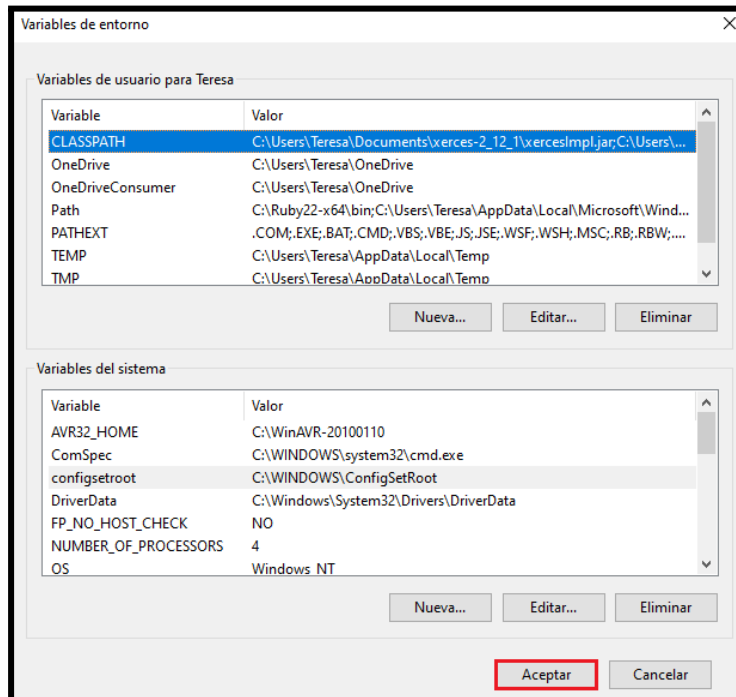


- c. En “Nueva variable de usuario” definir la variable “CLASSPATH”, cuyo valor tienen que ser los paths de las librerías java (xml-apis.jar, xercesImpl.jar y xercesSamples.jar) separados por “;” (punto y coma). Ese decir, que el valor de la variable tendría el siguiente valor (reemplazar <Usuario> por el nombre de usuario correspondiente):
- ```
C:\Users\<Usuario>\Documents\xerces-2_12_1\xercesImpl.jar;C:\Users\<Usuario>\Documents\xerces-2_12_1\xml-apis.jar;C:\Users\<Usuario>\Documents\xerces-2_12_1\xercesSamples.jar
```

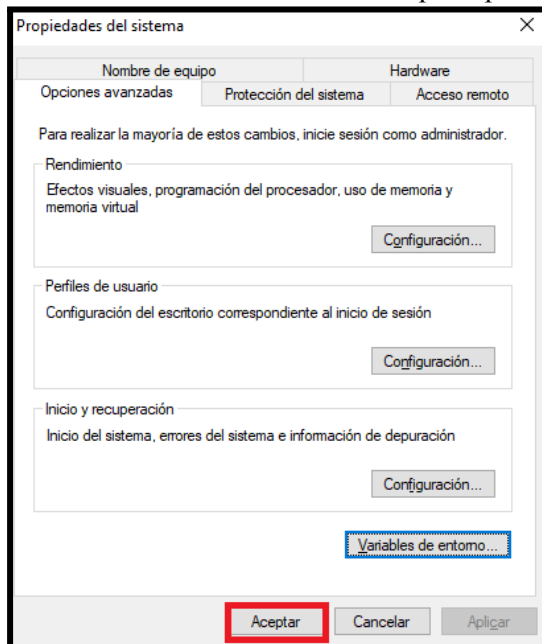


Luego, hacer click en “Aceptar”.

- d. Hacer click en “Aceptar” para cerrar la ventana de “Variables de entorno”.



Y hacer nuevamente click en “Aceptar” para cerrar la ventana de “Propiedades del sistema”.



- e. Una vez creada la variable, abrir una nueva terminal”. La validación (evaluar si el XML está bien formado), se realiza de alguna de estas 2 formas según se quiera usar parser DOM o parser SAX:

```
$ java dom.Writer nombreDocumento.xml
```

O bien

```
$ java sax.Writer nombreDocumento.xml
```

## Pasos para su instalación en Windows con WSL

1) Abrir una terminal de WSL y crear un directorio llamado “XML”.

**Nota:** NO crear la carpeta XML dentro de directorios que se encuentren dentro de iCloud/OneDrive/Google Drive (o cualquier otro sistema de archivos en la nube). En su reemplazo, utilizar el directorio home local del usuario.

2) Bajar a dicha carpeta, desde el Campus ITBA (campus.itba.edu.ar), los archivos: **Xerces-J-bin.2.12.1.zip**, **ejemplo\_bashrc** y **pruebaTP1.xml**. El archivo **Xerces-J-bin.2.12.1.zip** también puede bajarse de <https://dlcdn.apache.org/xerces/j/binaries/>

3) Descompactar el archivo **Xerces-J-bin.2.12.1.zip** en la carpeta “XML”. Las bibliotecas java que precisamos son: **xercesImpl.jar**, **xml-apis.jar** y **xercesSamples.jar**

Debemos tener instalado java y seteadas las variables de entorno PATH y CLASSPATH. En el paso siguiente se explica cómo hacerlo en Linux.

### 4) Configuración de las variables de entorno:

- Mostrar los archivos ocultos de “Carpeta Personal” mediante el explorador de Ubuntu:  
Ver->Mostrar Archivos ocultos.
- Asumimos que hemos creado el directorio XML en nuestra carpeta. Si no es así reemplazar lo indicado en el ítem 3-c) por el path dónde haya descomprimido los 3 archivos que mencionados en el punto 2): **xercesSamples.jar**, **xml-apis.jar** y **xercesImpl.jar**.
- Editar el archivo **.bashrc** de nuestro home. Si no existe, crearlo.

**EL USUARIO “**SALEarno**” DEBE SER REEMPLAZADO POR EL USUARIO DE CADA UNO**

Incluir las siguientes 2 líneas (ver archivo de Campus **ejemplo\_bashrc**)

```
export CLASSPATH=$CLASSPATH:/home/salearno/XML/xerces-2_12_1/xercesSamples.jar:/home/
/salearno/XML/xerces-2_12_1/xml-apis.jar:/home/salearno/XML/xerces-2_12_1/xer
cesImpl.jar
export PATH=$PATH:/usr/lib/jvm/java-7-openjdk-amd64/
```

- Si las variables ya existen y se está editando el archivo (no creándolo), agregar lo indicado usando : (dos puntos).

**Nota:** NO es conveniente llevar al **.bashrc** las líneas del recuadro del punto c) de este documento pdf, mediante **copy&paste**. Se pueden copiar estas líneas desde el archivo **ejemplo\_bashrc.txt** bajado de Campus en el punto 2)

- Grabar el archivo **.bashrc** y reiniciar la sesión para abrir y cerrar la terminal para que los cambios tengan efecto.
- Una vez ejecutado el script, la validación (evaluar si el XML está bien formado), procede de alguna de estas 2 formas según se quiera usar parser DOM o parser SAX:

```
$ java dom.Writer nombreDocumento.xml
```

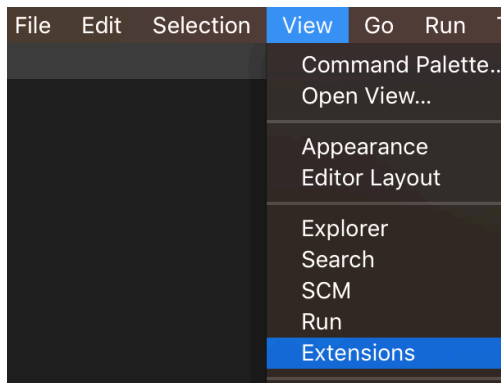
O bien

```
$ java sax.Writer nombreDocumento.xml
```

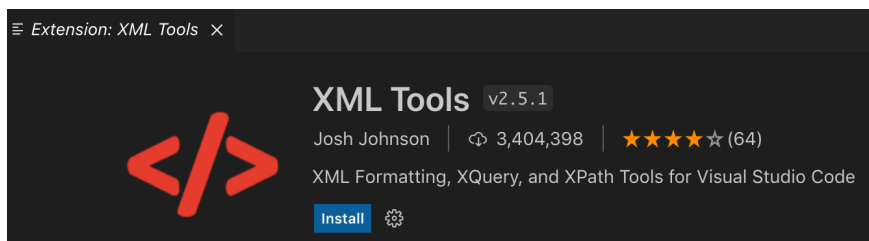
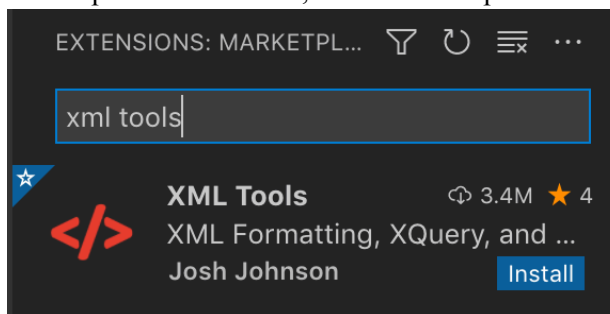
### Instalación de extensiones de XML en Visual Studio Code

En caso de no lograr la instalación correcta del parser con los pasos anteriores (según cada Sistema Operativo), se pueden instalar las siguientes extensiones en [Visual Studio Code](#): XML Tools y XML..

- 1) Abrir Visual Studio Code
- 2) Ir al menú “View”->”Extensions”.

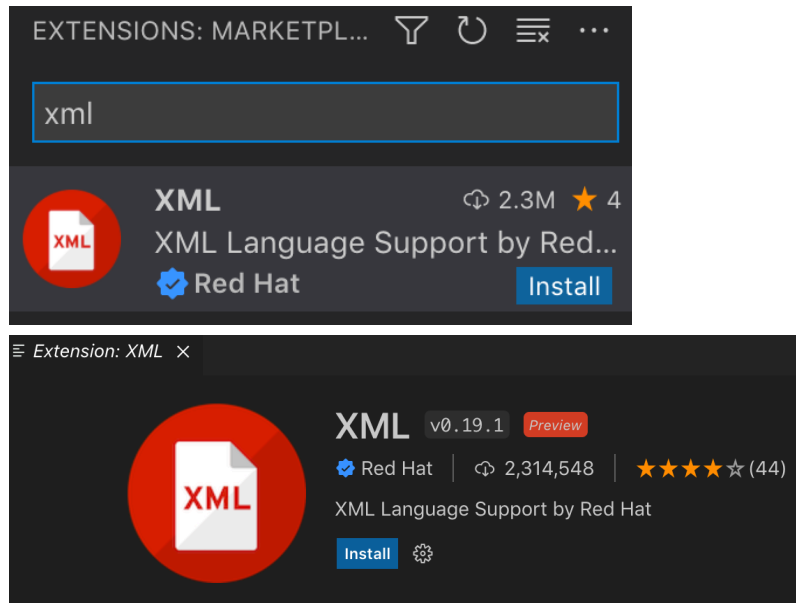


- 3) Buscar por “XML Tools”, seleccionar la primera respuesta de la lista y hacer click en “Install”.



- 4) Buscar por “XML”, seleccionar la primera respuesta de la lista y hacer click en “Install”.





- 5) La validación (evaluar si el XML está bien formado), se realiza creando un documento de XML y verificar que la tab de “Problems” esté vacía.

## Ejercicio 1

**Prueba del parser:** Analizar el contenido del archivo *pruebaTP1.xml* (abriéndolo con cualquier editor de texto), y luego ejecutar desde la terminal:

```
$ java dom.Writer pruebaTP1.xml
```

Notar que no estamos usando la opción `-v` ya que sólo queremos comprobar la estructura del documento xml.

**Prueba desde Visual Studio Code (si no se instaló el parser):** Analizar el contenido del archivo *pruebaTP1.xml* (abriéndolo con Visual Studio Code).

Luego de hacer la prueba:

- 1.1) Si se produce un error, corregirlo hasta obtener como resultado el documento xml correcto.
- 1.2) Identificar la estructura del documento indicando elementos y atributos.

## g. Ejercicio 2

Se quiere definir el tipo de documento *emails.xml* que permita armar un *template* para correos electrónicos. No hay límite en la cantidad de correos electrónicos que puede aparecer en un documento con raíz *emails*. Cada uno de los correos, contiene:

- **date:** fecha de envío del correo electrónico, en formato YYYY-MM-DD. La fecha no puede ser posterior a la fecha actual. Dato obligatorio.
- **header:**
  - **from**
    - **name:** nombre del remitente. Compuesto por nombre/s y apellido/s de la persona.
    - **address:** dirección de correo electrónico válida.<sup>1</sup> Dato obligatorio.
  - **to**
    - **name:** nombre del destinatario. Compuesto por nombre/s y apellido/s de la persona.
    - **address:** dirección de correo electrónico válida. Dato obligatorio.
- **subject:** título del correo electrónico. Texto alfanumérico de hasta 256 caracteres.
- **body:** mensaje. Texto libre.

**2.1** Construir un documento *emails.xml* con al menos tres correos electrónicos y que esté bien formado.

---

<sup>1</sup> <http://tools.ietf.org/html/rfc2822#section-3.4.1>